The Core Mouse Response to Infection by *Neospora Caninum* Defined by Gene Set Enrichment Analyses

John Ellis¹, Stephen Goodswen¹, Paul J Kennedy² and Stephen Bush³

¹School of Medical and Molecular Biosciences and the I3 Institute, University of Technology, Sydney, Broadway, Australia. ²School of Software and Centre for Quantum Computation and Intelligent Systems, University of Technology, Sydney, Broadway, Australia. ³School of Mathematical Sciences, University of Technology, Sydney, Broadway, Australia. Corresponding author email: john.ellis@uts.edu.au

**Abstract:** In this study, the BALB/c and Qs mouse responses to infection by the parasite *Neospora caninum* were investigated in order to identify host response mechanisms. Investigation was done using gene set (enrichment) analyses of microarray data. GSEA, MANOVA, Romer, subGSE and SAM-GS were used to study the contrasts *Neospora* strain type, Mouse type (BALB/c and Qs) and time post infection (6 hours post infection and 10 days post infection). The analyses show that the major signal in the core mouse response to infection is from time post infection and can be defined by gene ontology terms Protein Kinase Activity, Cell Proliferation and Transcription Initiation. Several terms linked to signaling, morphogenesis, response and fat metabolism were also identified. At 10 days post infection, genes associated with fatty acid metabolism were identified as up regulated in expression. The value of gene set (enrichment) analyses in the analysis of microarray data is discussed.

**Keywords:** mouse model, microarray, gene set, host response, immunity, neospora

*Bioinformatics and Biology Insights* 2012:6 187–202
doi: 10.4137/BBI.S9954

This article is available from http://www.la-press.com.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.
Introduction

Microarray technology has been used extensively for the study of host responses to infection by a variety of Apicomplexa. Such studies have used well established methods for mining of data to extract the identity of genes, pathways and other host mechanisms differentially expressed within the data sets under study.\(^1\)\(^-\)\(^4\) The involvement of changes in expression of genes associated with specific pathways is typically investigated by the application of enrichment tests applied to gene ontology terms, thereby providing evidence for enrichment of pathways by association with these terms.\(^5\) Pathway and network analyses potentially offer a number of advantages for analyzing microarray datasets, notwithstanding the identification of specific pathways, and hence networks of genes that are affected by the experimental treatment under study.\(^6\)\(^,\)\(^7\)

More recently, a variety of methods for gene set (enrichment) analysis have evolved, providing an alternative way of mining the transcriptional changes present in microarray data.\(^8\)\(^-\)\(^12\) Gene set analysis essentially asks whether the expression of a list of genes in a microarray data set is positively or negatively associated with one of the two experimental groups (eg, infected and uninfected mice). Gene sets represent lists of genes that are typically associated with a chromosomal location, biochemical pathway, or other computationally derived biological processes or experimentally studied systems such as cancer. Most published studies use gene sets downloadable from the Molecular Signatures (MSigDB) database (http://www.broadinstitute.org/gsea/msigdb/index.jsp), but personalized gene sets can also be created through tools such as WhichGenes.\(^13\)

Various mice types support \textit{N. caninum} infection. These include the BALB/c and Qs that have been described previously.\(^14\)\(^,\)\(^15\) The BALB/c mouse is susceptible to infection by \textit{N. caninum}, whereas the Qs is considerably more resistant to infection by both highly pathogenic and less pathogenic strains such as NC-Liverpool and NC-Nowra respectively. The outcome of infection of BALB/c and Qs mice with \textit{N. caninum} therefore differs. BALB/c mice typically become lethargic and their fur becomes ruffled, after which they may lose weight and may die,\(^14\) whereas the Qs mouse shows none of these signs of infection.\(^15\)

In this study, gene set analysis was used to mine microarray data derived from a previous study that investigated changes in transcription in mice in response to \textit{N. caninum} infection.\(^3\) In that study both BALB/c and Qs mice were used and their response to infection by NC-Liverpool and NC-Nowra investigated. This dataset was previously analyzed by significance analysis of microarrays, clustering, and gene ontology enrichment and these analyses have shown, notwithstanding, the complexity of the mouse response to infection. Here it is shown that gene set analyses provides a simple, yet effective, method by which to mine microarray data. For the record, elements of the core mouse response to infection by \textit{N. caninum} are identified and discussed.

Methods

Data

Microarray data was derived from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) series record GSE23520. The data set is derived from a two color study containing two different mice types (BALB/c and Qs) and two different \textit{N. caninum} (NC-Nowra and NC-Liverpool) isolates, studied at 6 hours post infection (HPI) and 10 days post infection (DPI). The experimental design, which included at least three mice per experimental group, is described elsewhere.\(^3\) Essentially mice were infected with \textit{N. caninum} and RNA prepared from spleen removed at 6 HPI or 10 DPI. mRNA was converted into cDNA, which was hybridized to microarrays containing the NIA15K gene set, using RNA from uninfected BALB/c or Qs mice as a control. The Kooperberg model-based background correction was applied to the raw microarray data, after which the expression log ratios were normalized within the slide (using print-tip loess). Array data were then RMA normalized within treatments prior to linear model fitting.

As gene set analysis requires gene annotation information, all data for genes in the array data set that did not contain annotation were removed, as were data for multiple probes representing the same gene. Annotation information was obtained from ID converter.\(^16\)

Gene sets (GMT format) were obtained from the MSigDB for all categories (human, C1–C5). Gene identifiers used in the array dataset and gene sets were gene symbols. Mouse gene sets were also obtained

---

\(^{1}\) Ellis et al

\(^{2}\) Bioinformatics and Biology Insights 2012:6

---

\(^{1}\) Ellis et al

\(^{2}\) Bioinformatics and Biology Insights 2012:6
from WhichGenes in the GMT format and the WEHI website (http://bioinf.wehi.edu.au/software/MSigDB/index.html). The WEHI gene sets are derived from the MSigDB by conversion of human gene symbols to mouse orthologs, with reference to the Jackson Laboratory Human and Mouse Orthology Report. In these analyses, the mouse gene sets analyzed were restricted to those sets containing between 10 and 500 genes as recommended. A gene set compiled for tissues from the Stanford Microarray Database was also used (http://www.stat.stanford.edu/~tibs/GSA/).

Gene set analysis
All comparisons made involved investigating those gene sets that are associated with either (1) Mouse: meaning mouse type (BALB/c vs. Qs); (2) Type: meaning N. caninum strain (NC-Nowra vs. NC-Liverpool), or (3) Time: meaning time post infection (6 HPI vs. 10 DPI). Data from the GEO series record GSE23520 was pooled for the different groups (Table 1) so that such comparisons could be made across the species N. caninum. For example, in order to compare the response to Qs and BALB/c mice, data from all Qs groups (infected with either N. caninum type and the two different time points) were considered as “Qs”. Similarly all groups from the BALB/c set of infections (infected with either N. caninum type and the two different time points) were considered as “BALB/c”.

Initially gene set enrichment analysis (GSEA) was performed using the JAVA desktop application with the human C1–C5 gene set collections (from the MSigDB), with 1000 permutations of the gene set.

Parametric (PAGE) analyses were performed using the webservice GAzer. 1263 gene sets using the gene ontology (GO biological process and molecular function categories) pathway and chromosome categories included on the server were investigated. Analyses were done with the fold change option, specifying a minimum of five genes/significant gene set to be reported. More specific parametric gene set analyses (PGSEA from Bioconductor) of mouse gene sets were performed in R using gene sets obtained from WhichGenes, reporting only gene sets returning a $P < 0.05$.

Four additional methods of gene set enrichment analysis were used to analyze the microarray data. The first method used to analyze the data was SAM-GS which was introduced as an extension of the significance analysis of microarray approach to accommodate gene sets. In SAM-GS the sum of the squares of the test statistics from a $t$-test (or similar) for individual genes is used to calculate a gene set test statistic. A sample permutation is then used to obtain $P$ values for each gene set. To do this, the R implementation of SAM-GS was used. It is available at http://www.ualberta.ca/~yyasu/homepage.html.

The second method used was multivariate analysis of variance. In the case where there are two genotype groups, a Hotelling’s $T^2$ test is used to determine whether there are any differences in gene expression between the two groups. A number of non-equivalent tests are available if more than two genotype groups are to be compared. Shrinkage covariance matrix estimators are used to overcome the problem of having fewer samples of each phenotype than there are genes in a gene set. An R implementation of MANOVA was used that can be found at http://mail.cmu.edu/~catsai/research.htm.

The third method that was used to analyze the microarray data was subGSE. In this method the genes within each gene set are ranked based on their association with the phenotype group and “strict subsets” of the top $i$ genes are permuted to obtain raw $P$ values, the smallest of which is used to test gene set significance. The advice of the authors of this paper was adopted in that a minimum strict subset size of $C = 5$ genes was used to ensure that the tests are not too sensitive to single genes with strong association. The C++ implementation of this algorithm was used, which can be found at http://www.rcf.usc.edu/~fsun/Programs/SubGSEWebPages/SubGSEMain.html.

The final method that used was Romer, the rotation mean rank version of GSEA for linear models. This method uses a rotation based simulation to obtain $P$ values. This simulation method was introduced in and is discussed in terms of gene analysis elsewhere.
Ten thousand rotations of the mouse gene sets (categories C1–C5 limited to 10 to 500 genes/gene set) were used, which appeared to give reasonably stable gene set rankings between simulations. The limma package in R was used to perform this analysis.

Once the $P$ values for each test were obtained, the $Q$ value package in R was used to obtain the estimated false discovery rates for each gene set within each test. The method in the $Q$ value package is based on the theory of $Q$ values.25,26

In all the analyses described here, gene sets returning a $Q$ value of $<0.25$ were initially recorded and lists of gene sets were sorted according to their $Q$ and $P$ values. In order to reduce the lists down from thousands of gene sets to a more manageable number, gene sets ranked by $P$ values (Manova, SAM-GS, Romer, subGSE) were selected representing the top 10% quantile of the cumulative frequency distribution. Gene lists representing those genes in the gene set were used in the GeneSearch excel macro to extract expression data from the original microarray expression matrix. Gene_Search is a second generation version of the Find_Gene macro described elsewhere.3

The gene lists were then filtered by value so as to identify those genes that met certain criteria based on M values. For example, genes were extracted where $M > 0.1$ and $M < -0.1$ for all four groups in the array data representing 6 HPI or 10 DPI or all eight groups from 6 HPI and 10 DPI together. Venn diagrams were constructed using Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Network analyses

Results obtained from gene set analyses using subGSE (ranked by $P$ value) were displayed by enrichment mapping in cytoscape.17 Descriptors for node clusters were generated using the WordCloud plugin.

Results
GSEA

The expression dataset (GSE23520) has data for 15151 gene probes that after collapsing into gene symbols gave data for 10384 genes. A summary of the number of gene sets that were significantly associated with the three analyses conducted is shown in Table 2. The contrast Time gave a rich number of gene sets for study (258, 111 with $Q < 0.1$), followed by Mouse, showing the strongest informative signal occurring in the dataset at 6 HPI. The main pathways identified were Glycolysis and Gluconeogenesis, Oxidative Phosphorylation and Proteasome Pathway (Table 3). The highest ranked (by NES) gene set (ICHIBA GVHD; not shown) contained 102 genes including Irf1, H2-D1, pbeff, S100A9, Nfkbia, H2-DMA, C1qb, Ly6a, Irf8, Tgtp and Gp49a. C5 gene sets gave results associated with a range of pathways such as Regulation of $IκB$ kinase NF-$κB$ Cascade (Ltbr, Myd88, Litafl, Bcl10, Ikbke), Structural Constituent of Ribosome (39 ribosomal proteins), Immune System Process (Ifitm3, Cd34, Il18, Irf8, Bcl10, Il2rg, Il6st, Cd164, Csf1, Cxcr4, Cd26), Inflammatory Response (S100A9, Nfkb1, Cxcl1, Cxcr4, Tnfap6), and 21 genes of the Mitochondrial Envelope.

A comparison of the C1 sets associated with Type (Neospora) identified a single gene set (CHR13Q14, $Q = 0.0044$) that was negatively associated with NC-Nowra and positively correlated with NC-Liverpool. The leading edge genes in CHR13Q14 were Lcp1, Tpt1, Wbp4, Kpna3, Gtf2f2, and Itm2b. II1R pathway

| Table 2. Summary of GSEA results for $Q < 0.25$. |
| --- |
| **Comparison** | Gene set collections | **Total** |
| **Number of gene sets** | C1 | C2 | C3 | C4 | C5 |
| Neospora type | 94 | 910 | 774 | 565 | 541 | 2884 |
| NC-Nowra | 1 | 32 | 4 | 41 | 10 | 88 |
| NC-Liverpool | 0 | 0 | 0 | 1 | 0 | 1 |
| Mouse type | 0 | 35 | 0 | 119 | 0 | 154 |
| BALB/c | 0 | 0 | 7 | 0 | 0 | 7 |
| Qs | 258 (111$^a$) | 7 | 158 (112$^a$) | 59 | 484 |
| Time post infection | 6HPI | 0 | 0 | 0 | 6 | 0 |
| 10DPI | 0 | 0 | 6 | 0 | 6 |

Note: $^a(Q < 0.1)$.
Table 3. Summary of GSEA results.

| Set collection | Gene seta | Descriptionb | Nc | NESc | Q value |
|---------------|-----------|--------------|----|------|---------|
| **Neospora type** | | | | | |
| C1 | CHR13Q14 | Genes in cytogenetic band chr13q14 | 16 | 1.950 | 0.032 |
| C2 | IL1RPATHWAY | Signal transduction through IL1R Pathway | 15 | 2.080 | 0.043 |
| C2 | HSA03010.RIBOSOME | Ribosome | 54 | 2.030 | 0.042 |
| C2 | VANASSE.BCL2.TARGETS | Differentially genes expressed in murine CD19+ B cells overexpressing Bcl-2 | 37 | 1.990 | 0.050 |
| C3 | GGCNNMSMYNTT.G.UNKNOWN | Genes with promoter regions around transcription start site containing motif GGCNNMSMYNTT-G | 36 | 2.000 | 0.068 |
| C4 | MORF.TPT1 | Neighborhood of TPT1 | 83 | 2.040 | 0.038 |
| C4 | MODULE.32 | Genes in cancer module_32 | 149 | 1.960 | 0.060 |
| C5 | STRUCTURAL.CONS.UTIEN.T. OF.RIBOSOME | Ribosome | 65 | 2.190 | 0.003 |
| C5 | STRUCTURAL.MOLECULE.ACTIVITY | Structural integrity of a complex or assembly | 107 | 1.990 | 0.044 |
| **Mouse type** | | | | | |
| C2 | H2O2.CSBR.ESCU.D.C1.UP | Unregulated by H2O2 in CSB-rescued fibroblasts | 22 | 2.100 | 0.040 |
| C3 | V$AML1.01 | Genes with promoter regions around transcription start site containing a motif which matches annotation for RUNX1 | 72 | 1.970 | 0.036 |
| C3 | V$AML1.Q6 | Genes with promoter regions around transcription start site containing a motif which matches annotation for RUNX1 | 72 | 1.930 | 0.041 |
| C3 | GCAAGAC,MIR-431 | Targets of MicroRNA GCAAGAC,MIR-431 | 15 | 1.840 | 0.088 |
| C4 | GNF2.SMC4L1 | Neighborhood of SMC4L1 | 55 | 2.260 | 0.001 |
| C4 | GNF2.RFC4 | Neighborhood of RFC4 | 38 | 2.240 | 0.001 |
| C4 | GNF2.CKS2 | Neighborhood of CKS2 | 33 | 2.210 | 0.001 |
| C4 | GNF2.BNIP3L | Neighborhood of BNIP3L | 28 | 2.190 | 0.001 |
| C4 | MODULE.54 | Genes in cancer module_54 | 122 | 2.170 | 0.001 |
| C4 | GNF2.FEN1 | Neighborhood of FEN1 | 34 | 2.160 | 0.002 |
| C4 | GNF2.RFC3 | Neighborhood of RFC3 | 27 | 2.090 | 0.004 |
| C4 | GNF2.ANP32B | Neighborhood of ANP32B | 28 | 2.070 | 0.004 |
| C4 | GNF2.RRM1 | Neighborhood of RRM1 | 54 | 2.060 | 0.004 |
| C4 | GNF2.MAP2K3 | Neighborhood of MAP2K3 | 30 | 2.030 | 0.006 |
| C4 | GNF2.DEK | Neighborhood of DEK | 42 | 2.030 | 0.005 |
| C4 | GNF2.MCM4 | Neighborhood of MCM4 | 33 | 2.010 | 0.007 |
| C4 | GMC.RAF1 | Neighborhood of RAF1 | 29 | 1.980 | 0.010 |
| C4 | GNF2.TAL1 | Neighborhood of TAL1 | 26 | 1.980 | 0.010 |
| **Time post infection** | | | | | |
| C1 | CHRXP22 | Genes in cytogenetic band chrXp22 | 22 | 1.780 | 0.239 |
| C1 | CHRSQ22 | Genes in cytogenetic band chr8q22 | 20 | 1.770 | 0.130 |
| C3 | TTCNRGNNNNTT.G.V$HSF.Q6 | Genes with promoter regions around transcription start site containing motif TTCNRGNNNNTT-G | 66 | 1.940 | 0.090 |
| C3 | ATCMNTCCG.Y.UNKNOWN | Genes with promoter regions around transcription start site containing motif ATCMNTCCG-Y | 21 | 1.910 | 0.065 |
| C5 | IMMUNE.SYSTEM.PROCESS | GO:0002376 | 84 | 1.905 | 0.084 |
| C5 | POSITIVE.REGULATION.OF.I.KAPPAB.KINASE.NF.KAPPAB.CASCAD | Activates kappaB kinase/NF-kappaB induced cascade | 33 | 1.887 | 0.078 |

(Continued)
### Table 3. (Continued)

| Set collection | Gene set | Description | q value | \(Q\) value |
|---------------|---------|-------------|---------|--------------|
| C5            | C5      | REGULATION.OF.BINDING | 16 1.899 0.068 | 0.069 |
| C5            | C5      | MITOCHONDRIAL.envelope | 54 1.899 0.067 | 0.067 |
| C5            | C5      | InFLAMMATOrY.reSPOnSe | 26 1.879 0.069 | 0.069 |
| C5            | C5      | POSITIVe.regULATIOn.OF.SIgnAL.TrAnSDUcTIOn | 43 1.821 0.080 | 0.080 |
| C5            | C5      | ceLL.PrOJecTIOn | 45 1.818 0.082 | 0.082 |
| C5            | C5      | IMMUne.reSPOnSe | 56 1.805 0.087 | 0.087 |
| C5            | C5      | Defense.reSPOnSe | 55 1.786 0.094 | 0.094 |
| C5            | C5      | I.KAPPAB.KInAse.nF.KAPPAB.cAScADe reactions initiated by the activation of the transcription factor NF-kappaB | 39 1.781 0.090 | 0.090 |

Notes: A selective representation of gene sets are shown to emphasize the diversity and nature of the host responses affected by *N. caninum*. Description of gene set. The GO term is provided when the gene set name is descriptive. \(Q\) value normalized enrichment score.

Abbreviation: NES, normalized enrichment score.

The GO term is given the highest NES with C2 gene sets and contained genes Map3k7, Ikkbk, Tgfb2, Mapk14, Mapk8, Map2k3, Nfk1, Chuk and Nkkaia as contributing to the core enrichment. MORF_TPT1 had the highest NES of the C4 sets and contained a large number of genes associated with translation including ribosomal proteins Rps5, Rpl5, Rpl7, Rpl31, Rpl15a, Rpl10, Rpl24, Rpl27, Rpl15, Rps25, Rpl12, Rpl22, Rpl38, Rpl11, Rps3a, Rps7, Rps27, Rps4x, Rpl32, Rps14, Rpl18a, Rpl19, Rps10, Rpl10a, Rpl30, Rpl27a and elongation factors Eif4a2, Eif4 g2, Eef1b2, Eef1 g. Gene sets Structural Constituent of Ribosome and Structural Molecule Activity of C5 also gave \(Q\) < 0.1 and also contained ribosomal proteins.

Investigations of Mouse identified gene sets primarily in the C2 and C4 categories. In C4 many of the gene sets contained genes associated with the cell cycle, such as cyclins A2, H, B2 and F plus other associated regulatory molecules such as Cdc6, Cdc28, Cdk2c, Cks1b, Cks2, and Tgfr3.

### Parametric gene set analysis (PAGE)

From the three analyses conducted using GAzer, 166 gene sets were associated with one or more of the contrasts, 66 were associated with Time and Mouse, and 34 were associated with Type. Those gene sets with a \(Q\) < 0.1 are listed in Table 4. Seven of the sets featured in two or more of the PAGE analyses outcomes: Hemoglobin complex, Heme Biosynthesis_GenMAPP, Glutathione metabolism_KEGG, Proteasome KEGG, Proteasome Core Complex (sensu Eukaryota), Immune Response and Mitochondrial fatty acid betaoxidation_GenMAPP.

Both these initial studies using GSEA and PAGE had indicated that Time was rich in correlation with gene sets, indicating that the expression profiles obtained at 6 HPI and 10 DPI are quite different. Further analyses to confirm this observation was performed using PGSEA (in R) with a cut-off \(P\) value of 0.05. Using the Tissues gene set, 34 of them were significantly associated with Time. Gene expression data for genes present in these 34 gene sets were extracted; 20 genes were up-regulated in most (7/8 or 8/8) experimental groups studied. Two of the genes (Bst2, Mc11) are associated with the GO term Regulation of Apoptosis while three were associated with Cell Cycle (Atm, Lats2, Weel). Eighty six genes were primarily up-regulated at 6 HPI while 132 showed...
Table 4. GAzer results for $Q < 0.1$.

| Gene set                                       | N$^a$ | Z$^a$  | Q$^a$  |
|------------------------------------------------|-------|--------|--------|
| **Mouse type**                                 |       |        |        |
| HEMOGLOBIN COMPLEX                             | 6     | 4.900  | 0.0002 |
| POLY(A) BINDING                                | 7     | 4.383  | 0.0031 |
| RECEPTOR BINDING                               | 33    | $-4.113$ | 0.0052 |
| chr1p36.12                                     | 5     | 4.525  | 0.0069 |
| MRNA METABOLISM                                | 7     | 4.301  | 0.0077 |
| OXYGEN TRANSPORTER ACTIVITY                    | 6     | 3.913  | 0.0082 |
| PEROXIDASE ACTIVITY                            | 15    | 3.546  | 0.0263 |
| INORGANIC ANION EXCHANGER ACTIVITY             | 5     | 3.437  | 0.0316 |
| GLUTATHIONE METABOLISM.KEGG                   | 16    | 3.387  | 0.0410 |
| HEME.BIOSYNTHESIS.GenMAPP                     | 8     | 3.516  | 0.0410 |
| CHROMOSOME                                     | 35    | 3.477  | 0.0459 |
| METALLOEXOPEPTIDASE ACTIVITY                   | 6     | 3.2    | 0.0485 |
| OXYGEN BINDING                                 | 8     | 3.186  | 0.0485 |
| RNA SPlicing FACTOR ACTIVITY, TRANSESTERIFICATION MECHANISM | 25    | 3.206  | 0.0485 |
| NUCLEOSOME                                     | 16    | 3.299  | 0.0588 |
| CALCIUM ION BINDING                            | 193   | $-3.019$ | 0.0758 |
| TRANSFORMING GROWTH FACTOR BETA RECEPTOR ACTIVITY | 9     | 2.983  | 0.0766 |
| ANAPHASE-PROMOTING COMPLEX                     | 11    | $-3.125$ | 0.0806 |
| OXIDATIVE.STRESS.GenMAPP                      | 12    | 3.047  | 0.0895 |
| VOLTAGE-GATED POTASSIUM CHANNEL COMPLEX        | 19    | 3.001  | 0.0977 |
| **Neospora type**                              |       |        |        |
| chr12D1                                        | 10    | $-5.290$ | 0.0002 |
| HEMOGLOBIN COMPLEX                             | 6     | $-3.987$ | 0.0123 |
| PROTEASOME.KEGG                                | 14    | 3.866  | 0.0128 |
| chrBo32                                        | 23    | $-4.209$ | 0.0146 |
| IMMUNE RESPONSE                                | 112   | 4.147  | 0.0153 |
| CYTOSKELETAL ANCHORING                         | 5     | $-3.935$ | 0.0189 |
| EPITHELIAL CELL DIFFERENTIATION                | 6     | $-3.801$ | 0.0219 |
| PROTEASOME CORE COMPLEX (SENSU EUKARYOTA)      | 14    | 3.401  | 0.0613 |
| GLUTATHIONE METABOLISM.KEGG                   | 16    | $-3.070$ | 0.0864 |
| HEME.BIOSYNTHESIS.GenMAPP                     | 8     | $-3.006$ | 0.0864 |
| MITOCHONDRIAL.FATTY.ACID.BETA OXIDATION.GenMAPP | 17    | $-2.970$ | 0.0864 |
| FATTY ACID.DEGRATION.GenMAPP                   | 20    | $-2.895$ | 0.0879 |
| FATTY ACID BETA-OXIDATION                      | 13    | $-3.346$ | 0.0934 |
| chr14A2                                        | 9     | $-3.609$ | 0.0992 |
| chr7D2                                         | 6     | 3.577  | 0.0992 |
| **Time post infection**                        |       |        |        |
| PROTEASOME.KEGG                                | 14    | 4.815  | 0.0002 |
| PROTEIN BIOSYNTHESIS                           | 220   | 4.856  | 0.0005 |
| RESPONSE TO DRUG                               | 8     | $-4.681$ | 0.0006 |
| IMMUNE RESPONSE                                | 112   | 4.490  | 0.0010 |
| PROTEASOME CORE COMPLEX (SENSU EUKARYOTA)      | 14    | 4.185  | 0.0012 |
| RIBONUCLEOPROTEIN COMPLEX                      | 163   | 4.388  | 0.0012 |
| RIBOSOME                                       | 133   | 4.185  | 0.0012 |
| RIBOSOMAL.PROTEINS.GenMAPP                    | 147   | 4.216  | 0.0014 |
| SPINDLE                                        | 26    | $-3.911$ | 0.0029 |
| JAK-STAT CASCADE                               | 8     | 4.104  | 0.0044 |
| INOSITOL-TRISPHOSPHATE 3-KINASE ACTIVITY        | 5     | 4.096  | 0.0046 |
| POLY(A) BINDING                                | 7     | $-4.185$ | 0.0046 |
| THREONINE ENDOPEPTIDASE ACTIVITY               | 14    | 4.095  | 0.0046 |
| chr7p22                                       | 9     | 4.571  | 0.0055 |
| ENDOPEPTIDASE ACTIVITY                         | 22    | 3.805  | 0.0117 |
| STRUCTURAL CONSTITUENT OF RIBOSOME             | 152   | 3.738  | 0.0122 |

(Continued)
raised expression at 10 DPI (not shown). These gene lists were further investigated by GO enrichment. At 6 HPI the gene list was linked to terms such as Natural Killer Cell Mediated Cytotoxicity, Glycolysis and Cell Redox Homeostasis amongst others. At 10 DPI the gene list was linked to GO terms associated with a range of Cell Division activities, as well as terms such as Protein Amino Acid Phosphorylation, Cell Respiration and Signaling Pathways.

**SAM-GS and MANOVA**

No gene sets were identified by SAM-GS that correlated with either of the contrasts Mouse or Type, with all gene sets giving $Q > 0.25$. In contrast, all gene sets studied (8965) were significantly associated with Time (post infection), with all $Q$ values less than 0.06 with most being $< 0.003$. Similarly with MANOVA, no gene sets were identified that correlated with either of the contrasts Mouse or Type, with all gene sets giving $Q > 0.32$. 6378 gene sets were significantly associated with Time, with all $Q$ values $< 0.034$.

**Identification of core host response**

Experience with SAM-GS and MANOVA produced very long lists of gene sets that were significantly correlated with Time. In order to identify those gene sets that were highly correlated with Time, gene set analyses was repeated using Romer and subGSE and gene sets in the 10% quantile of each list were extracted and the lists compared. Thirty seven gene sets were common to the lists generated by MANOVA, SAM-GS, Romer and subGSE (Table 5). Visual inspection of the gene lists associated with each gene set showed that some gene sets shared genes for which data was responsible for the enrichment observed. For example, Response to UV and Response to Light Stimulus shared similar lists of genes when the microarray data was considered. Similarly, Cytoplasmic Vesicle Part and Cytoplasmic Vesicle Membrane also shared gene lists. Consequently the presence of unique genes in the 37 gene sets was investigated. Extraction of the gene data was performed using the excel macro Gene_Search and 1951 unique genes were identified in these 37 gene sets. The gene data was further filtered according to value using criteria specific for either 6 HPI, 10 DPI or both (using $M > 0.1$ or $< -0.1$; Table 6). Figure 1 shows a Venn diagram that summarizes the number of genes meeting these criteria. The 10 DPI had higher numbers of genes showing altered expression compared to the 6 HPI time point. The identity of the genes affected was also different.

A small number of genes were identified whose expression involved a greater than two fold change in expression in these 37 gene sets (Table 7). Four genes

### Table 4. (Continued)

| Gene set                                              | N$^a$ | Z$^a$   | Q$^a$   |
|-------------------------------------------------------|-------|---------|---------|
| FERTILIZATION (SENSU METAEOA)                         | 9     | −3.743  | 0.0157  |
| INOSITOL OR PHOSPHATIDYLINOOSITOL KINASE ACTIVITY      | 8     | 3.617   | 0.0163  |
| CARBON FIXATION.KEGG                                   | 11    | 3.358   | 0.0228  |
| OXIDATIVE PHOSPHORYLATION.KEGG                         | 38    | 3.371   | 0.0228  |
| CYTOSOLIC SMALL RIBOSOMAL SUBUNIT (SENSU EUKARYOTA)   | 27    | 3.270   | 0.0270  |
| chr4o42                                               | 18    | 4.030   | 0.0319  |
| PROTEIN FOLDING                                       | 126   | 3.494   | 0.0327  |
| RESPONSE TO TOXIN                                      | 6     | −3.464  | 0.0327  |
| chr18o21                                              | 5     | −3.832  | 0.0362  |
| chrBA1.1                                              | 11    | 3.292   | 0.0535  |
| CASPASE ACTIVATION                                    | 39    | 2.914   | 0.0747  |
| CYTOSOLIC LARGE RIBOSOMAL SUBUNIT (SENSU EUKARYOTA)   | 7     | −3.157  | 0.0763  |
| mRNA METABOLISM                                       | 17    | −2.933  | 0.0778  |
| MITOCHONDRIAL.FATTY.ACID.BETAOXIDATION.GENMAPP        | 21    | −3.118  | 0.0818  |
| CORECEPTOR ACTIVITY                                   | 81    | 3.092   | 0.0818  |
| ELECTRON TRANSPORTER ACTIVITY                         | 173   | −3.038  | 0.0871  |
| NUCLEOTIDE BINDING                                    | 7     | 3.072   | 0.0916  |
| CELLULAR PROTEIN METABOLISM                           | 94    | 2.983   | 0.0939  |

**Note:** *N*, number of genes in set.

**Abbreviations:** Z, Z statistic; Q, q value.
| Gene set                                      | Description                                                         | MANOVA  | SAM-GS | Romer  | subGSE |
|----------------------------------------------|---------------------------------------------------------------------|---------|--------|--------|--------|
| Module.38                                    | Genes in cancer module 38                                            | 0.001   | 0.035  | 0.088  | 0.029  |
| FINETTI.BREAST.CANCER.BASAL.VS.LUMINAL       | Protein kinases in cancer                                            | 0.008   | 0.035  | 0.043  | 0.029  |
| FINETTI.BREAST.CANCER.KINOME.RED             | Protein kinases in cancer                                            | 0.008   | 0.035  | 0.043  | 0.029  |
| BROWNE.HCMV.INFECTION.1HR.DN                 | Genes down-regulated in HCMV infected fibroblasts                   | 0.009   | 0.035  | 0.003  | 0.029  |
| PROTEIN.KINASE.ACTIVITY                      | GO:0004672                                                          | 0.013   | 0.035  | 0.003  | 0.029  |
| HORIUCHI.WTAP.TARGETS.UP                     | Genes up-regulated in primary endothelial cells after knockdown of WTAP | 0.014   | 0.035  | 0.078  | 0.029  |
| Module.37                                    | Genes in cancer module 37                                            | 0.014   | 0.035  | 0.044  | 0.029  |
| CELL.PROLIFERATION                           | GO:0008283                                                          | 0.015   | 0.035  | 0.086  | 0.029  |
| ANATOMICAL.STRUCTURE.MORPHOGENESIS          | GO:0009653                                                          | 0.016   | 0.035  | 0.088  | 0.029  |
| CELLULAR.LIPID.METABOLIC.PROCESS             | GO:0044255                                                          | 0.016   | 0.035  | 0.084  | 0.029  |
| CYTOPLASMIC.vesicle.membrane                 | GO:0030659                                                          | 0.016   | 0.035  | 0.099  | 0.029  |
| CYTOPLASMIC.vesicle.part                     | GO:004433                                                           | 0.016   | 0.035  | 0.099  | 0.029  |
| DNA.RECOMBINATION                            | GO:0006310                                                          | 0.016   | 0.035  | 0.070  | 0.029  |
| DOUBLE.STRAND.BREAK.REPAIR                   | GO:0006302                                                          | 0.016   | 0.035  | 0.029  | 0.029  |
| ENZYME.LINKED.RECEPTOR.PROTEIN.              | GO:0007167                                                          | 0.016   | 0.035  | 0.025  | 0.029  |
| SIGNALING.PATHWAY                            |                                                                     |         |        |        |        |
| GINESTIER.BREAST.CANCER.ZNF217.              | Genes down-regulated in non-metastatic breast cancer                | 0.016   | 0.035  | 0.006  | 0.029  |
| AMPLIFIED.DN                                 |                                                                     |         |        |        |        |
| GLYCEROPHOSPHOLIPID.METABOLIC.PROCESS        | GO:0006650                                                          | 0.016   | 0.035  | 0.036  | 0.029  |
| GLYCOPROTEIN.BIOSYNTHETIC.PROCESS           | GO:0009101                                                          | 0.016   | 0.035  | 0.094  | 0.029  |
| GLYCOPROTEIN.METABOLIC.PROCESS              | GO:0009100                                                          | 0.016   | 0.035  | 0.110  | 0.029  |
| LIPID.BIOSYNTHETIC.PROCESS                  | GO:0008610                                                          | 0.016   | 0.035  | 0.060  | 0.029  |
| MEMBRANE.LIPID.BIOSYNTHETIC.PROCESS         | GO:0046467                                                          | 0.016   | 0.035  | 0.061  | 0.029  |
| MEMBRANE.LIPID.METABOLIC.PROCESS            | GO:0006643                                                          | 0.016   | 0.035  | 0.016  | 0.029  |
| MICROTUBULE                                  | GO:0005874                                                          | 0.016   | 0.035  | 0.115  | 0.029  |
| ORGAN.MORPHOGENESIS                         | GO:0009887                                                          | 0.016   | 0.035  | 0.056  | 0.029  |
| PHOSPHOLIPID.BIOSYNTHETIC.PROCESS           | GO:0008654                                                          | 0.016   | 0.035  | 0.102  | 0.029  |
| PHOSPHOLIPID.METABOLIC.PROCESS              | GO:0006644                                                          | 0.016   | 0.035  | 0.023  | 0.029  |
| POSITIVE.REGULATION.OF.IMMUNE.SYSTEM.PROCESS| GO:0002684                                                          | 0.016   | 0.035  | 0.017  | 0.029  |
| REGULATION.OF.CELL.ADHESION                  | GO:0030155                                                          | 0.016   | 0.035  | 0.109  | 0.029  |
| RESPONSE.TO.LIGHT.STIMULUS                  | GO:0009416                                                          | 0.016   | 0.035  | 0.067  | 0.029  |
| RESPONSE.TO.UV                              | GO:0009411                                                          | 0.016   | 0.035  | 0.006  | 0.029  |
| SPHINGOLIPID.METABOLIC.PROCESS              | GO:0006665                                                          | 0.016   | 0.035  | 0.000  | 0.029  |
| SPINDLE.MICROTUBULE                          | GO:0005876                                                          | 0.016   | 0.035  | 0.052  | 0.029  |
| SPINDLE.POLE                                 | GO:000922                                                           | 0.016   | 0.035  | 0.003  | 0.029  |
| TIGHT.JUNCTION                              | GO:0005923                                                          | 0.016   | 0.035  | 0.052  | 0.029  |
| TRANScription.INITIATION                    | GO:0006352                                                          | 0.016   | 0.035  | 0.028  | 0.029  |
| WOUND.HEALING                               | GO:0024060                                                          | 0.016   | 0.035  | 0.116  | 0.029  |

Notes: *Description of gene set. The GO term is provided when the gene set name is descriptive. P values are sorted by MANOVA results.
were up-regulated at 6 HPI and four at 10 DPI. None were identified in any other possible group. Aak1, which functions in clathrin mediated endocytosis, showed the highest fold change (81 fold at 10 DPI) in BALB/c mice infected with NC-Liverpool. Expression of Acadvl, associated with mitochondrial long chain fatty acid beta-oxidation in the mouse, was raised in mice at 10 DPI infected with NC-Liverpool. The significance of fatty acid metabolism at 10 DPI is discussed below.

Discussion
Gene set enrichment is a method that investigates whether predefined sets of genes are differentially expressed between two sets of conditions (eg, infected vs. uninfected) present in microarray data. The most commonly used method is gene set enrichment analysis (GSEA). However, debate and the recognition of shortcomings in gene set analyses has led to the emergence of other approaches such as SAM-GS and MANOVA for gene set analysis.

The microarray data sets used in this study were derived from a series of experiments that involved infection of Qs and BALB/c mice with N. caninum (NC-Nowra or NC-Liverpool strains). Two different time points (6 HPI and 10 DPI) were studied, providing an insight into the host responses occurring early during infection. Qs mice are relatively resistant to infection by N. caninum, whereas responses in BALB/c are strain specific. NC-Liverpool, for example, is very pathogenic in the BALB/c mouse leading to weight loss, appearance of clinical signs such as head tilting and limb paralysis, and death. The comparison of these groups (Type: Nc-Nowra v NC-Liverpool, Mouse: Qs v BALB/c and Time: time post infection) should provide further understanding of an animal’s responses to infection by N. caninum and the mechanisms associated with disease and resistance.

Using the same data set, it was previously demonstrated that the transcriptional responses occurring in the spleen of mice was dependent on a number of factors including the strain of N. caninum used, as well as the mouse type and time post infection. The methods of differential gene analyses used included significance of microarrays, ANOVA and clustering methods. Alternatively, Bayes statistics using the functions lmFit, eBayes and topTable found in limma were used in association with gene enrichment methodologies that measured functional enrichment (of gene ontology terms). These approaches identify lists of genes that are assigned to biological processes and functions via the gene ontology language. In contrast the gene set approaches described here represent an alternative approach for mining microarray data. It is argued that informative signals, derived from multiple genes associated with a pathway for example, may be more easily identified than those associated with single genes alone. Such approaches may be beneficial in analyzing data sets where an association with a treatment or phenotype has yet to be identified. Consequently the mouse response to infection by N. caninum was examined here, in anticipation that gene set analyses would provide further insight into identifying host responses that are associated with neosporosis.

GSEA and PAGE were initially used to mine the expression data. The main reason behind this choice was that easily used web servers are available that can

| Change in expression | Time post infection |
|----------------------|---------------------|
|                      | Six | Ten | Six and ten |
| Up                   | 170 | 329 | 29 |
| Down                 | 120 | 219 | 9 |

Note: *Defined as a change in M value as <0.1 or >–0.1.
be used to rapidly analyze data by gene set analyses. Despite the limitations of these approaches, including use of human gene symbols in the gene sets, they identified that the largest number of gene sets were correlated with Time (post infection) rather than Mouse or Type. Subsequently, gene set analyses were conducted by SAM-GS, MANOVA, Romer and subGSE. The number of gene sets detected that correlated with the microarray expression data was very much dependent on the method used for gene set (enrichment) analyses, as well as the definition of the minimum number of genes to be included. SAM-GS, for example, found no correlations with Mouse or Type. Using expression data merged from both Qs and BALB/c mice types infected with either NC-Liverpool or NC-Nowra, the analyses showed that the host response is quite different at these two time points. Similar observations were made with all methods of analysis. For example, GSEA identified a range of gene sets with an immunological basis such as inflammatory responses and NF-κB signaling.

The two time points chosen (6 HPI and 10 DPI) were based on the previous observations of others concerning the mechanisms of innate and adaptive immunity to *N. caninum* in the mouse. For example, γ-interferon is known to be one response molecule produced at these time points. PAGE identified the Jak-Stat cascade as one of the significant gene sets. Overall, these results indicated the timing of the host response (Time) needed to be further investigated for its importance in determining infection outcomes in terms of disease.

GSEA and PAGE both identified significant differences in the expression data derived from mice infected with NC-Nowra or NC-Liverpool (that is they were correlated with strain of *N. caninum*) suggesting that the mouse response to infection by these two strains is different. The two methods identified very different gene sets that differed between the groups. GSEA identified differences in molecules affecting translation (eg, ribosomal proteins) along with MAP kinase activity whereas PAGE suggested fatty acid metabolism differs along with proteasomal activity (plus others). SAM-GS and MANOVA found no associations in this category. The idea that fatty acid metabolism is influenced by the *Neospora* strain represents just one example where gene set analyses has provided new hypotheses to explore. The BALB/c and Qs mice differed, according to GSEA, by the expression of genes associated with the cell cycle. GAzer identified haemoglobin/heme-metabolism/oxygen-related gene sets as being significantly different between these mice types. Peroxidase and glutathione metabolism are also in the list produced by PAGE, identifying that redox metabolism differs between mouse types.

Overall the results obtained by GSEA and PAGE were similar with those obtained previously by analyses of individual gene data by SAM, clustering and ANOVA, followed by enrichment analyses of gene lists based on GO. The advantages of gene set analysis are, however, evident—unlike analyses of individual genes, it is advantageous to identify several genes of a pathway (gene set) that is altered by the experimental treatment, thereby flagging those pathways for future study. There are also, however, several drawbacks associated with gene set analysis. In the first instance, the presence of a differentially

### Table 7. List of genes with a twofold (or greater) change in expression identified from the 37 gene sets of Table 5.

| Gene   | BALB/c Liverpool | BALB/c Nowra | Qs Liverpool | Qs Nowra |
|--------|------------------|--------------|--------------|----------|
| 6 HPI (up) |                  |              |              |          |
| B3gnt5  | 2.2              | 2.4          | 2.1          | 4.3      |
| Cnn3    | 2.4              | 2.7          | 2.2          | 3.3      |
| S100a9  | 2.5              | 3.0          | 2.6          | 4.6      |
| Txn1    | 2.4              | 3.5          | 2.0          | 4.1      |
| Aak1    | 1.0              | 0.8          | 0.9          | 1.1      |
| Acadvl  | 1.2              | 1.3          | 1.8          | 1.1      |
| Adipor2 | 1.1              | 1.2          | 1.0          | 1.0      |
| Mest    | 0.9              | 1.1          | 1.5          | 1.2      |
| Ttk     | 0.8              | 0.9          | 1.1          | 0.7      |

**Note:** The fold change is determined from M values and represents the fold change when comparing infected to un-infected mice.
expressed gene in more than one gene set means that several of the associations found can occur simply because of the impact of gene membership on a gene set. An example can be found here in those gene sets that contain ribosomal proteins, which occurred in more than one gene set. The algorithms themselves have also come up for criticism. GSEA was shown to be subject to false positive and negative findings, and PAGE ignores gene-specific variances.

Methods for gene set (enrichment) analyses are typically grouped as two types, competitive or self-contained, with the later gaining widespread popularity based on logical criteria. SAM-GS, MANOVA and subGSE are examples of self-contained methods for finding gene sets associated with two groups under study. The approach adopted here for identifying the mouse core responses was to select the top ranking gene sets identified by each of these analyses (including Romer) and to simply determine those present in the top 10% quantile of each. In this manner 37 gene sets containing 1521 unique genes were identified as featuring in the mouse response to *N. caninum*.

Host responses to *N. caninum* are known to be of the Th1-type and the present dogma is that resistance to infection is mediated via IFN-γ. Similar to those anti-parasitic mechanisms observed in *T. gondii*, host responses to *N. caninum* are shown in this and the accompanying studies to be extremely diverse in their nature. Of note is the statistical significance supporting claims for involvement of pathways associated with MyD88 and NF-κB, as well as Jak-Stat signaling in the mouse response, for which experimental evidence is now present. With *T. gondii*, mouse responses are also based on toll-receptor MyD88, NF-κB, and MAP kinase signaling, resulting in defined inflammatory responses. It is reassuring that gene set analyses has identified similar pathways, thereby providing a high degree of confidence in the results presented here and the claims behind the association of other pathways and mechanisms in the mouse response to *N. caninum*.

Finally, it is now possible to provide a more detailed, albeit general, summary of the core responses of the mouse in response to infection by *N. caninum*. The influence of γ-interferon on gene expression is extensively described and linked to a vast number of responses including those of dendritic and other antigen-presenting cells, natural killer cells, macrophages and T helper and Treg cells, to name just a few. Systems biology approaches have led to the curation of the widespread influence of γ-interferon on gene expression; 31 of the top 50 network hubs (genes) of the γ-interferon network were present in the dataset studied here. The fold changes in expression associated with them were relatively small (generally in the 1.1–3 range, eg, Nfkbia and Irf8 were increased across all the groups studied). Five of the hub genes (Irf1, Irf3, Ctnnb1, Raf1, Map3k7) were reduced in expression at 10 DPI by up to 50% of the level shown by uninfected mice (not shown). Text mining using SciMinder identifies 1562 genes linked to a search through the keyword “gamma interferon”; 355 were present on the arrays used here. Only five (Stat1, Irf1, Cnd2, Lap3, Nod1) showed a greater than twofold increase in expression at either of the two time points studied and all were at 6 HPI.

Table 5 summarizes the identity of 37 gene sets associated with Time, identified by MANOVA, SAM-GS, Romer and subGSE, which define the core mouse response to infection by *N. caninum*. From a GO perspective there are a number of significant terms in this list, such as Protein Kinase Activity, Cell Proliferation and Transcription Initiation, which reflect core activities differing between the two time points post infection. The word clouds in Figure 2 attempt to summarize the simple terms associated with the core responses identified by just one of the methods used for gene set analyses (subGSE). Although the different methods of gene set enrichment are likely to generate slightly different word clouds as a result of the different results obtained from the enrichment analyses, subGSE was selected for illustration purposes only. Using KEGG based gene sets, two major nodes are observed in the enrichment map composed of transcription and regulation of metabolic process. Gene ontology gene sets provide word clouds focused on nodes related to regulation and protein. In the latter, regulation is linked to a variety of nodes describing functions such as Apoptosis, Programmed Cell Death, Signaling, Transduction, Kinase, Cascade and ikappaB. The protein node is connected to a wide number of other nodes describing protein functions such as Transport, Localization, Modification and Metabolic.

At 6 HPI four genes were identified that showed at least a twofold change in expression in response to
Host responses to neospora caninum infection

Figure 2. Wordcloud network derived from the subGSE analysis of time. 651 gene sets ($P = 0.029$) were selected and an enrichment map derived using (A) gene ontology or (B) KEGG as the source of gene sets.

**Notes:** The Wordcloud plugin was used to generate a network of predominant words linked to the nodes present in the enrichment map. The size of the node term is scaled to the frequency of the term usage for that node; the width of the edges linking nodes reflects the word similarity score between the nodes.

Infection by *N. caninum*. Thioredoxin (Trx1) is a fundamental component of the pathways that maintain redox homeostasis. B3gnt5 is crucial for development of B cells in spleen. Cnn3 regulates phagocyte motility, and S100A9 is secreted by both neutrophils and monocytes early during infection.

An interesting outcome of these analyses is the observed mouse response associated with fatty acid metabolism at 10 DPI. Acadvl, Adipor2 and Mest were all raised significantly in expression at 10 DPI in comparison to uninfected mice. Acadvl is a mitochondrial, very long-chain specific acyl-CoA dehydrogenase involved with the initial steps of fatty acid β oxidation that generates ATP in mitochondria. Adipor2 is a receptor for adiponectin, an anti-inflammatory adipocytokine produced by adipocytes. Adipor2 is typically expressed in the liver and disruption of Adipor2 results in decreased PPAR-α signaling and increased inflammation and oxidative stress, ultimately leading to glucose intolerance. Mest is induced in response to dietary fat and knock down of Mest expression prevents adipogenesis. Aak1 also showed raised expression in response to infection by *N. caninum* at 10 DPI; this Ser/Thr kinase triggers clathrin assembly during clathrin-mediated endocytosis, which is also a feature of adiponectin signaling. Such observations suggest a direct effect of infection by *N. caninum* on fat cells, as well as metabolism of fatty acids.

Coincidentally, recent research using a new animal model (the fat-tailed dunnart) has provided direct evidence that the mass of body fat is dramatically reduced during the course of infection of a susceptible animal by *N. caninum*. Obviously there are important leads here to investigate further. For example, BALB/c mice infected with NC-Liverpool also tend to loose body weight rapidly from about day 10 DPI and this may also be associated with loss of body fat.

Another of the novel observations made here is the relatively large number of genes in these 37 gene sets that are associated with mammalian development and embryogenesis. This is demonstrated in the word cloud of Figure 2A as the group of nodes in the bottom left hand corner linked to development. Module 38 and the sets linked to Morphogenesis are examples of gene sets containing such genes. Mest is extensively expressed in fetal tissues, while Acadvl is also widely expressed. S100A9 is a proinflammatory mediator secreted by leukocytes at sites of infection.
or injury. Studies have also implicated the molecule in control of intrauterine infections, as well as the onset of labor.

A developing theme is therefore that genes associated with host responses to pathogens are also associated with reproduction and fetal development. The multifunctional role of proteins such as those discussed here shows this to be a reasonable assertion. Another example is that of TGF-β, which is produced in response to infection but is also involved in a wide range of other processes including remodeling of the feto-maternal interface. A mechanism of molecules possessing many different functions may well represent a means for preserving the health of the pregnant animal during infection, at the potential expense of the unborn fetus. That pregnant mice often resorb fetuses in response to infection is well known and a simple illustration of this.

Fetal death and abortion are the main clinical signs observed in cattle following infection by *N. caninum* and it is believed that the route and timing of infection determines the outcome of the pregnancy. Recently, progress in this area from studies on cattle indicates placental function may contribute to control of infection by *N. caninum* rather than simply being deleterious to fetal survival. Similarly, studies on fetal immunity suggest the timing of infection in relation to development of fetal immune competence is a key process in determining the outcome of infection on pregnancy. The studies described here and elsewhere provide additional leads to explore during investigations of cattle responses to *N. caninum*, especially during pregnancy. In cattle, there is evidence that liver Acadvl expression is correlated with serum nonesterified fatty acid levels. A duodenal infusion of alpha-linolenic acid into dairy cattle was also shown to have immunomodulating activity that was associated with changes in γ interferon. The link between fatty acid metabolism and inflammation is obviously one of the more important areas to be explored further. As it is recognized that immune responses differ between adult and fetus, studies on fetal immunity may also provide the clues needed to understand the link between neosporosis and fetal death and abortion.

Acknowledgements

We thank Qi Liu and Yutaka Yasui (University of Alberta) and Xiting Yan (Yale University) for their patience and assistance with installing SAM-GS and subGSE respectively.

Funding

This research was funded by the University of Technology, Sydney.

Competing Interests

JE’s institution has patented previous work on Neospora caninum. Other authors disclose no competing interests.

Author Contributions

Conceived and designed the experiments: JE, SB. Analyzed the data: JE, SB, SG. Wrote the first draft of the manuscript: JE, SB. Contributed to the writing of the manuscript: JE, SB, SG, PK. Agree with manuscript results and conclusions: JE, SB, SG, PK. Jointly developed the structure and arguments for the paper: JE, SB, SG, PK. Made critical revisions and approved final version: JE, SB, SG, PK. All authors reviewed and approved of the final manuscript.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Blader IJ, Manger ID, Boothroyd JC. Microarray analysis reveals previously unknown changes in Toxoplasma gondii-infected human cells. *J Biol Chem*. Jun 29, 2001;276(26):24223–31.
2. Morrison DA, Ellis JT. The design and analysis of microarray experiments: applications in parasitology. *DNA Cell Biol*. Jun 2003;22(6):357–94.
3. Ellis J, Sinclair D, Morrison D, et al. Microarray analyses of mouse responses to infection by Neospora caninum identifies disease associated cellular pathways in the host response. *Mol Biochem Parasitol*. Dec 2010;174(2): 117–27.
53. McMinn J, Wei M, Sadovsky Y, Thaker HM, Tycko B. Imprinting of PEG1/MEST isoform 2 in human placenta. *Placenta*. Feb-Mar 2006;27(2–3):119–26.

54. Oey NA, Ruiter JP, Ijlst L, et al. Acyl-CoA dehydrogenase 9 (ACAD 9) is the long-chain acyl-CoA dehydrogenase in human embryonic and fetal brain. *Biochem Biophys Res Commun*. Jul 21; 2006;346(1):33–7.

55. Park SJ, Yoon WG, Song JS, et al. Proteome analysis of human amnion and amniotic fluid by two-dimensional electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proteomics*. Jan 2006;6(1):349–63.

56. Havelock JC, Keller P, Muleba N, et al. Human myometrial gene expression before and during parturition. *Biol Reprod*. Mar 2005;72(3):707–19.

57. Sugawara K, Kizaki K, Herath CB, Hasegawa Y, Hashizume K. Transforming growth factor beta family expression at the bovine feto-maternal interface. *Reprod Biol Endocrinol*. Oct 15, 2010;8:120.

58. Reichel MP, Ellis JT. Neospora caninum—how close are we to development of an efficacious vaccine that prevents abortion in cattle? *Int J Parasitol*. Sep 2009;39(11):1173–87.

59. Williams DJ, Hartley CS, Bjorkman C, Trees AJ. Endogenous and exogenous transplacental transmission of Neospora caninum—how the route of transmission impacts on epidemiology and control of disease. *Parasitology*. Dec 2009;136(14):1895–900.

60. Rosbottom A, Gibney H, Kaiser P, et al. Up regulation of the maternal immune response in the placenta of cattle naturally infected with Neospora caninum. *PLoS One*. Jan 19, 2011;6(1):e15709.

61. Loor JJ, Dann HM, Everts RE, et al. Temporal gene expression profiling of liver from periparturient dairy cows reveals complex adaptive mechanisms in hepatic function. *Physiol Genomics*. Oct 17, 2005;23(2):217–26.

62. Sun P, Wang J, Yang G, Khas E, Liu Q. Effects of different doses of free alpha-linolenic acid infused to the duodenum on the immune function of lactating dairy cows. *Arch Anim Nutr*. Dec 2010;64(6):504–13.

63. Witkin SS, Linhares IM, Bongiovanni AM, Herway C, Skupski D. Unique alterations in infection-induced immune activation during pregnancy. *BJOG*. Jan 2011;118(2):145–53.

64. Williams DJ, Trees AJ. Protecting babies: vaccine strategies to prevent foetopathy in Neospora caninum-infected cattle. *Parasite Immunol*. Mar 2006;28(3):61–7.