12-2008

Altered developmental programming of the mouse mammary gland in female offspring following perinatal dietary exposures : a systems-biology perspective.

Caleb Deen Bastian
University of Louisville

ALTERED DEVELOPMENTAL PROGRAMMING OF THE MOUSE MAMMARY
GLAND IN FEMALE OFFSPRING FOLLOWING PERINATAL DIETARY
EXPOSURES: A SYSTEMS-BIOLOGY PERSPECTIVE

By

Caleb Deen Bastian
B.S., University of Tennessee, 2006
M.S., University of Tennessee, 2006

A Thesis
Submitted to the Faculty of the
Graduate School of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science

School of Dentistry
Master of Science in Oral Biology Program
University of Louisville
Louisville, Kentucky

December 2008

ALTERED DEVELOPMEMENTAL PROGRAMMING OF THE MOUSE
MAMMARY GLAND IN FEMALE OFFSPRING FOLLOWING PERINATAL
DIETARY EXPOSURES: A SYSTEMS-BIOLOGY PERSPECTIVE

By

Caleb Deen Bastian
B.S., University of Tennessee, 2006
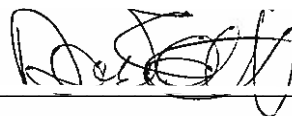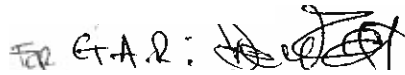M.S., University of Tennessee, 2006

A Thesis Approved on

November 21st, 2008

By the following Thesis Committee:

_____

Thesis Director

_____

_____

DEDICATION

This thesis is dedicated to my parents

John S. Bastian, DDS, PC

and

Paula Bastian

who have given me invaluable support and educational opportunities

ACKNOWLEDGEMENTS

ABSTRACT

ALTERED DEVELOPMEMENTAL PROGRAMMING OF THE MOUSE
MAMMARY GLAND IN FEMALE OFFSPRING FOLLOWING PERINATAL
DIETARY EXPOSURES: A SYSTEMS-BIOLOGY PERSPECTIVE

CALEB DEEN BASTIAN

November 21st, 2008

Mishaps in prenatal development can influence mammary gland development
and, ultimately, affect susceptibility to factors that cause breast cancer. This research was
based on the underlying hypothesis that maternal dietary composition during pregnancy
can alter developmental (fetal) programming of the mammary gland.We used a
computational systems-biology approach and Bayesian-based stochastic search variable
selection algorithm (SSVS) to identify differentially expressed genes and biological
themes and pathways. Postnatal growth trajectories and gene expression in the mammary
gland at 10-weeks of age in female mice were investigated following different maternal
diet exposures during prenatal-lactational-early-juvenile development. This correlated a
decrease in expression of energy pathways with a reciprocal increase in cytokine and
inflammatory-signaling pathways. These findings suggest maternal dietary fat exposure
significantly influences postnatal growth trajectories, metabolic programming, and
signaling networks in the mammary gland of female offspring. In addition, the
adipocytokine pathway may be a sensitive trigger to dietary changes and may influence
or enhance activation of an immune response, a key event in cancer development.

Table of Contents

LIST OF TABLES

1. INTRODUCTION

Over fifteen years ago, Barker and colleagues hypothesized that adverse environmental factors in early life, particularly nutritional deficiencies, disrupt fetal growth and development, leading to a more susceptible adult phenotype prone to cardiovascular disease (Barker, Bull et al. 1990; Dobbing 1993; Barker 1995). After a worldwide series of epidemiological studies, in which fetal growth restriction was demonstrated to be correlated with chronic disease in adulthood, this hypothesis became known as the developmental origins of health and disease (DOHAD) or Barker hypothesis. It states that environmental factors act in early life to program the risks for major chronic diseases, such as cardiovascular disease, stroke, hypertension, type 2 diabetes, and obesity later in life.

It is well recognized that during development, there are critical periods of vulnerability to suboptimal conditions when developmental programming may permanently modify disease susceptibility (Lucas 1998). Developmental programming involves structural changes in important organs: altered cell number, imbalance in distribution of different cell types within the organ, and altered blood supply or receptor numbers (Armitage, Khan et al. 2004). Tissues that have been identified as targets for programming include heart, kidney, liver, pancreas, and adipocytes (McMillen and Robinson 2005). For the mammary gland, data on developmental programming are still limited. Changes in maternal diet have been demonstrated to alter mammary gland

differentiation (Hilakivi-Clarke, Stoica et al. 1998; De Assis and Hilakivi-Clarke 2006). This might influence further development of the mammary gland and, ultimately, play an important role in determining later risk to breast cancer through epigenetic mechanisms (i.e. ubiquitinization, phosphorylation, chromatin methylation, histone acetylation, sumovlation). A link between maternal nutrition in pregnancy and breast cancer risk in adult life has been proposed but not established.

The goal of the study presented here was to identify cellular-response pathways underlying fetal programming of the murine mammary gland by nutrition. The Barker hypothesis has mainly focussed on the impact of maternal undernutrition. However, maternal and postnatal nutrition are either sufficient or excessive in developed countries. In fact, not only undernutrition, but also a disbalance of nutrition in the opposite direction correlates with adult diseases, such as cancer. Breast cancer incidence rates are about fourfold higher in Europe and North America than in Asia and Africa (Hortobagyi, de la Garza Salazar et al. 2005; Parkin, Bray et al. 2005). Besides other lifestyle factors, consumption of a Western diet, and more specifically dietary fat intake, has been demonstrated to be a major risk factor (Kato, Tominaga et al. 1987; Mattisson, Wirfalt et al. 2004; Binukumar and Mathew 2005). Epidemiological studies have suggested a positive relation between dietary fat intake and breast cancer risk (Howe, Hirohata et al. 1990; Prentice and Sheppard 1990; Boyd, Stone et al. 2003), and this has also been demonstrated in rodents. Here, we investigated the effect of different perinatal diet exposures, using high-fat diets high in n-6 or in n-3 polyunsaturated fatty acids, on programming of the murine mammary gland (Figure 1) by studying gene expression profiles in relation to body weight trajectories. This thesis comprises a computational

analysis of a recent dataset from a collaboration between the University of Louisville and

RIVM, the Netherlands.
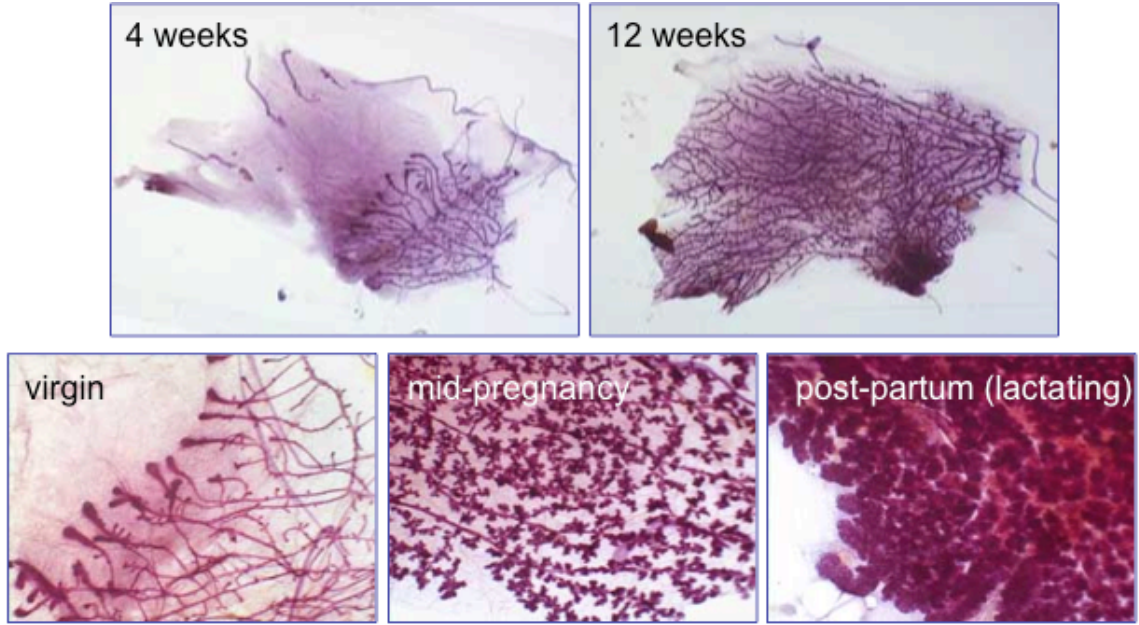
Figure 1. Mouse mammary tree

4

2. THEORY (Singh, Rouchka et al. 2007)

2.1. Systems Biology Research Strategies

Systems biology is an interdisciplinary science that aims at system-level understanding of biological systems (biosystems). It has been only recently, through progress in the human genome project and advancement of newer technologies for high-content data generation, that a system-level analysis has been enabled at the molecular level. Grounded in the coupling of experimentation (in vivo, in vitro) with computation (in silico), this approach has applications toward building predictive models that are integrative, quantitative, and dynamic in nature. A general workflow paradigm is diagrammed in Figure 2 .
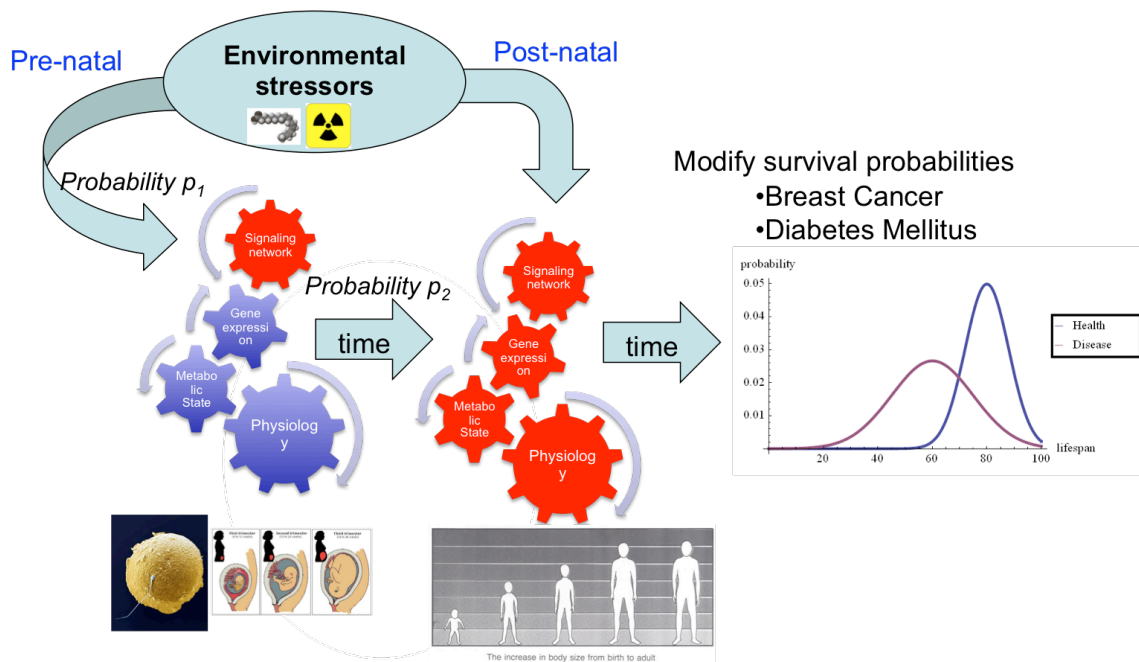
Figure 2. Workflow for inference and prediction in a systems biology paradigm

Biological attributes observed from hypothesis-driven studies on the natural system are used to make inferences at different informational levels; this information is used to formalize the system with mathematical models. These models are tested with simulated and empirical data to then predict behaviors in the natural system. The process is iterated to improve performance and accuracy (adapted with permission from Dr. Wolkenhauer).

To integrate a biological system both top-down and bottom-up strategies must be considered. The former starts with the system as a whole and decomposes it to smaller modular entities, e.g., phenotypes molecules. The strength in a top-down view of birth defects is that overall knowledge of the system can be broken down based on clinical phenotype, independently of detailed molecular knowledge. The weakness, however, is that not all critical components or interactions may be known at the cellular and molecular levels. A bottom-up strategy rather starts with basic entities (genes, proteins) and integrates them into relevant patterns and functions (pathways, networks). The strength in this approach is that detailed molecular knowledge is available for a large number of molecular entities, and the weakness is the limited capacity to project the effect of perturbations in pathway models onto the cell as a whole.

Given that a cell is the basic functional unit of a tissue, an ideal modeling framework should serve both strategies by providing a perspective that can be scalable down to the level of individual genes and scalable up to the level of the developing system or embryo. Fundamental epistemological issues facing systems biology were explored by O'Malley and Dupré in their recent BioEssay article (2005). These authors discuss two fundamental philosophies: pragmatic and theoretical. In the pragmatic paradigm, scientists use local data to construct pathways and to structure networks based on current knowledge. In Figure 2, for example, we represent the pragmatic school by the vertical integration of data from signaling pathways to physiological states. Exposure to various teratogens that give rise to a particular malformation or syndrome of malformations can be monitored by high-content technology platforms yielding diverse

data. This philosophy can be applied in top-down (phenotypes → molecules) or bottom-up (molecules → phenotypes) mode, although in principle the genome is given the highest causal and informational priority over other informational levels.

In the theoretical paradigm, priority is given to higher-level processes and properties of a biosystem and the genome is deprioritized. In Figure 2, for example, the natural system is observed by experimentalists using a traditional hypothesis-driven scientific method to yield information on various biological attributes at different information levels, e.g., molecular abundance profiles, gene-product association networks, metabolic and regulatory pathways, biomarkers of cellular processes and function, and clinical (physiological or anatomical) phenotypes. The morass of such data require expertise in bioinformatics, statistics, and applied mathematics to make inferences about the overall system properties and build models of the formal system that can lead to meaningful predictions of the behavior of the system under different parameters.

In their BioEssay paper, O'Malley and Dupré (2005) raise three fundamental questions with regard to the basic theme of integration common to both schools of systems biologists: what is a system (structure); what biological units map onto a system (modules); and how are modular response characteristics regulated (control) and/or constrained by the system (logic)? A system may be defined generally as a group of entities comprising a whole, in which each component interacts with or is related to at least one other component functioning together to achieve the same objective. Enumerating components of the system is important for understanding of structure of the system, such as network topologies; however, the essence of a system resides in its dynamics. These dynamics must be studied in quantitative terms to construct models with

predictive capability. It would, however, be misleading to focus on system structure-dynamics without paying attention to the diversity and functionality of its component parts. In the real world, objects may be part of the system environment but not necessarily required by the system to operate. Consider a light bulb inside a refrigerator unit: function of the light bulb may enhance the unit's environment but it is not critical for the unit's main function.

2.2. Scaling, Details, and Integration

Complex biological systems may be viewed from the perspective of integrated modules whose relationships and properties are determined by function of the system as a whole. The concept of modularity implies an assemblage of minimal functional units (subsystems) that interact with one another to give rise to emergent properties (higher order properties that arise from the interaction of fundamental processes in the system) of the system. A challenge for systems biology is to define the structure of the system and its modularity in terms of the inherent properties of a self-organizing system: energy, control, and robustness.

Just as the discovery of the cell as the microscopic unit of life changed the way life itself is understood, it is logical to consider the cell as the computational unit of an embryo. Embryos are composed of diverse cells that are each made up of multiple interacting entities or parts (e.g., genes, proteins, and metabolites). Information in biological systems moves both up and down the scale, from molecules phenotype and phenotype molecules (Figure 2). Data-driven abstraction in biosystems also moves up and down in scale, from genes genome and metabolites metabonome. Whether based on high information content or high data content, the cell provides a midlevel focal plane

from which one can move upward or downward in scale. Because newer technologies for live-cell imaging enables the study of single cells in action, model-building can be anchored to the cell as a scale factor and the basic properties of a self-organizing system (energy, control, robustness) can be applied whether the scale is multicellular or subcellular.

Developing embryos can be abstracted as a collection of interacting autonomous modules in which the behavior of each biomodule is controlled by an internal genetic program that can respond to local and systemic signals. We may assume that a system entails groups of cells that continuously interact with one another and with their local environment. Understanding the properties of a system at a multicellular dimension thus requires knowledge of biophysical constraints, such as composition of the extracellular milieu, the chemical nature of signal molecules that are being exchanged between cells, and the kinetics of chemical diffusion and receptor interaction. Similar knowledge is needed for modeling subcellular systems, although scaling requires consideration of signal transduction pathways that modulate gene expression.

Several resources are publicly-available that utilize a universal standard exchange format to represent pathway-based models in general terms. To name a few: Systems Biology Markup Language (SBML; http://www.sbml.org) is well-suited for building programs to model biological processors as a wiring diagram; BioTapestry (http://www.biotapestry.org) meets the informational needs for computational modeling of developmental gene regulatory networks using the SBML workbench; CellDesigner (http://www.CellDesigner.org) creates process diagrams that can be linked to computer-readable SBML files; and CytoScape (http://www.cytoscape.org) can be applied to

visualizing molecular interaction networks and integrating these interactions with gene expression profiles. These resources have broad applications for mapping connections between entities (genes, proteins, metabolites) in a cell to visualize how specific molecular mechanisms are linked with cellular physiology, and to build data-driven mathematical models for the analysis and prediction of druggable-targets or environmental-response pathways.

2.3. Comparative bioinformatics

Constructing analytical and predictive models for birth defects research is constrained by an incomplete understanding of the fundamental parameters underlying embryonic susceptibility, sensitivity, and vulnerability. To understand a developing organ system from a systems biology perspective, key developmental milestones must be parameterized in terms of system structure and dynamics, the relevant control methods, and the overall design logic. This is predicated on the availability of high-content data from studies in developmental biology and toxicology, coupled with the availability of bioinformatics resources to help interpret these data.

Generally, the classification of genes that are up- and downregulated under specific experimental conditions or disease states using microarray-based analysis remains an essential part of this strategy. The availability of public resources for bioinformatics studies, coupled with the acquisition of contextual data from in-house experimentation or national data repositories, has popularized the effort to unravel complex biological networks using a workflow such as in Figure 3.
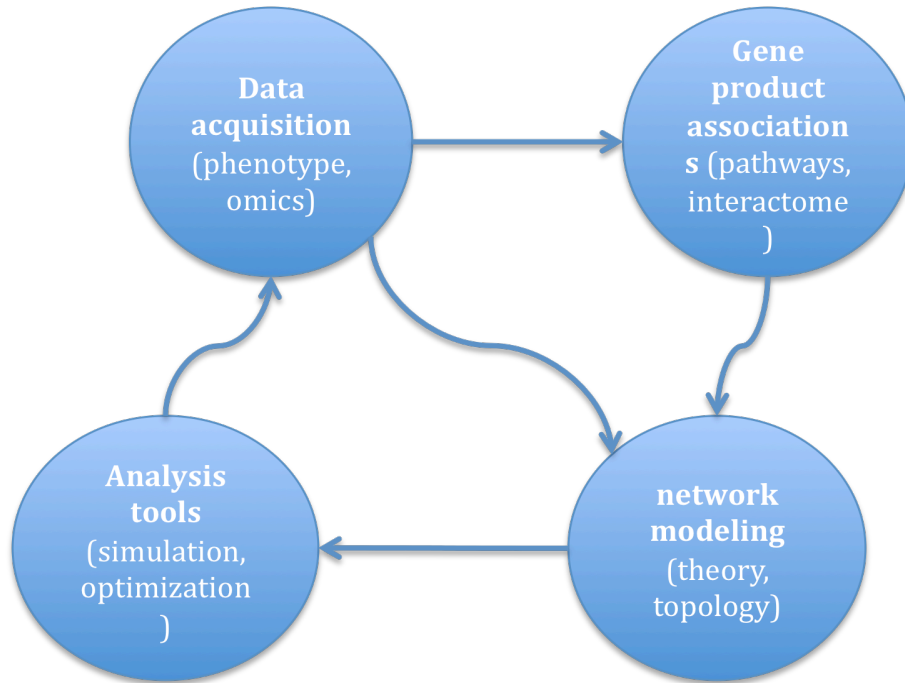
Figure 3. System-level workflow for empirical and computational definition of a network.

Newer technologies in the genomic sciences have facilitated systems biology. A challenge is to assign quantitative values to multiple entailments (genes, proteins, metabolites) under different biological states. Because these technologies are focused on data generation, they emphasize scale at the expense of mechanistic understanding. This fundamental problem requires the integration of gene expression with functional information. Two general classification approaches have been used for this integration: class enrichment and phenotypic enrichment. Class enrichment entails a non a priori stratification of biological themes across a list of genes ordered on the basis of expression difference between biological conditions. This data-driven method usually starts with some type of multivariate statistical method that clusters the genes in each sample by expression profile. An advantage is the ability to discover coarse biological themes without prior knowledge of the system, under the assumption that products of functionally related genes have similar expression profiles and hierarchical relationships. Phenotypic enrichment on the other hand entails an a priori assessment as to whether specified, predefined sets of genes are enriched in a list of genes ordered by expression difference between biological conditions. An advantage here, of course, is the ability to anchor new information with existing (prior) knowledge. In both cases, hierarchical clustering can find simpler developmental trajectories. Most clustering approaches do not account for the serial dependencies of time-course gene expression data. For that reason they may have difficulty determining the optimal number of trajectories that accurately describe time-series data. Several statistical methods have been devised to solve this problem. The use of these methods in conjuction with network-based analysis will allow for identification of cellular-response pathways in the considered dataset of fetal

programming of the murine mammary gland. This approach may have sufficient

resolution to establish the hypothesis that maternal dietary fat balance during pregnancy-

lactation has a lasting effect on mammary gland development in her daughter.

2.4. Modeling
2.4.1. Mathematics Pipeline

Toward the goal of prioritization of chemicals for screening and regulation, the

scientific research and regulatory policy enterprise requires a process model for

developmental toxicity based on a reference connectivity map similar to what was

described by Lamb et al. (2006). As mentioned earlier, part of this effort will depend on

model-building functional networks from qualitative network topologies such as directed

acyclic graphs (DAGs) of interactions based on expression data. Many methods of

adapting qualitative models to learn the continuous-time character of gene regulatory

networks have been proposed, including: discrete models that assume the network entities

can be represented in binary state (on or off); continuous-time models with pairwise and

weighted sum networks based on relationships between entity pairs; and Bayesian or

belief networks that allow model-building with hidden variables and incorporate prior

knowledge.

Discrete models in which the precise outcomes are accurately known have been

successful for modeling simpler genetic networks based on Boolean operators (gene is on

or off). Gastrulation in sea urchin, for example, is built on a computational framework of

50 or so genes affecting modules of cell fate. Boolean models generally assume only two

possible states (expressed, not expressed) for each gene or protein in the network; this

may be difficult to capture with typical microarray data, because discrete state transitions

are more likely to be conditioned on the level of gene activity, or sequence of changes over time, rather than whether a gene is on or off. On the other hand, methods of incorporating the continuous-time character of regulatory networks have been applied to discrete models of the segment polarity gene network in Drosophila.

While many methods have been developed for the qualitative portrayal of gene product-association networks, few methods exist for quantifying the interrelated behavior of genes within these networks. DAGs of interactions based on limited expression data have been used to portray the quantitative interrelationships of genes as Bayesian networks. This approach has advantages to formalizing a network because it requires empirical gene expression data on relatively few samples, it incorporates prior knowledge about entities in the network, and it can address unknown (hidden) variables in the potential interactions. A limitation is in modeling relations among a relatively small number of genes, mainly due to the dimensionality problem of microarray data whereby the number of gene measurements far exceeds the number of samples being measured. This tends to restrict analytical models to represent only a small subset of the measured genes in the network. For example, Bayesian networks were used to model the complex effects of dioxin on human lung epithelial cells using empirical data collected on 12 genes using eight samples of 27 genes to model a segment polarity gene network in Drosophila. Other exemplar models used 28 samples of 65 genes and 14 samples of 113 genes. Large numbers of genes (800) are possible when the goal is to identify regulatory connections rather than to specify network topology. As pointed out in Huang et al. (2007), the poor scalability of existing techniques typically reveals only an incomplete structure of a gene regulatory network and these authors proposed the use of scalable

gene regulatory network learning algorithms based on Bayesian networks and association rules.

Bayesian networks can be implemented in the R package freely available from the R Development Core Team (2006). A Bayesian network is a DAG of nodes representing genes with variable attributes (expression value) and arcs representing gene-dependency relations. Joint distribution is represented by the variable parameters $X(1)$, , $X(n)$ and parents (A) of node A. The joint distribution for $X(1)$ through $X(n)$, for example, is represented as the product of the probability distributions for $i = 1$ to n. If X has no parents, then its probability distribution is unconditional, otherwise it is conditional. The extension to a stochastic process is often employed where we consider the joint distribution of $X_t(1)$, , $X_t(n)$ and where t is the time variable. Such a dynamic Bayesian network is appropriate for time-series microarray data. To structure the network, it is necessary to specify, for each node, the probability distribution of X conditional on its parents. Conditional distributions depend on unknown or missing parameters; therefore, iterative fitting techniques such as the expectation-maximization (EM) algorithm may be used for network optimization. A Markov Chain Monte Carlo (MCMC) algorithm can be applied to compute posterior distribution, giving a statistic for prediction error. Analysis of sensitivity-specificity employs parameters specific to each treatment group to run computer simulations that are conditional on actual experimental data (Figure 4).
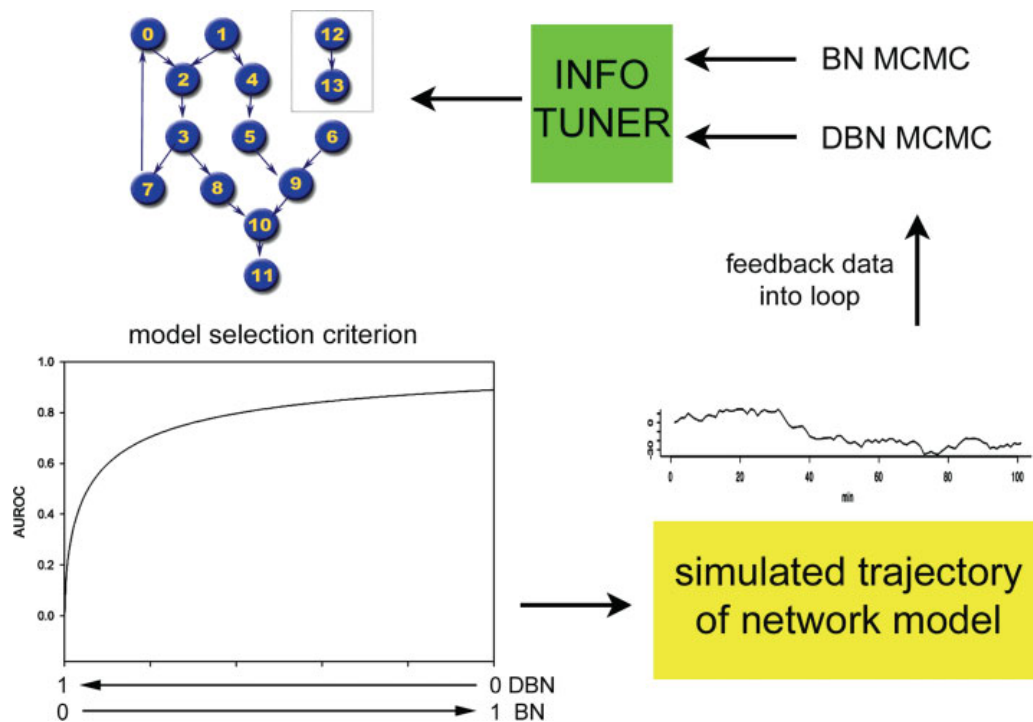
Figure 4. Schema of Bayesian network (BN) and dynamic Bayesian network (DBN) kernels for gene network learning.

(modeled after E. Paquet, University of Montreal, http://www.iro.umontreal.ca/lisa/workshop2004/slides/paquet_slides.ppt). Time-series (sequential) and dichotomous (nonsequential) data are run through MCMC algorithm. Information from the two types of data are coupled by Info tuner to construct a filtered DAG of nodes representing genes and arrows representing dependence relations among them. An optimal DAG topology is selected from sensitivity-specificity of the scoring criterion, such as area under the receiver-operator curve (AUROC). The network trajectories are then simulated for each of the vertices and the results are fed into the data loop for optimization.

Model refinement using Info tuner couples the simulated results with empirical data to give significant improvement in both speed and accuracy of the learning algorithm by replacing the greedy search with novel implementation of the Metropolis-Hastings-type random walk. Computational models must be validated against empirical data. Bayesian techniques are particularly suited to treating some of the major difficulties associated with validation: quantifying multiple sources of error and uncertainty in reaction models; combining multiple sources of information; and updating validation assessments as new information is acquired. Several standard methods of sampling from the probability distribution are available based on a variety of MCMC-type schemes. Gamma and beta priors for the hypothesized system parameters can give posterior assessment of the model predictive power and assess how well the model performs. With that information the model deficiencies could be addressed and the initial model refined.

2.4.2. Critical decisions of a bionetwork

Given that most high-content genomics data is collected on blocks of cells or tissues, it is important to think about how to anchor Bayesian networks to specific cellular processes that operate jointly across time and in different parts of the embryo. Using experimental manipulations such as small interfering RNAs (RNAi), genetic knockouts, mutations, and drugs or chemicals, one can begin to explore the quantitative response to specific changes in the network entities (nodes) and relations (arcs). One approach is to assume that the fundamental molecular interactions have both deterministic and stochastic components, then simulate networks using a standard stochastic simulation algorithm and analyze complex behaviors in the perturbed system. Because disrupting gene function (e.g., knockouts, knockdowns, mutations, and altered

expression) can perturb embryogenesis, and because various teratogens may reprogram

developmental networks, it is useful to begin to run different simulations on the Bayesian

networks to predict which genes and genetic dependencies can have the greatest impact

on the flow of molecular regulatory information through the network. Thus, once a gene

network structure has been predicted by the bioinformatics and mathematical pipelines

described above, a pathway can be represented as a weighted, directed graph (Young,

Nolte et al. 2008) with a set of vertices, V, representing the genes in the case of a gene

expression network, and a set of edges, E, representing their formal interactions. The

network can then be portrayed with the weights representing the level of importance of

each interaction as measured through the correlation of gene expression values, along

with additional supporting data that may be obtained from external sources such as

transcription factor databases, genotype-phenotype databases and so forth. Figure 5

represents such a gene interaction network for the 14 genes (labeled 0-13) determined

using a Bayesian network approach (see Figure 4). In this simplified example, the

modeled relations indicate that genes 12 and 13 interact with each other, but not with any

of the remaining genes in the network. Given a directed, weighted graph representation, a

number of properties can be determined regarding the interactions. Using graph theory, it

is possible to determine all possible paths beginning at one node and ending at another. In

Figure 5A, two possible paths can link nodes 1 and 11: 123 81011 and 145910 11, based

on current knowledge and empirical data. Since each edge is directed and weighted, the

minimum spanning tree can be determined in more complex schema using Kruskal's

algorithm. This returns the fewest number of interactions needed to have a functional

path.
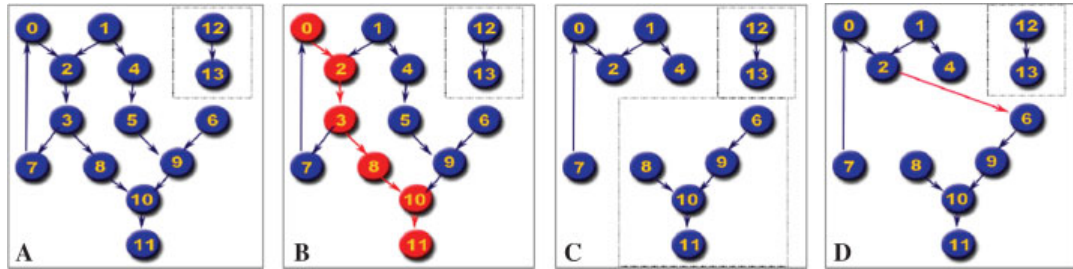
Figure 5. Computational perturbation of modeled gene product association networks.

A: Network model constructed from the Bayesian effort illustrated in Figure 4, with a genetic dependency flow for genes 0 to 11. **B**: Shortest path of information flow between nodes 0 11 shown in red. **C**: Model for the removal (e.g., genetic knockout, miRNA knockdown) of two nodes, in this case nodes 3 and 5, leaving three disjoint gaps in the pathway. **D**: Emergence of a new connection between nodes 2 and 6 (red arrow), enabling a new information path from nodes 0 11 and 1 11.

In addition, the shortest cost and minimum cost paths can be calculated according to either the number of genes involved in the path, or the weights of the interactions between the genes. Figure 5B shows a shortest path between genes 0 and 11 in red. Each of the individual paths in the network potentially represents how the flow of information might react to physiological signals or perturbations related to development. In the case of fetal alcohol syndrome, for example, we may envisage qualitative network topologies that identify sensitive genes and their potential connections, then limit the networks to those interactions with the strongest weight of evidence for genetic dependencies, construct Bayesian models to arrange the nodes and edges in the pathway (Figure 4), and finally impute how a teratogen might disrupt the flow of information across the network to alter downstream processes. Another step in examining such a graph might be to determine the set of vertices (V) and edges (E) most vital to network integrity. In other words, which nodes, when removed, result in a disjoint set that will not allow for traversal from the first node in the pathway (source) to the final product (sink)? Figure 5C shows what might happen when two critical nodes are removed, while Figure 5D indicates a single relation that can potentially restore the interaction network. The resulting effect on the network as a whole can be predicted by traversing the pathway from input to the output using a shortest cost path approach as with Dijkstra's algorithm, in which the cost of the path is inversely proportional to the weighted edge. Representing the pathway as an adjacency matrix of a graph as such can enable each of these steps to be incorporated and run in a relatively short period of time for a thorough analysis of individual pathways and to eventually predict outcomes computationally.

2.4.3. Toward a predictive understanding of developmental toxicity

Multistudy expression profiles have been run across systems of the mouse embryo during normal development and following teratogen exposure. Although those studies were limited to microarray gene expression data, an integrative analysis of the mouse embryonic transcriptome revealed hidden relationships that might not be discovered without bioinformatics resources. This motivates a continuation of research efforts to model developmental toxicity at the level of the embryonic transcriptome. Two important goals are to find a minimal set of genes with adequate discriminating power to categorize the samples by experimental parameter, and to design a prediction system using the selected genes to accurately classify unknown samples. Figure 6 presents a conceptual framework for intelligent model selection in comparative toxicogenomics of birth defects and developmental disease.
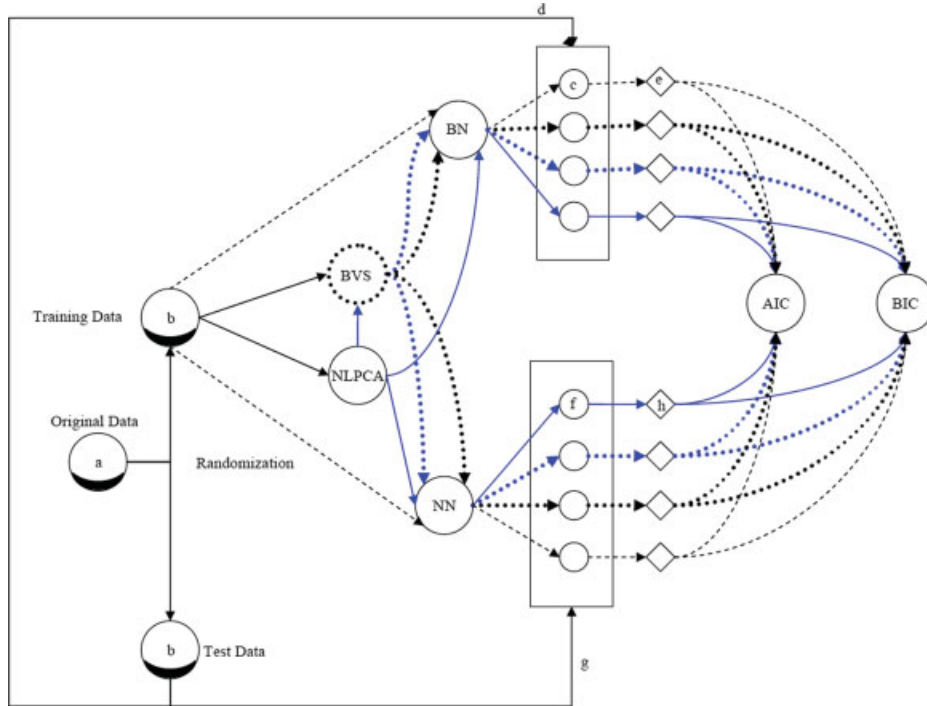
Figure 6. Conceptual framework for intelligent model selection in comparative toxicogenomics of birth defects and developmental disease.

a: The original microarray data representing the quantitative descriptions of gene expression levels given a specific developmental stressor. **b:** The original data are randomly divided into Training Data and Test data; each data set has similar sample numbers. Blue lines represent cleansed data processed through nonlinear principal components analysis PCA (NLPCA); black lines refer to the uncleansed data. The various paths these data can take are shown by the arrows, e.g.,: Bayesian variable selection (BVS) network analysis through Bayesian networks (BN) or neural networks (NN). **c:** The corresponding models from each path are depicted as a manifold in which each circle represents a DAG containing the stochastic dependencies of the genes on one another and on the phenotype as quantified by probability density functions. The BN manifold defines an order of network models having a class of networks linked to the different input conditions (e.g., different sets of significant genes). **d:** Test data used as input into the DAGs representing the BN models of (c). **e:** For an unknown sample, the probability of developing the phenotype of interest is represented by the diamonds and is calculated from the input Test data sets; this probability represents output of (d) when input with test data set. It allows calculation of the true negative and true positive rates by comparison with the actual phenotypic expression/nonexpression of the Test data set. **f:** The NN manifold defines an order of network models having a class of networks linked to the different input conditions; each circle is a NN model consisting of nonprobabilistic nodes (neurons), contrasted to the probabilistic nature of the models in (c), representing nonlinear regression. This establishes a nonprobabilistic mathematical relationship of

23

gene interdependencies and gene dependency on the phenotype. **g:** Test data set used as input into the NN models of (f). **h:** Prediction of phenotypic expression or nonexpression based on input Test data set: allows calculation of the true negative and true positive rates by comparison with the actual phenotypic expression/nonexpression of the Test data set. AIC and BIC: because multiple data sets may be run for each order of network models, there will be many model classes to compare. Each class of models can be compared within order (e.g., BN-BN, NN-NN) as well as between order (e.g., BN-NN, NN-BN) and the consensus species of network models will be used for further validation. Akaike information criterion (AIC) and Bayesian information criterion (BIC) assist in selection of the optimal species by essentially scoring the parsimony of the data and finding the optimal number of variables weighed against the model's error.

Predictive modeling in developmental toxicology starting from microarray gene expression data will require a series of steps, from denoising algorithms to optimal variable selection and model comparison. The starting data are assumed to have inherent structure that is related to the biological condition of the test samples and that is expressed by many variables (e.g., high-dimensional), some having significance to the phenotype and others not contributing significantly. Many analysis methods of unsupervised clustering have been described to find the more compact description of the data (e.g., low-dimensional) and will not be described here. The problem here, however, is to define an underlying structure in the regulation of the embryonic transcriptome where the standard learning task identifies gene networks significantly anchored to the risk of developmental phenotypes.

Nonlinear principal component analysis (NLPCA) is powerful method of data cleansing that is conceptually a nonlinear extension of the popular principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its Cartesian coordinates along each axis. The nonlinear generalization of PCA has been described to transform high-dimensional data into a low dimensional code (encoder network) and to recover the structured data (decoder network) under the constraint that initial weights of the data should be close to a solution. NLPCA is accomplished using an autoassociative neural network, which is a neural network that has exactly as many output neurons as input neurons, and is trained (an optimization routine) so that its outputs match its inputs for all data presented. The concept in Figure 6 contains a modified feed-forward multilayer layer perceptron neural network, whereby input and output layers have linear activation functions to allow recovery of the input

space; three hidden layers with nonlinear activation functions (sigmoid, tansigmoid, etc.) give the network nonlinear mapping capability. The middle hidden layer functions as a bottleneck because it contains fewer neurons than the two other hidden layers or input and output layers. This limits the autoassociation and thus removes information that does not contribute to the feature space (e.g., risk for a particular phenotype). Such networks are very useful in cleansing the data because the network can, knowing what gene interaction signatures look like, separate the gene pathway data from statistically insignificant information in a nonlinear way. The output of NLPCA is cleansed or denoised data, which is then used as input into optimal variable selection techniques based on Bayesian principles.

Optimal variable selection using the Bayesian paradigm is another powerful tool to lower dimensionality to focus on entities (variables) that are significant contributors to the group structure of the data. Two inherent assumptions are as follows: 1) that an optimal subset of genes can be found (solution) that can account for a particular phenotype, such as fetal alcohol syndrome; and 2) that most genes do not contribute significantly to the syndrome. The validity of these assumptions must be confirmed although they appear to be reasonable given microarray data for experimental fetal alcohol syndrome in mice. (Tadesse, Sha et al. 2005) developed a data-driven method that selects discriminating variables into G groups, where G is unknown. This clustering approach is a multivariate mixture model with an unknown number of components and uses a binary latent vector that effectively searches through all subsets using the Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm. The reversible jump aspect of this particular MCMC simply allows movement around different dimensional

26

subspaces. Furthermore, this method can predict the class membership of future observations. The general approach for estimating the parameters of interest uses Metropolis moves and the RJMCMC, as follows: 1) update parameters from their full conditions (Gibbs sampler); 2) split one mixture component into two, or merge two into one (split/merge moves); and 3) birth or death of an empty component (birth/death moves). The birth/death moves specifically deal with creating/deleting empty components and do not involve reallocation of the observations (e.g., the entire data space does not need to be recomputed), hence avoiding doubling-back to unproductive models and reaching a faster convergence to a posterior. Optimal variables are identified by selecting those with the largest marginal posterior density. This Bayesian variable selection (BVS) method is complicated but has been utilized successfully in benchmark plant (iris) data and DNA microarray data from endometrial cancer patients (Tadesse, Sha et al. 2005).

To compare different models for critical gene solutions, a manifold of Bayesian networks (BN) and of neural networks (NN) are computed from the sample data using the paradigm suggested in Figure 6. The BN approach is described above; within the field of NN, there are a variety of network architectures built in MATLAB NN toolbox software that include feed-forward multilayer perceptron networks (The Mathworks, Inc., Natick, MA). An optimal network of genes associated with a particular risk for malformation, defined as a species, might be theorized at the convergence of different class models onto a minimal set of entities (genes) having maximized interactions (relations). In this case, a tradeoff of complexity in a model versus how well a model fits the data can be assessed using the Akaike information criterion (AIC) and Bayesian information criterion (BIC)

that reward goodness of fit and penalize parameters. These statistical methods are rooted in information theory and are indicative of the parsimony of the final model, where a decrease in value corresponds to an improved model.

3. METHODS

3.1. Experimental (animal) studies at RIVM, the Netherlands

Wild type FVB/NHanHsd mice, six to eight-weeks old, were obtained from Harlan (Horst, NL). All mice were housed in the animal facility in a climate-controlled room with a 12 h on/off light cycle. Tap water and diets as described below were provided *ad libitum*. Animals were monitored daily for general health. The animal protocol used in this research was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Louisville. The research was also approved by the Animal Ethics Committee (IACUC) of the Dutch National Vaccine Institute (NVI) at Bilthoven, which also serves as the IACUC of the National Institute for Public Health and the Environment (RIVM) where the animal work was performed for this study.

All diets used in this experiment were obtained from AB Diets (Woerden, NL). Mice in the control group were fed a low-fat (5%) regular diet. The high-fat diets contained 24% fat at the expense of carbohydrates and were based on corn oil (high ratio of n-6/n-3 PUFAs) or flaxseed oil (low ratio of n-6/n-3 PUFAs) (Chempri Oleochemicals, Raamsdonksveer, NL).

The study presented here consisted of two dietary exposure groups and one control group, with differing proportions of polyunsaturated fatty acids (PUFAs). Each dietary group consisted of 6 females and 3 males. After one week of acclimation to the

animal room and two weeks prior to mating the dams and sires were shifted from standard mouse chow to one of the treatment diets. Females were allowed to breed during one week with the males. Subsequently, dams were housed individually and maintained on their respective diets through birth and until weaning of their litters. The number of pups born alive was recorded at the day of parturition, designated postnatal day (PND) 0. Body weights of all pups were recorded on PND 1, PND 7, and PND 14. After weaning on PND 21, male pups were euthanized by carbon dioxide asphyxiation. F1 female pups were kept on the appropriate treatment diets until 6 weeks of age. At 6 weeks the F1 female mice were shifted to the standard low-fat (5%) chow until terminal sacrifice at 10 weeks of age. Body weights of F1 female pups ($n = 21$ for the control group, $n = 12$ for the high-fat corn oil group, and $n = 15$ for the high-fat flaxseed oil group) were recorded weekly from PND 21 until 10 weeks of age. At autopsy, blood and major organs were isolated from each animal, including the fourth abdominal mammary gland on the right side of the mouse, liver, kidney and spleen. The mammary gland was kept in RNA*later* RNA stabilization Solution (Ambion, Austin, Texas) at 4 °C for maximally 14 days; other organs were snap frozen in liquid nitrogen. A schematic overview of the study design is depicted in Figure 7.
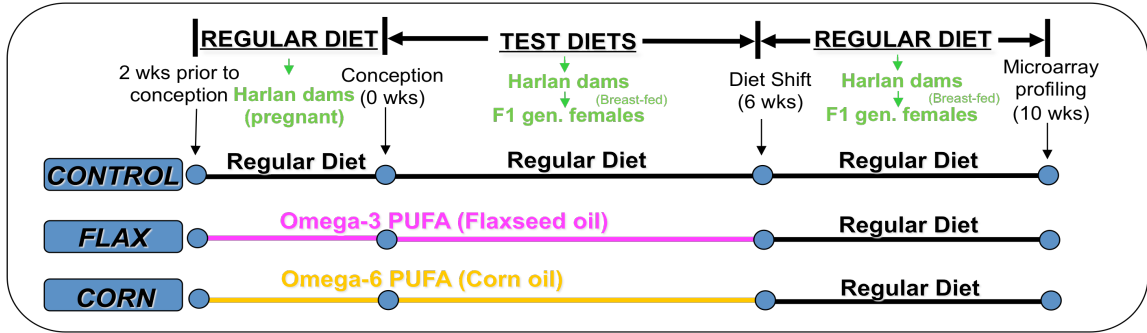
Figure 7. Study design

3.2. RNA isolation and microarray analysis

DNA/RNA was extracted by using AllPrep DNA/RNA mini isolation kit (Qiagen, Valencia, CA, USA). RNA samples were treated with the RNase-Free DNase set (Qiagen) and concentrations were measured using a NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). RNA was assessed for quality with the Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Both the RNA integrity number and the presence of degradation products were checked.

The mouse oligonucleotide libraries (catalog no. MOULIBST and MOULIB384B) were obtained from Sigma-Compugen, Inc. Technical support was supplied by LabOnWeb (http://www.labonweb.com/cgi-bin/chips/full_loader.cgi). The libraries represent in total 21,766 LEADS™ clusters plus 231 controls. The oligonucelotide library was printed with a Lucidea Spotter (Amersham Pharmacia Biosciences, Piscataway, NJ, USA) on commercial UltraGAPS slides (amino-silane-coated slides, Corning 40017) and processed according to the manufacturer's instructions. The slides contained 65-mer oligonucleotides, and the batch was checked for the quality of spotting by hybridizing with SpotCheck Cy3-labeled nonamers (Genetix, New Milton, Hampshire, UK).

Total RNA samples were hybridized in randomized batches, according to a common reference design without dye swap. An RNA pool of all samples isolated was used as a common reference. From the total RNA samples with an RNA integrity number value of >7, 1.0 μg was amplified using the Amino Allyl MessageAmp amplified RNA kit (Ambion) and labeled with Cy5 (experimental samples) and Cy3 (common reference)

reactive dye according to the manufacturer's instructions. The microarrays were hybridized overnight with 200 μl hybridization mixture, consisting of 50 μl Cy3- and Cy5-labeled amplified RNA (with 150 pmol Cy3 and 150 pmol Cy5), 100 μl formamide, and 50 μl 4× RPK0325 microarray hybridization buffer (Amersham Pharmacia Biosciences), at 37°C, washed in an automated slide processor (Amersham Pharmacia Biosciences), and subsequently scanned at two wavelengths using a ScanArray 4000XL microarray scanner (Perkin-Elmer, Waltham, MA, USA). Median Cy3 and Cy5 signal intensities per spot were determined using Array Vision software (Imaging Research, St. Catharines, Ontario, Canada).

In total, 28 two-color mouse genome (*M. musculus*) cDNA microarrays were used in this study, representing 4 batch groups: Groups A and B ($n = 6$ each) were from the 5% fat diet (control) groups, and Groups C and D ($n = 8$ each) were from the treatment groups. The microarray processing was done on 4 different dates, and all groups were processed on each of these days; therefore, treatment is not nested within batch. For purposes of the analysis, the two control groups were combined ($n = 12$).

3.3. Pre-processing of microarray dataset

Microarray quality control was performed on raw data by means of visual inspection of the scanned images, as well as a check on the scatter and MA plots. Image plots, box plots, MVA plots, and histograms were used to assess data and microarray processing quality. The image plots show that several arrays had large areas with cold spots during some of the runs. One of the arrays (A3) had close to 30% of its area taken up by cold spots. There is a potential that these spots had an adverse affect upon the analysis. Normalization methods cannot be relied upon to eliminate the effect of large

cold spots. The BioConductor software packages arrayQuality and marray were used to assess array quality. An offset of 50 was used for background correction to avoid negative or zero corrected intensities and to preserve variance for analysis of differential expression. Locally weighted regression was used to normalize within arrays. Finally, quantile normalization was performed across arrays because of the large variation between arrays seen in the box plots (Figure 8).

Boxplots Before Background and
Normalization
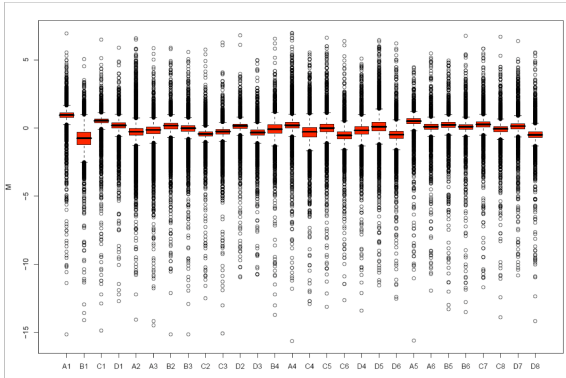
Boxplot after Normalization
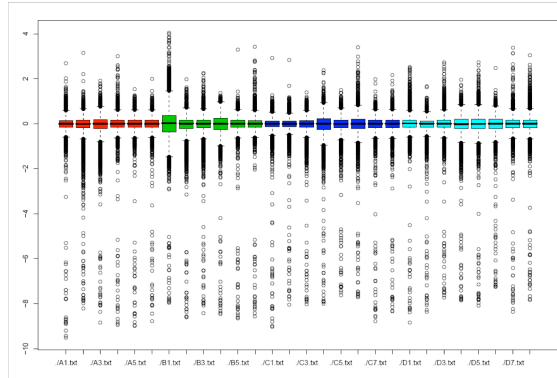
Figure 8. Boxplots before and after normalization

3.4. Microarray data analysis

Several data analysis models were used to identify genes that were differentially expressed between the treatment groups and the control group.

3.4.1. Method 1 – analysis of variance, hierarchical clustering

In the first model, two-way analysis of variance (ANOVA) was employed that takes into account the batch effect (date of microarray processing) as well as the treatment effect. Treatment effect was included last in the model to account for the unbalanced design. To correct for multiple testing, Benjamini-Hochberg's false discovery rate (FDR) (Benjamini and Hochberg 1995) was used with a cut-off of 0.05. The F1 statistics were used for gene set enrichment analysis (GSEA) (Mootha, Lindgren et al. 2003). All pathways present in the c4 database of the Molecular Signatures Database (MSigDB 2.0; http://www.broad.mit.edu/gsea/msigdb/msigdb_index.html) were tested for significance.

Hierarchical clustering was employed using Euclidian distance as a dissimilarity measure. In addition, the DAVID 'Functional Annotation Clustering' tool (Dennis, Sherman et al. 2003) was used to cluster the 99 genes into different groups (more than 15). The method allows genes to belong to more than one group.

3.4.2. Method 2 – principle component analysis, analysis of variance

In the second model, ratiometric values (test/reference) were transformed to $\log_2$ and normalized with Lowess smoothing. Batch and litter-size effects were removed using Partek Genomics Suite v 6.3 (Partek Inc.; www.partek.com) and the residual variance

was analyzed by Principal Components Analysis (PCA), one-way ANOVA ($P \leq 0.005$) and specific group comparisons. All data was mean-centered and quantile normalized to make the gene expression distributions across the different microarrays similar. This analysis returned 670 genes for subsequent cluster analysis and functional annotation.

3.4.3. Method 3 – Bayesian variable selection

In the third model, a Bayesian-based variable selection (BVS) algorithm was applied. A systematic methodology for identifying optimal discriminative gene subsets in the analysis of this data (controls and flax) is adopted via a Bayesian-based variable selection algorithm (BVS) utilizing the stochastic search technique. BVS analysis was implemented in MATLAB (The MathWorks, Inc. 3 Applehill Drive, Natick, MA 01760-2098) using code available from Marina Vannucci (Bayesian variable selection in multinomial probit models 2004 - http://www.stat.rice.edu/~marina/codes.html) (Sha, Vannucci et al. 2004), additional coding written in MATLAB. The MATLAB Bioinformatics Toolbox 2.5 was utilized in generating heatmaps and gene ontology enrichment.

Vannucci (P. J. Brown 1998) implemented a BVS methodology that builds upon the stochastic search variable selection technique (SSVS). The idea of SSVS is to embed a linear regression model into a hierarchical Bayes model, where regression coefficients are given mixture priors. Vannucci's code extended the model to the general multivariate case for probits and multinomial classification (Sha, Vannucci et al. 2004). This is accomplished through the use of latent (unobserved) variables representing the categorical response. Introduction of latent data (data augmentation) into the model allows the probit model to be transformed into a linear regression problem in the

multivariate normal context. Essentially, the entire regression model is embedded into a

hierarchical Bayes normal mixture model, where latent variables are used for subset

selection. A key idea for variable selection in this context is introduction of a binary p-

vector γ that indexes all possible variable subsets. More specifically, the jth element of

the latent binary p-vector γ ,γ_j , is 0 or 1, corresponds to inclusion or exclusion of a

particular variable in the gene subset. This vector takes 2^p values. Elaborating on the

priors of the regression coefficients by assuming the coefficients come from a normal

mixture facilitates assessment of the importance of a variable in the model. The prior

distributions of γj's are set as independent Bernoulli (collectively Binomial).

Through appropriate prior elicitation and integration of parameters out of the joint

posterior, the posterior distribution of all possible subsets given the data can be derived

but is of unknown form. This posterior space is searched via Metropolis-Hastings using

add/delete/swap Metropolis moves. More specifically, a fast-inference scheme is adopted

by deriving the marginal posterior distribution of the latent vector γ given the data by

integrating out other parameters from the joint posterior distribution (Sha, Vannucci et al.

2004). Again, the distribution of γ is not of known form and is explored using Markov

Chain Monte Carlo (MCMC), where the Markov Chain is constructed using the

Metropolis-Hastings algorithm. The composition of the variable subset (i.e. genes)

evolves as we iterate through in the MCMC context by searching the 2^p space using

add/delete/swap Metropolis moves (add/delete: randomly choose one of the p indices in

γ_{old} and change its value from 0 to 1 (add), or from 1 to 0 (delete), to become γ_{new}; swap:

independently choose and at random a 0 and a 1 in γ_{old}, switch their values to get γ_{new}).

Subsets and variables with a higher joint posterior probability can be identified by more

frequent occurrence in the MCMC output. As a result, posterior inference of γ, and thus

the composition of the variable subsets, is completely data-based after hyperparameter

specification. Prediction performance of these gene subsets is assessed using a sampling-

based cross-validation prediction procedure due to the small number of microarrays in

relation to the large number of variables.

A key attractive feature of this technique is that it assesses the joint effects of the

variables. Two variables independently may not be descriptive given the observed data,

but in conjunction, are valuable predictors. Because variable subsets are selected on the

basis of maximal joint posterior density, the subsets will inherently attempt to avoid

correlation of included variables, as inclusion of an additional highly-correlated variable

(at the expense of a potentially valuable predictor) into a model brings no additional

predictive power, leading to no increase in joint posterior probability. Data for this

technique must be pre-processed by mean-centering and rescaling by variable range.

Advantages of this implementation are that the subset selections assess the joint effect of

variables in attempting to account for the discriminative changes seen in the response

data and that this procedure truly determines how "valuable" a variable is. Disadvantages

are computational demands due to the vast and highly complex posterior space (defined

on all possible subsets, 2^p), where it may not be possible to efficient search over the space

in extremely large datasets. Also, there is risk that the MCMC procedure does not

converge to the posterior distribution within a reasonable timeframe. This can be

minimized by running multiple chains for long periods, discarding all but the last

iterations, and pooling the MCMC. Prior specification is also an issue, although the

authors give procedures for calibration of the priors from the data and from theoretical

constraints (avoiding Lindley's Paradox). The SSVS schema is shown in Figure 9.

Variable selection is achieved through introduction of *latent binary vector*
• jth element is either 1 or 0
    • 1 = inclusion of jth variable in the model
    • 0 = exclusion of jth variable in the model

γ

Prior

$P[\alpha, B, \Sigma]$

Decompose Prior

$P[\alpha, B \mid \Sigma] * P[\Sigma]$

Conditional Independence

$P[\alpha \mid \Sigma] * P[B \mid \Sigma] * P[\Sigma]$

Elaborate on Regression coefficients

$P[\alpha \mid \Sigma] * P[B, \gamma \mid \Sigma] * P[\Sigma]$

Hierarchical Bayes

$\mathbf{Y}_i = \alpha' + \mathbf{X}_i'B + \varepsilon_i, \quad \varepsilon_i \sim N(0, \mathbf{\Sigma}), \quad i = 1, \ldots, n$

$\mathbf{X}$ = Microarray Data

"First-pass" ANOVA
Mean Center
Quantile Normalize

$z_i = \begin{cases} 0 & \text{if } y_i \leq 0 \\ 1 & \text{otherwise.} \end{cases}$

$$\iiint P[X, Y, Z \mid \alpha, B, \Sigma, \gamma] * P[\alpha \mid \Sigma] * P[B \mid \Sigma, \gamma] * P[\gamma] * P[\Sigma] \, d\alpha \, dB \, d\Sigma$$

Likelihood   Prior

Markov Chain Monte Carlo
• 500,000 iterations
•100,000 burn-in

Metropolis-Hastings Algorithm
Explore space using **Metropolis Moves**
Add/Delete
Swap

Posterior $P[\gamma \mid X, Y, Z]$ (set of all subsets given the data)

*Hyperparameter specification
•Model is sensitive to specification
•Can calibrate hyperparameters from data

Top Genes by Marginal Probability or Joint Posterior Probability

Optimal Gene Subsets

Figure 9. Layout of stochastic search variable selection technique

Hyperparameters controlling model shrinkage, prior probabilities of variable inclusion, and model expectation size were all specified. The parameter related to total relative prior to posterior precision was taken as the average of the values corresponding to 0.1 and 0.005. This specific hyperparameter regulates the amount of shrinkage in the model and is related to the ridge regression. Prior probabilities of variable inclusion are assigned independent Bernoulli trials where p is assigned ½. Because it is expected and desired to have few genes in the gene subset, the binomial prior for the chosen variables (genes) in γ is set to have an expectation of 10, so that the MCMC procedure tends toward the space of smaller models. MCMC was run for 500,000 iterations, where the first 100,000 were discarded as burn-in. The starting γ vector was taken with 10 randomly selected genes included.

Vannucci et al. have extended variable selection to incorporation in clustering high-dimensional data. Future work may be done to explore the cluster structure and associated significant variables (Sinae, Mahlet et al. 2006) (Tadesse, Sha et al. 2005).

3.5. Fatty acid profiling of exposure diets and mouse sera (RIVM)

Fatty acid composition was analyzed both in the various animal diets and in a representative number of sera obtained from mice fed these diets. Methods were adapted from procedures described earlier (Mamalakis, Jansen et al. 2006). Fatty acids have been expressed as percent of the total fatty acids present in the chromatogram.

3.6. Statistical analyses

General statistical procedures were performed with either SPSS version 12.0.1 or S-PLUS 2000 statistical software packages. All values were expressed as means ± standard deviations when appropriate. Reproductive parameters and pup weights of PND 1 were analyzed using one-way ANOVA. Body weights of female pups from 3 to 10 weeks of age were analyzed using a nonlinear mixed effects model. The nonlinearity is formed by an $y_{ij}(t) = Asym_i + (R0_i - Asym_i) \cdot \exp[-\exp(lrc_i) \cdot t_j] + \varepsilon_{ij}$ asymptotic growth model, in which the weight y_{ij} of mouse i at time t_j is modeled as:

4. Results

4.1. Phenotype anchors (RIVM)

Maternal dietary exposure to a high-fat diet, rich in either n-6 or n-3 PUFAs, did not result in significant effects on pregnancy rate, litter size or sex ratio (data not shown). The number of female pups born in each dietary group was $n = 21$ for the control group, $n = 12$ for the high-fat corn oil group, and $n = 15$ for the high-fat flaxseed oil group.

4.2. Fatty acid profiling

To investigate whether perinatal exposure to treatment diets high in n-6 or n-3 PUFAs was still detectable at time of sacrifice (i.e. 4 weeks after cessation of treatment), fatty acid profiling of a representative number of mouse sera was performed. The n-6/n-3 PUFA ratio is about threefold higher in mice fed a diet based on corn oil as compared with mice fed a diet based on flaxseed oil. These results indicate that perinatal exposure to high-fat diets with either high or low n-6/n-3 PUFA ratios has a lasting effect on the fatty acid profile of the serum.

4.3. Postnatal body weight gain (RIVM)

Postnatal body weight trajectories are shown in Figure 10. At PND1, no statistically significant difference in body weight of the pups between any of the groups was observed. However, at 3 weeks of age, female pups from the high-fat diet groups, based

44

on either corn or flaxseed oil, were significantly heavier than female pups from the low-fat diet ($P < 0.001$). In addition, body weights of F1 female pups from the high-fat diet groups increased more rapidly than from the control group ($P < 0.001$). At 10 weeks of age, body weights of the F1 female pups perinatally fed a high-fat diet based on corn oil were no longer significantly different from the control group. In contrast, female pups perinatally fed a high-fat diet based on flaxseed oil were still heavier than pups from the control group ($P < 0.001$), despite the fact that they were shifted to a low-fat diet from 6 weeks of age onwards. Postnatal body weight trajectories of F1 females sustained on the high-fat diet based on flaxseed oil paralleled the normal growth curve, but these test animals remained heavier by a constant scalar. In contrast, the growth curve of females sustained on the high-fat diet based on corn oil approached the regular trajectory. The slight dip in trajectories at 3 weeks of age is accounted for by the removal of male pups, which are slightly heavier than females.
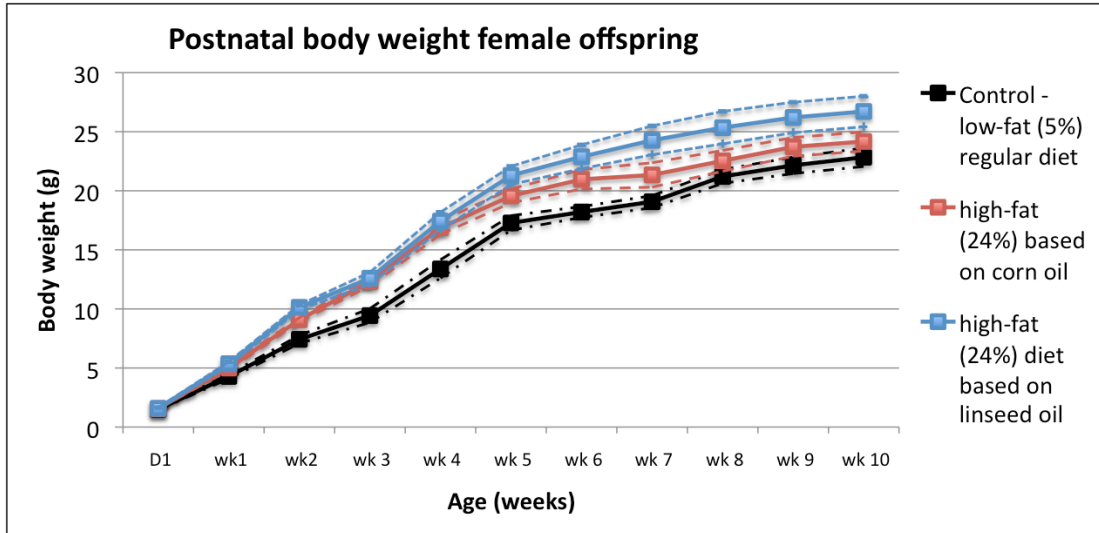
Figure 10. Developmental programming of growth trajectories in female FVBn mice

Test diets were given maternally, from preconception through lactation (3-weeks), and after weaning to early juvenile life (6-weeks of age). The increased weights of high-fat diets were significant versus regular diet at 3-weeks and beyond; however, females from the fatty corn oil diet shifted to the regular trajectory between 6- and 10-weeks of age.

4.4. Mammary gland gene expression

Gene expression profiling of the 4th right abdominal mammary gland was determined on postnatal week 10. We used two different data analysis models. In the first model, statistical analysis using a two-way ANOVA and false discovery rate cut-off at $P=0.05$ returned 99 differentially expressed genes across the dietary groups. Pathway inference performed using the R package GSA (Gene Set Analysis) focused on a gene set from Molecular Signature Databases (MSigDB) defined by expression neighborhoods centered on 380 human cancer-associated genes, of which there were 28 gene sets with fdr less than or equal to 0.10. Functional annotation chart from DAVID showed weak representation of only one KEGG pathway at that same level (natural killer cell mediated cytotoxicity).
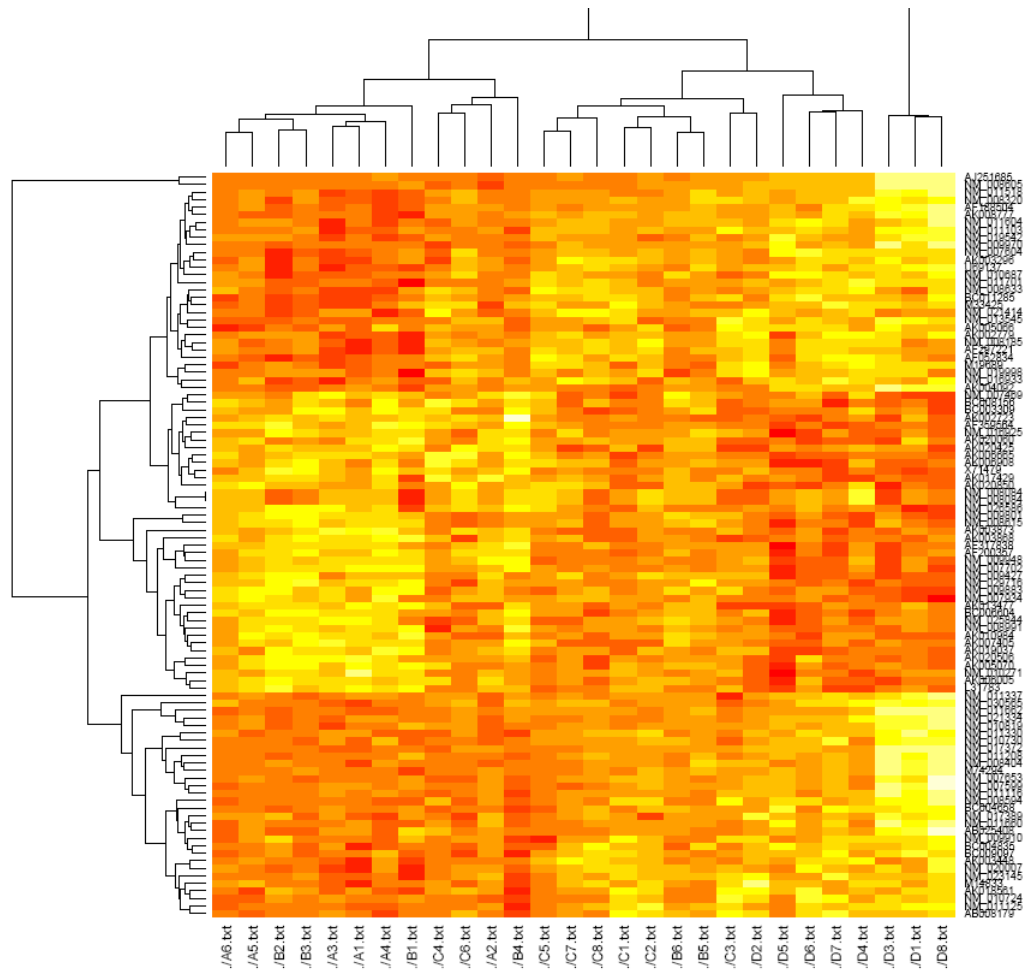
Figure 11. Heatmap of 99 differentially expressed genes
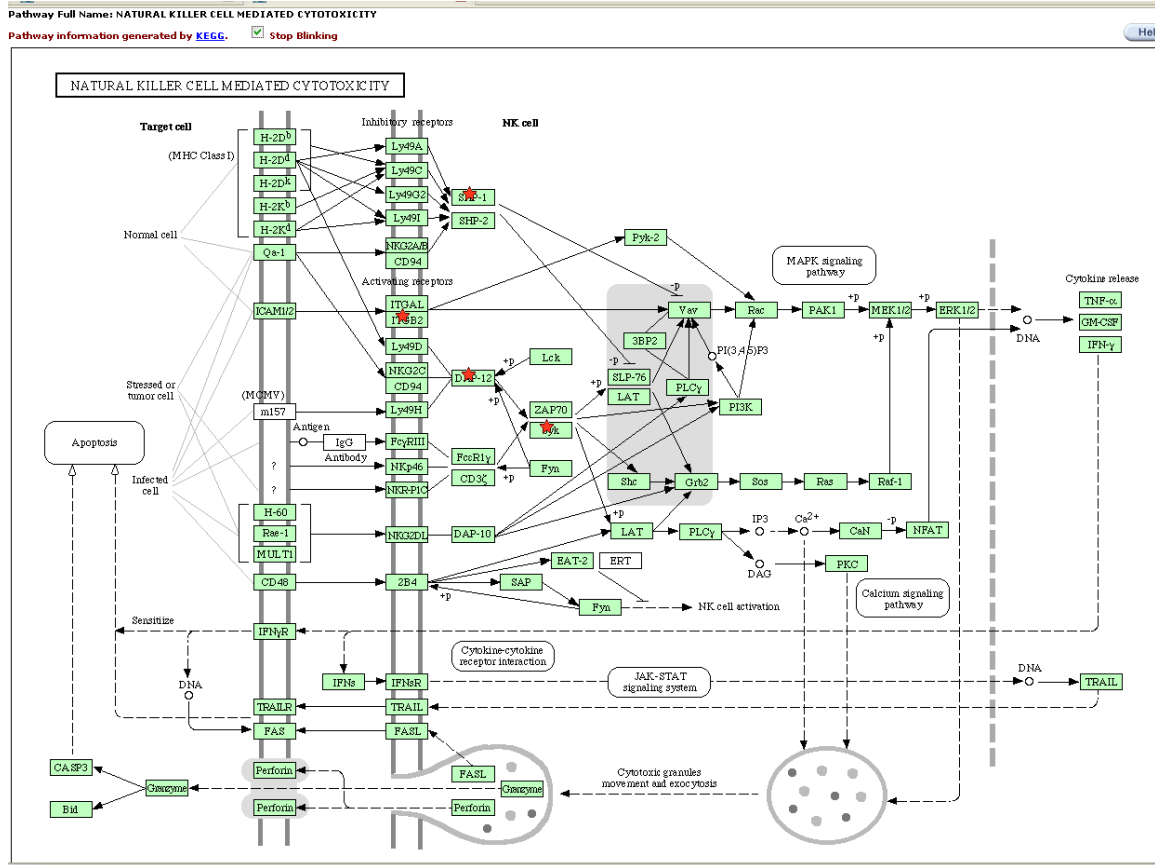
Figure 12. DAVID results showing one KEGG pathway (using significant genes as input)

Note: one limitation of the pathway analysis performed using GSA is that the gene set from the Broad Institute contains genes from the human genome, and the microarray experiment contains genes from the mouse genome. The genes from the KEGG pathways contain gene symbols from both human and animal. In this analysis, all gene symbols were converted to upper-case, so that the analysis was across species.

49

In the second statistical model, one-way ANOVA identified 670 genes that accurately classified samples by maternal diet exposure and PCA showed clear evidence of clustering by diet group (Figure 13B). Class differences were greatest between regular (5% fat) diet and high-fat (24%) based on flaxseed oil groups, with the corn oil group being intermediate in its response. This pattern was similar to that observed for body growth curves (Figure 10). Differential transcript abundance profiles could be partitioned neatly into two clusters of genes (Figure 13C). One cluster had 240 genes with downward programming by flaxseed oil. Gene Ontology (GO) scores enriched this cluster for biological themes to metabolism pathways: oxidative phosphorylation, TCA cycle, fatty acid metabolism, glyocolysis/gluconeogenesis, propanoate metabolism, and pyruvate metabolism. The other cluster had 430 genes with upward programming by flaxseed oil. GO scores enriched this cluster for morphoregulatory pathways: cell adhesion molecules and cytokine pathways. We may conclude from these results that maternal dietary fat exposure significantly influenced metabolic programming, and signaling networks in the mammary gland of female offspring.
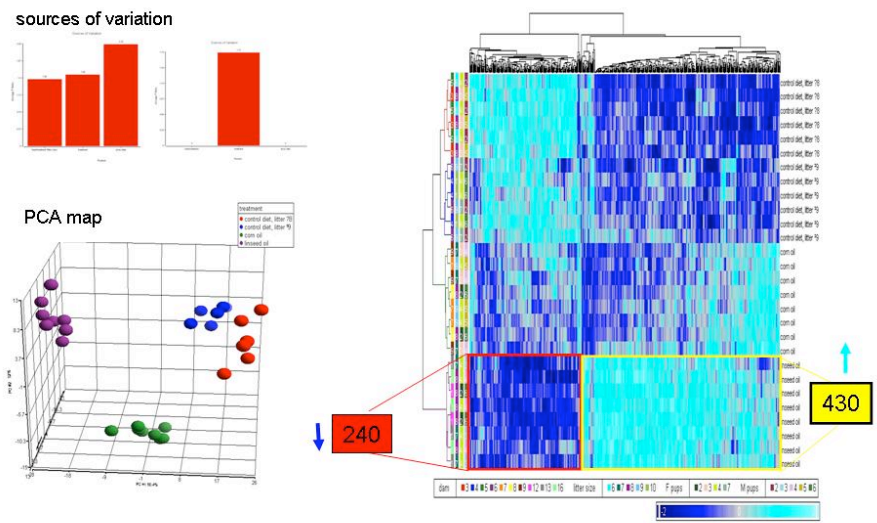
Figure 13. Developmental programming of gene expression in the mammary gland

Test diets (Figure 1) given to 6-weeks of age, and regular diet thereafter. Microarray analysis performed at 10-weeks of age; sources of variation identified for litter-size, treatment, and batch. After removing litter / batch effects the source of variance was plotted for regular diet (blue, red; n=12), fatty-diet corn oil (green, n=8) and fatty-diet flaxseed oil (magenta, n=8) using PCA. The heat-map shows clustering of 670 differentially-regulated genes (horizontal) and samples (vertical)

4.5. Discriminatory genes identified through BVS

Because PCA is a popular dimension-reduction tool in statistical models but as a compression technique basically categorizes data by variance, using a linear transformation to shift coordinate systems and compute values in the rotated coordinate system. By doing so, PCA decorrelates the inputs and makes them orthogonal; however, this linear decomposition is not inimical to gene interactions that are highly nonlinear and correlated. Also, PCA does not assess the joint effect of multiple variables and does not allow for evaluation of the original variables. Singular value composition, a very closely related concept (from which PCA can be derived), has also been applied to gene expression data to reduce dimensionality, but again, with the same problems.

Because the BVS algorithm does not scale well above ~1000 variables and has not been rigorously tested above this threshold, BVS was run on an ANOVA ($P <$ 0.0055) gene set comprising 914 genes. The algorithm was run using the control (isocaloric) and flaxseed groups only. This was done because the algorithm performs optimally under binary classification, although the method has been extended to the multinomial setting. Because of the restriction to the binary classification setting, the two groups were chosen by those with largest group weight divergence (isolcaloric vs. flaxseed). Distinct visited gene models (subsets) were sorted according to normalized joint posterior probability. Interesting gene subsets were chosen as the union of the top 10 models via normalized joint posterior probability. Gene subsets can also be selected via top marginal density. In fact, optimal gene selection via both joint posterior density and

marginal density resulted in very similar gene sets. The union of the top 10 models via

joint posterior density resulted in a 17 gene subset (Table 1).

Table 1. Nomenclature of the 17 gene subset (Bult CJ 2008)

Input	MGI Gene/Marker ID	Symbol	Name	Marker Type
M33425	MGI:96628	Jak1	Janus kinase 1	Gene
NM_025543	MGI:1913655	Mcts2	malignant T cell amplified sequence 2	Gene
AB008179	MGI:1298369	Pcdha7	protocadherin alpha 7	Gene
NM_011073	MGI:97551	Prf1	perforin 1 (pore forming protein)	Gene
NM_008404	MGI:96611	Itgb2	integrin beta 2	Gene
AF188504	MGI:2447166	Cmtm7	CKLF-like MARVEL transmembrane domain containing 7	Gene
NM_008615	MGI:97043	Me1	malic enzyme 1, NADP(+)-dependent, cytosolic	Gene
AF294811	MGI:87961	Agrn	agrin	Gene
NM_007378	MGI:109424	Abca4	ATP-binding cassette, sub-family A (ABC1), member 4	Gene
BC008156	MGI:2444541	Usp43	ubiquitin specific peptidase 43	Gene
AF154337	MGI:3710533	Gpr137b-ps	G protein-coupled receptor 137B, pseudogene	Gene
NM_011116	MGI:1333782	Pld3	phospholipase D family, member 3	Gene

AF317838	MGI:1931123	Inpp4a	inositol polyphosphate-4-phosphatase, type I	Gene
NM_010877	MGI:97284	Ncf2	neutrophil cytosolic factor 2	Gene
NM_007377	MGI:1197518	Aatk	apoptosis-associated tyrosine kinase	Gene
NM_007781	MGI:1339760	Csf2rb2	colony stimulating factor 2 receptor, beta 2, low-affinity (granulocyte-macrophage)	Gene
NM_011581	MGI:98738	Thbs2	thrombospondin 2	Gene
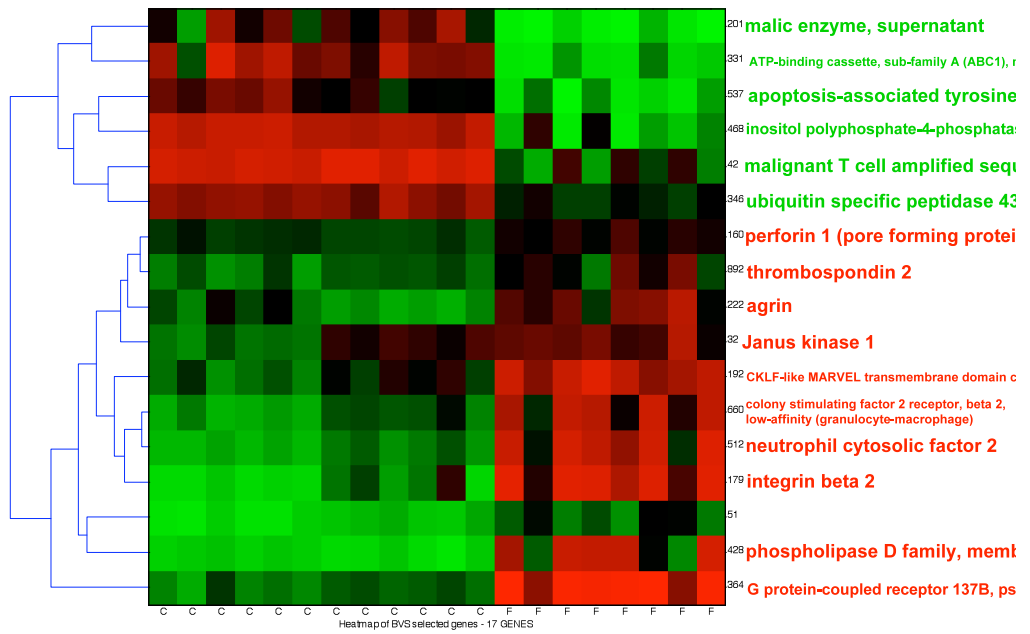
Figure 14. Heatmap of 17 genes from Bayesian variable selection, representing the union of the top 10 models via top posterior density

Table 2. List of genes in the 17 gene subset upregulated and their respective functions (per GeneGo) (Bult CJ 2008)

MGI	Gene Symbol	Gene Name	Goid	Code
MGI:96628	Jak1	Janus kinase 1	GO:0019221	cytokine and chemokine mediated signaling pathway
			GO:0007167	enzyme linked receptor protein signaling pathway
			GO:0018108	peptidyl-tyrosine phosphorylation
			GO:0007243	protein kinase cascade
			GO:0004713	protein tyrosine kinase activity
			GO:0005524	ATP binding
			GO:0005856	cytoskeleton
			GO:0004718	Janus kinase activity
			GO:0016301	kinase activity
			GO:0016020	membrane
			GO:0004715	non-membrane spanning protein tyrosine kinase activity
			GO:0000166	nucleotide binding
			GO:0006468	protein amino acid phosphorylation
			GO:0004672	protein kinase activity
			GO:0016740	transferase activity
			GO:0005515	protein binding
MGI:97551	Prf1	perforin 1 (pore forming protein)	GO:0019835	cytolysis
			GO:0016023	cytoplasmic membrane-bounded vesicle
			GO:0005509	calcium ion binding
			GO:0016021	integral to membrane
			GO:0016020	membrane
			GO:0005615	extracellular space
MGI:96611	Itgb2	integrin beta 2	GO:0009986	cell surface
			GO:0007229	integrin-mediated signaling pathway
			GO:0005624	membrane fraction
			GO:0045121	membrane raft

			GO:0005488	binding
			GO:0007160	cell-matrix adhesion
			GO:0008305	integrin complex
			GO:0016020	membrane
			GO:0005515	protein binding
			GO:0004872	receptor activity
			GO:0050798	activated T cell proliferation
			GO:0007155	cell adhesion
			GO:0045123	cellular extravasation
			GO:0030593	neutrophil chemotaxis
			GO:0016021	integral to membrane
MGI:2447166	Cmtm7	CKLF-like MARVEL transmembrane domain containing 7	GO:0006935	chemotaxis
			GO:0005125	cytokine activity
			GO:0005615	extracellular space
			GO:0016020	membrane
			GO:0016021	integral to membrane
MGI:3710533	Gpr137b-ps	G protein-coupled receptor 137B, pseudogene		
MGI:1333782	Pld3	phospholipase D family, member 3	GO:0003824	catalytic activity
			GO:0005783	endoplasmic reticulum
			GO:0016787	hydrolase activity
			GO:0016042	lipid catabolic process
			GO:0016020	membrane
			GO:0008152	metabolic process
			GO:0005624	membrane fraction
			GO:0004630	phospholipase D activity
			GO:0016021	integral to membrane
MGI:97284	Ncf2	neutrophil cytosolic factor 2	GO:0001669	acrosome
			GO:0005737	cytoplasm
			GO:0005488	binding
			GO:0005829	cytosol
			GO:0006742	NADP catabolic process
			GO:0006801	superoxide metabolic process
			GO:0016175	superoxide-generating

				NADPH oxidase activity
MGI:1339760	Csf2rb2	colony stimulating factor 2 receptor, beta 2, low-affinity (granulocyte-macrophage)	GO:0019221	cytokine and chemokine mediated signaling pathway
			GO:0004896	hematopoietin/interferon-class (D200-domain) cytokine receptor activity
			GO:0004907	interleukin receptor activity
			GO:0016020	membrane
			GO:0004872	receptor activity
			GO:0016021	integral to membrane
MGI:98738	Thbs2	thrombospondin 2	GO:0005509	calcium ion binding
			GO:0007155	cell adhesion
			GO:0005576	extracellular region
			GO:0008201	heparin binding
			GO:0005515	protein binding
			GO:0005198	structural molecule activity
			GO:0005615	extracellular space

Table 3. List of genes in the 17 gene subset downregulated and their respective functions (per GeneGo) (Bult CJ 2008)

MGI	Gene Symbol	Gene Name	Goid	Code
MGI:1913655	Mcts2	malignant T cell amplified sequence 2	GO:0005737	cytoplasm
			GO:0006355	regulation of transcription, DNA-dependent
			GO:0006350	transcription
			GO:0003723	RNA binding
MGI:109424	Abca4	ATP-binding cassette, sub-family A (ABC1), member 4	GO:0005524	ATP binding
			GO:0016887	ATPase activity
			GO:0016021	integral to membrane
			GO:0016020	membrane
			GO:0017111	nucleoside-triphosphatase activity
			GO:0000166	nucleotide binding
			GO:0050896	response to stimulus
			GO:0006810	transport
			GO:0006649	phospholipid transfer to membrane
			GO:0005548	phospholipid transporter activity
			GO:0004012	phospholipid-translocating ATPase activity
			GO:0045494	photoreceptor cell maintenance
			GO:0007601	visual perception
			GO:0042626	ATPase activity, coupled to transmembrane movement of

				substances
			GO:0005887	integral to plasma membrane
MGI:1931123	Inpp4a	inositol polyphosphate-4-phosphatase, type I	GO:0016787	hydrolase activity
			GO:0016316	phosphatidylinositol-3,4-bisphosphate 4-phosphatase activity
			GO:0006798	polyphosphate catabolic process

Gene ontology (GO) enrichment for the 17 genes was performed using both biological process and molecular function ontologies. Code written in MATLAB[4] grabbed GOids (biological process and molecular function) associated with each gene in the subset. To get a systems-level appreciation of these GOids, expression values of genes (in the original BVS analyzed dataset ($P < 0.0055$)) having membership to subset-derived GOids (from the 17 gene subset) were grabbed and for each subset-derived GOid, the constitutive gene expression levels were averaged. For example, a biological process GOid of the 17 gene subset is 'neutrophil chemotaxis.' The gene expression profiles of all genes having membership to this GOid in the $P < 0.0055$ geneset were averaged. This provides information on the behavior of this 'neutrophil chemotaxis' GOid on the basis of its constituent genes' expression profiles. Because the GOids have now been characterized with gene expression data (from their constituentive genes), the GOids can be visualized using heatmaps (Figure 15 and Figure 16). Furthermore, the 17 gene subset was able to correctly classify all control and flax pups as either a control or flax group member using a sampling based cross-validation procedure (

Table 4). This analysis clearly correlates a decrease in expression of metabolism (energy) pathways with a reciprocal increase in cytokine and inflammatory (adipocytokine) signaling pathways. More specifically, upward programming was seen in chemotaxis, cell adhesion, cytokine and integrin pathways, T-cell and neutrophil activity, and cytolysis. Downward programming was seen chiefly in metabolic pathways. These findings suggest a developmental reprogramming of pathways in energy metabolism (decreased) and adipocytokine (increased) signaling, linked with the body weight defect between the flaxseed and control diets in mammary tissue.

Table 4. 100% classification rate for control and flax pups using the 17 gene subset

Pup	Group Prediction 0= control 1 = flax	Group Prediction 0= control 1 = flax	Actual Group 0= control 1 = flax
1	0.0304	0	0
2	0.1214	0	0
3	0.0358	0	0
4	0.031	0	0
5	0.0355	0	0
6	0.0524	0	0
7	0.0406	0	0
8	0.0709	0	0
9	0.0462	0	0
10	0.0413	0	0
11	0.0591	0	0
12	0.0439	0	0
13	0.9786	1	1
14	0.9202	1	1
15	0.9386	1	1
16	0.9682	1	1
17	0.9565	1	1
18	0.8816	1	1
19	0.85	1	1
20	0.9779	1	1
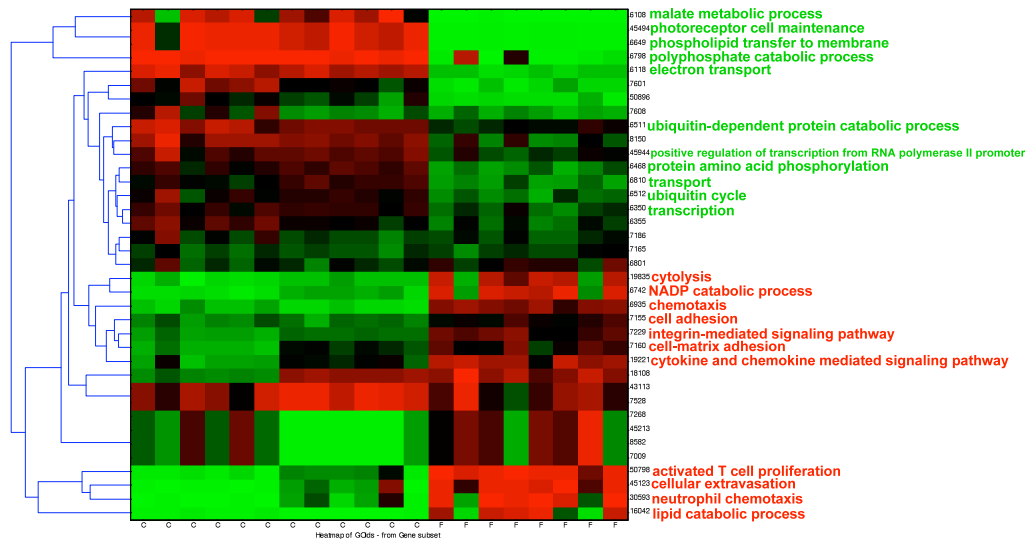
Figure 15. Heatmap of biological process GOids from the 17 genesubset (from Bayesian variable selection representing the union of the top 10 models according to posterior density)
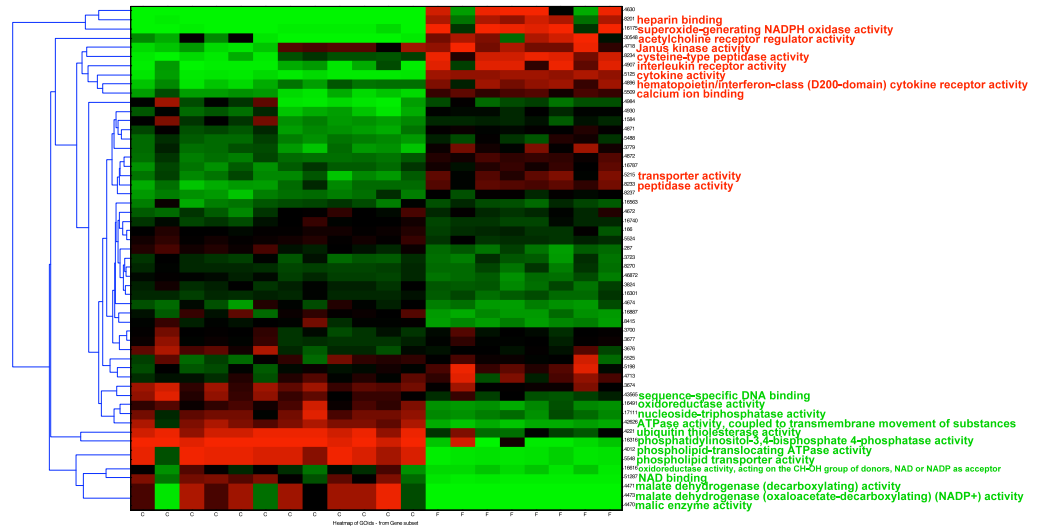
Figure 16. Heatmap of molecular function GOids from the 17 gene subset (from Bayesian variable selection representing the union of the top 10 models according to posterior density)
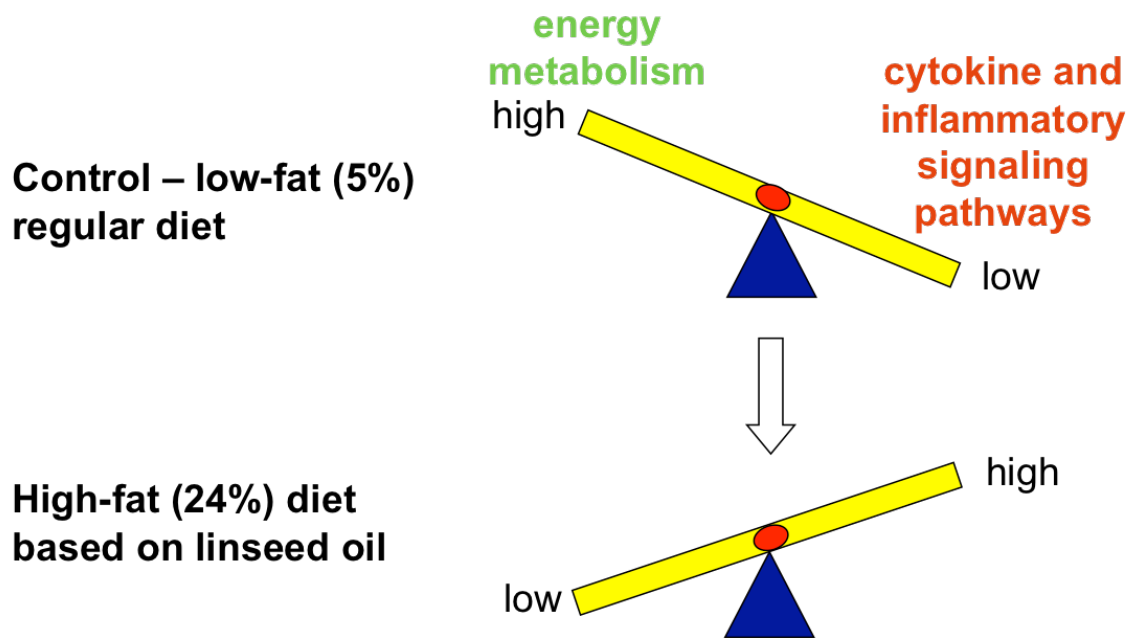
Figure 17. Reciprocal behavior between energy and immune pathways

To further explore this reciprocal relationship MetaCore^{TM} V 5.0 analysis Suite from GeneGO Inc. was used to generate a genetic network of the adipocytokine pathway based on gene expression values of pathway genes in the entire microarray data set (Figure 18). A canonical representation of the adipocytokine pathway in KEGG was colored for expression levels using GeneSpring GX 7.3, Agilent Technologies Inc. (**Figure 19**).
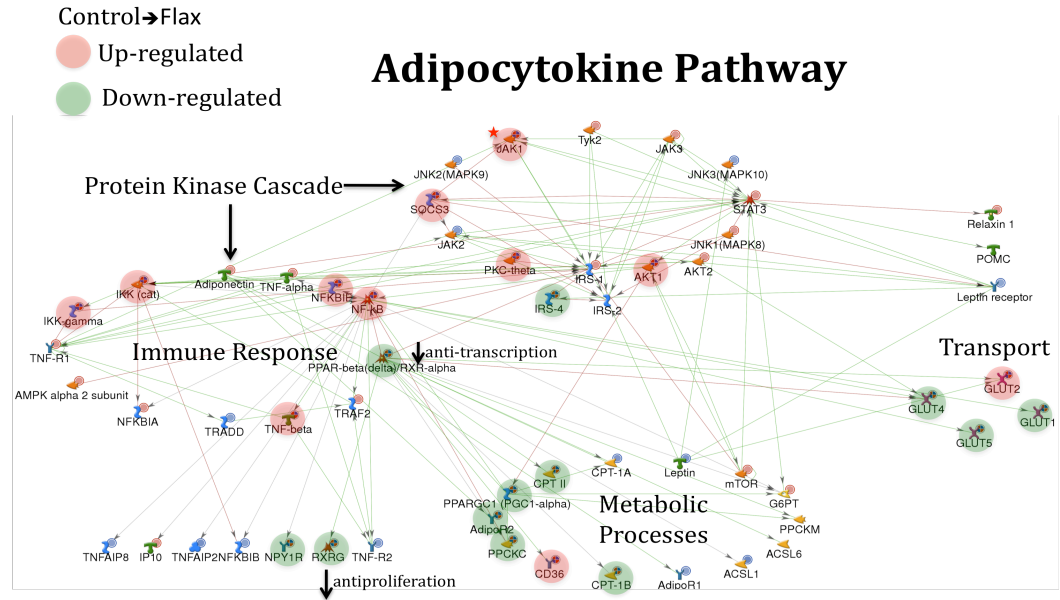
Figure 18. Adipocytokine pathway

In the pathway, primarily immune-related and signaling pathways were up-regulated, and in general, metabolic pathways were down-regulated (Figure 17). These are shown in Table 5 and Table 6 respectively.

Table 5. Up-regulated adipocytokine pathway genes

Up		
Gene Symbol	**Protein**	**Protein Name**
*JAK1	JAK1	Tyrosine-protein kinase JAK1
SOCS3	SOC3	Supressor of Cytokine Signaling
PKC-theta	KPCT	Protein kinase theta type
AKT1	AKT1	RAC-alpha serine/threonine-protein kinase
IKK-gamma	IKKB	Inhibitor of nuclear factor kappa-B kinase subunit beta
IKK (cat)(IKBKG)	NEMO	NF-kappa-B essential modulator
NFKBIE	IKB	NF-kappa-B inhibitor epsilon
NFkb		
- RELA	TF65	Transcription factor p65
- RELB	RELB	transcription factor RelB
- REL	REL	C-Rel proto-oncogene protein
- NFKB2	NFKB2	Nuclear factor NF-kappa-B p100 subunit
LTA	TNF-beta	Lymphotoxin-alpha precursor
CD36	CD36	Platelet glycoprotein 4
SLC2A2	GTR2 (GLUT2)	Solute carrier family 2, facilitated glucose transporter member 2

Table 6. Down-regulated adipocytokine pathway genes

Down		
Gene Symbol	**Protein**	**Protein Name**
PPAR-beta(delta)/RXR-alpha	PPARD	Peroxisome proliferator-activated receptor cells
IRS-4	IRS-4	Insulin-receptor substrate 4
SLC2A1	GTR2 (GLUT1)	Solute carrier family 2, facilitated glucose transporter member 1
SLC2A4	GTR4 (GLUT4)	Solute carrier family 2, facilitated glucose transporter member 4
SLC2A5	GTR5 (GLUT5)	Solute carrier family 2, facilitated glucose transporter member 4=5
NPY1R	NPY1R	Neuropeptide Y receptor type 1
RXRG	RXRG	Retinoic acid receptor RXR-gamma
CPTII	CPT2	Carnitine O-palmitoyltransferase 2, mitochondrial precursor
PPARGC1 (PGC1-alpha)	PRGC1	Peroxisome proliferator-activated receptor gamma coactivator 1-alpha'
AdipoR2	ADR2	Adiponectin receptor protein 2
PCK1	PPCKC	Phosphoenolpyruvate carboxykinase, cytosolic [GTP]
CPT-1B	CPT1B	Carnitine O-palmitoyltransferase I, muscle isoform

Genespring (VERSION GX7.3, 5301 Stevens Creek Blvd, Santa Clara CA 95051, United States) was used to generate a canonical representation of the adipocytokine pathway using expression data via KEGG and is represented in **Figure 19**. Of the 22680 Genbank Ids, 17487 and 21551 were mapped to unique gene IDs and EntrezGene IDs respectively.
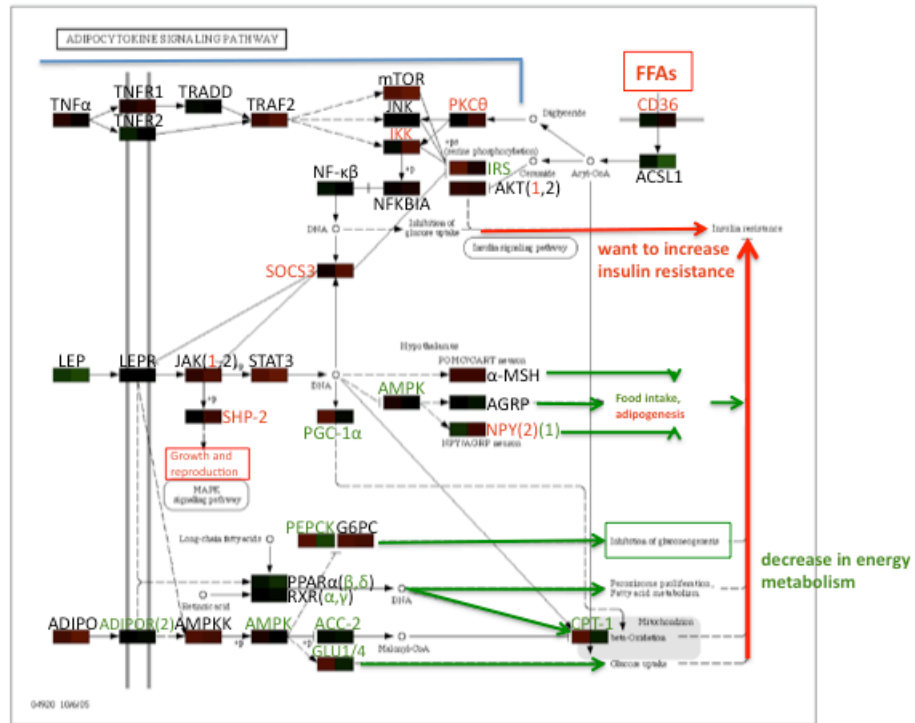
Figure 19. Canonical adipocytokine pathway per KEGG using Genespring

5. DISCUSSION

The adipocytokine pathway experienced a number of changes, with up and down-regulated groups of genes showing displaying biological themes. Up-regulated genes primarily present protein kinase casacde and immune response pathways, whereas down-regulated genes are indicative of metabolic pathways. These genes, their biological functions, and relevance to oncogenesis and obesity are included in Table 7 and Table 8.

More specifically, JAK-1 was independently implicated in potential relevance to high-fat fetal reprogramming by BVS and by pathway analysis, where the adipocytokine pathway displayed a large number of gene expression changes in constitutive genes. JAK-1 is a key component of the intracellular signaling cascade, and is an upstream effector of downstream transcript modulation of immune processes in particular. It is intimately involved in cytokine signaling and plays a role in initiaing such responses. JAK-1 stimulates stat5, a conventional prolactin signalling protein. Moreover, JAK-1 also recuits new signals, such as stat3 and ERK. These are considered to be tumor-promoting in breast tissue (Lee, Joung et al. 2006). It is interesting that JAK-1 retained such high expression in ex high-fat pups. This prolonged over-expression may effect many downstream changes, especially involving inflammatory processes, that are related to the pathogenesis of breast carinoma. As some of these genes and pathways have been implicated in promoting breast cancer development, JAK-1 inhibitors may be useful for breast cancer treatment. Another signal transduction element displaying over-expression

is protein kinase B (AKT-1), which is critical for transmitting growth signals. Also, AKT-1 is responsible for inactivating components of the apoptosis. Moreover, AKT-1 expression have been implicated in accelerated breast tumourigenesis (Young, Nolte et al. 2008). Increased AKT-1 expression is also found with increased serum insulin levels, as insulin activates AKT-1 and 2. High insulin levels will lead to insulin resistance.

Small heterodimer partner 2 (SHP-2) is a cytoplasmic SH2 domain and is involved in signaling pathways for growth factors and cytokines. This gene displayed up-regulation in pups dosed with high-fat diets. Particularly, it is a critical intracellular regulator in mediating cell proliferation and differentiation. Given this function, and its positive relationship with lymph node metastasis and tumor grade, it has been suggested that SHP2 promotes breast oncogensis (Zhou, Coad et al. 2008). It is important to note that SHP2 effects downstream transciption changes of epidermal growth factor receptor (EGFR) and enhances EGFR activity during mouse growth and development (Qu, Yu et al. 1999). EGFR has been implicated in the breast carcinoma pathogensis.

AMP-activated protein kinase (AMPK) was significantly down-regulated. AMPK's net effect is fatty acid oxiation, where lipogensis is inhibited. Interestingly, adiponectin expression is high in both control and dosed groups, where the dosed group had even higher expression, which is contrary to the observation that in obesity, adiponectin expression is typically markedly depressed. This observation is not well understood in infants and children, and as tissue sections were collected when mice were in puberty, the high adiponectin expression and obesity are probably not related. This is supported by the observation that adiponectin transcripts are over-expressed in adipogenic tissues (during development and differentiation). Adiponectin receptor

(ADIPOR-2) displayed marked decreased expression and is likely due to the diet-induced obesity and regulatory feedback from high expression of adiponectin. This would reduce adiponectin sensitivity, which would eventually lead to insulin resistance, a consequence of the adipocytokine pathway when showing decreased metabolic processes and increased cytokine/inflammatory pathways.

Neuropeptide Y receptor 2 (RPY-2R) expression was significatly increased, which effects decreased food intake (appetite) and adipogenesis. These, as a result of diet-induced obesity, also lead to insulin resistance. Most of the genes showing significant down-regulation are related to metabolic processes. CPT-1 was significantly down-regulated, and is responsible for fatty acid oxidation and transport of long chain fatty acids across the mitochondrial membrance (binds them to carnitine). Also, acetyl-CoA carboxylase 2 (ACC-2) is another key regulator of mitochondrial fat oxidation that displayed downward expression in the experimental groups. It has been noted in the litertaure that ACC-2 negative mice have a normal lifespan, and that ACC-2 targeted inhibition may be a viable therapeutic option for the treatment of obesity and type 2 diabetes mellitus.

These metabolic processes do not demonstrate the relationship to breast carcinogensis as do the signal tranduction and regulatory immune processes that underwent significant upward cellular response reprogramming. However, they do have a strong relationship with obesity, and more significantly, diabetes mellitus. Generally, this is a consequence of the increase in insulin resistance, which is the net outcome of the upwardly expressed signalling and immune processes and downwardly expressed metabolic processes in the adipocytokine pathways. It is important to note that these

changes reflect cellular responses and have been observed effecting transcription activity in the mammary tissue of female offspring via microarray profiling.

Although these changes were represented in the adipocytokine pathway, this is not the only pathway involved. These changes in the adipocytokine pathway are only an indication of statistical/bioinformatic programming taken place. Furthermore, the changes in the adipocytokine pathway may be interpreted from different persectives. For example, the changes seen per this analysis may comprise some portion of a larger system-wide regulatory change, or there may have been changes in cellularity that express this particular pathway. These potential changes in cellularity may reflect an imbalance cellularity of parenchyma and adipose tissues in the mammary gland.

Overall, inflammatory processes were statistically up-regulated, and metabolic processes were statistically down-regulated in the mammary gland. As shown earlier, some immune components are significantly upstream and effect many downstream transcript changes. It was observed that these changes in inflammatory processes were maintained in ex-fat period until 10 wks and may reflect a possible permanent immune dysregulation. The increased pup weight for ex-fat diets is likely due to increased adipogenesis because of the nature of fatty diet. It is important to note that the mammary glands themselves were not weighed. However, it is likely they experienced a weight defect in accordance with body weight trajectory.

Table 7. Processes and cancer implications for up-regulated genes of Adipocytokine pathway

Up-regulated Genes		
Name	**Biological Process**	**Information**
Protein kinase C theta (PKC)	signal transduction	• Members of the PKC family are divided into three groups based on their molecular structures and activating mechanisms: 1) Conventional PKC (alpha, beta, and gamma) requiring calcium, phosphatidylserine (PS), and diacylglycerol (DG) for activation; 2) novel PKC (sigma, epsilon, eta, and theta) activated independent of calcium; and 3) Atypical PKC (zeta and lamda), which are independent of both calcium and DG. High-level expression of PKC-theta was found in skeletal muscle, lung, T cells, and brain, and minimal expression in cardiac muscle, placenta, and liver. PKC theta is a autophosphorylated protein kinase.
Protein kinase B (AKT-1)	signal transduction	• Mice lacking Akt1 display a 25% reduction in body mass, indicating that Akt1 is critical for transmitting growth promoting signals, most likel via the igf1 receptor. Mice lacking atk1 are also resistant to cancer: they experience considerable delay in tumor growth initiated by the large T antigen or the Neu oncogene. • inactivates components of the apoptotic machinery. • Mammary carcinoma: Expression of activated Akt1 in MMTV-c-ErbB2 mice accelerates tumourigenesis with a reduced requirement for signalling through the EGFR family, as well as a reduced requirement for a subset of downstream signaling molecules with a metabolic shift in the tumours from bitransgenic mice. (Young, Nolte et al. 2008)
Suppressor of cytokine signaling 3 (SOCS 3)	cytokine-inducible negative regulators of cytokine signaling	• The expression of this gene is induced by various cytokines, including IL6, IL10, and interferon (IFN)-gamma. The protein encoded by this gene can bind to JAK2 kinase, and inhibit the activity of JAK2 kinase. Studies of the mouse counterpart of this gene suggested the roles of this gene in the negative regulation of fetal liver hematopoiesis, and placental development. • Elevated expression of SOCS genes is a specific lesion of breast-cancer cells that may confer resistance to proinflammatory cytokines and trophic factors, by shutting down STAT1/STAT5 signaling that mediate essential functions in the mammary gland. (Evans, Yu et al. 2006)
Janus kinase 1 (JAK1)	Signal transduction	• upstream effector of downstream transcript modulation (immune processes) • Signaling of type 1 and 2 cytokines (IL-2, IL-4, gp130, CNTF-R, NNT-1R, Leptin-R, IFN-alpha,beta,gamma interferon's, IL-10) • In breast cancer cells, Jak1 not only stimulates conventional prolactin signaling via proteins such as Stat5, but also that Jak1 recruited new signals, especially Stat3

		and ERK. Stat3 and ERK typically are considered tumor-promoting, so inhibitors of Jak1 may become useful in breast cancer treatment. • Results that may provide the basis for identifying another mechanism of breast tumorigenesis via the JAK [?]/STAT pathway in hypoxia.(Lee, Joung et al. 2006)
Small heterodimer partner 2 (SHP-2)	member of the nuclear receptor family of intracellular transcription factors	• SHP-2, a cytoplasmic SH2 domain containing protein tyrosine phosphatase, is involved in the signaling pathways of a variety of growth factors and cytokines. Recent studies have clearly demonstrated that this phosphatase plays an important role in transducing signal relay from the cell surface to the nucleus, and is a critical intracellular regulator in mediating cell proliferation and differentiation (Qu 2000) • Given SHP2's positive role in cell growth, transformation and stem cell survival, the positive relationship of its overexpression to lymph node metastasis, nuclear accumulation of hormone receptors and higher tumour grade suggests that SHP2 promotes breast oncogenesis (Zhou, Coad et al. 2008) • Feeds into MARK signaling pathway • Effects Downstream transcription of epidermal growth factor receptor (EGF-R) => Thus, we provide biological evidence here that protein-tyrosine phosphatase SHP-2 acts to enhance information flow from the EGF-R in mouse growth and development (Qu, Yu et al. 1999) **HUGE RELATIONSHIP TO BREAST CANCER**
Neuropeptide Y 2 (NPY-2)	peptide neurotransmitter found in the brain and autonomic nervous system	• Decrease in food intake, increase in adipogenesis
IKappa B Kinase gamma (IKK-gamma)	upstream NF-κB signal transduction cascade	• Regulatory subunit of IKK complex, which activates NF-KappaB. IKKbeta (and IKKgamma) are essential for rapid NF-kappaB activation by proinflammatory signaling cascades, such as those triggered by tumor necrosis factor alpha (TNFalpha) or lipopolysaccharide (LPS) (Hacker and Karin 2006) • IKK/NF kappa B system is generally overexpressed in breast cancer cells and there is heterogeneity in expression levels of individual members between different cell lines
Tumor necrosis factor beta (lymphotoxin beta) (TNF-b)	Immune response (cytokine)	• TNF-beta is involved in the regulation of various biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, coagulation, and neurotransmission. TNF-beta is secreted as a soluble polypeptide, but can form heterotrimers with lymphotoxin-beta, which effectively anchors the TNF-beta to the cell surface. TNF-beta is cytotoxic to a wide range of tumor cells.
peroxisome proliferator-activated receptors beta-delta (PPAR-	nuclear hormone receptors that bind peroxisome proliferators and control the size	• differentiation, lipid accumulation, directional sensing, polarization, and migration in keratinocytes • After fertilisation, PPAR-gamma and PPAR-beta/delta are essential regulators of placentation and the subsequent development of key metabolic tissues such as skeletal

bd)	and number of peroxisomes produced by cells	muscle and adipose cells (Rees, McNeil et al. 2008)
Glucose Transporter 2 (GLUT-2)	carrier protein and gene which is involved in passive glucose transport	

Table 8. Processes and cancer implications for down-regulated genes of Adipocytokine pathway

Down-regulated Genes		
Name	**Biological Process**	**Information**
Insulin receptor substrate (IRS)	elicits insulin's actions	
Phosphoenolpyruvate carboxykinase (PEPCK)	Gluconeogensis	
Adiponectin receptor 2 (ADIPO-R(2))	receptor for adiponectin	• Both of the receptors activate AMPK and PPAR alpha metabolic pathways leading to an increase in fatty acid oxidation, glucose uptake and a decreased rate of gluconeogenesis, thus enhancing insulin sensitivity. Moreover effects of adiponectin mimic many metabolic actions of insulin such as augmenting blood flow and glucose disposal in NO-dependent manner. The precise mechanism of regulation of plasma adiponectin level is unknown. Recently the mechanism of transcriptional activation of adiponectin gene via PPAR gamma was described. Its level seems to be decreased by TNFalfa and beta-adrenergic agonists (Szopa, Malczewska-Malec et al. 2004) • Obesity decreased expression levels of AdipoR1/R2, thereby reducing adiponectin sensitivity, which finally leads to insulin resistance, the so-called "vicious cycle." (Kadowaki and Yamauchi 2005)
AMP-activated protein kinase (AMPK)		• Recently, low adiponectin levels are significantly associated with an increased breast cancer risk (Takahata, Miyoshi et al. 2007) • The net effect of AMPK activation is stimulation of hepatic fatty acid oxidation and ketogenesis, inhibition of cholesterol synthesis, lipogenesis, and triglyceride synthesis, inhibition of adipocyte lipolysis and lipogenesis, stimulation of skeletal muscle fatty acid oxidation and muscle glucose uptake, and modulation of insulin secretion by pancreatic beta-cells.
Acetyl-CoA carboxylase 2 (ACC-2)		• key regulator of mitochondrial fat oxidation • Taken together with previous work demonstrating that Acc2(-/-) mice have a normal lifespan, these data suggest that Acc2 inhibition is a viable therapeutic option for the treatment of obesity and type 2 diabetes (Choi, Savage et al. 2007)
GLUT 1,4,5	carrier protein and gene which is involved in passive glucose transport	

CPT-1, 2		• transport of long chain fatty acids across the membrane by binding them to carnitine. • Genes for fatty acid oxidation (Sampath, Miyazaki et al. 2007)
PGC-1alpha		• regulates the genes involved in energy metabolism • The protection from obesity involves elevated oxygen consumption/energy expenditure and increased fatty acid oxidation in adipose tissue with concurrent increased mitochondria genesis, up-regulation of PGC-1alpha and UCP-2 [?], and down-regulation of perilipin [?]. (Bansode, Huang et al. 2008)
PPCKC		• main control point for the regulation of gluconeogenesis
NPY-1R		• <u>Y1-R-/- mice showed a moderate obesity and mild hyperinsulinemia without hyperphagia</u>. Although there was some variation between males and females, typical characteristics of Y1-R-/- mice include: greater body weight (females more than males), <u>an increase in the weight of white adipose tissue (WAT)</u> (approximately 4-fold in females), an elevated basal level of plasma insulin (approximately 2-fold), impaired insulin secretion in response to glucose administration, and a significant changes in mitochondrial uncoupling protein (UCP) gene expression (up-regulation of UCP1 in brown adipose tissue and down-regulation of UCP2 in WAT).(Kushi, Sasai et al. 1998)
RXRG		• involved in mediating the antiproliferative effects of retinoic acid (RA). • Retinoid X receptor gamma-deficient mice have increased skeletal muscle lipoprotein lipase activity and less weight gain when fed a high-fat diet.
IRS-4		• Protein phosphatase 4 interacts with and down-regulates insulin receptor substrate 4 following tumor necrosis factor-alpha stimulation.(Mihindukulasuriya, Zhou et al. 2004) • Acts as an interface between multiple growth factor receptors possessing tyrosine kinase activity, such as insulin receptor, IGF1R and FGFR1, and a complex network of intracellular molecules containing SH2 domains.

Our studies aim to identify new biomarker leads for altered developmental (fetal) programming that correlate with early susceptibility to breast tumors. Such biomarker leads may translate into molecular diagnostics for pre-malignant foci in human patients. Comparative bioinformatics can reveal how closely the murine data reflects the human response. These scenarios can be enhanced in the applications of prior knowledge of established biomarkers for breast cancer (eg, ER-alpha, PrgR, EgfR, Her2-Neu, Cox2) drawn from national databases, statistical reductions in the high-dimension data from cellular microgenomics profiles, and the generation of new hypotheses regarding how developmental programming of large genetic regulatory networks in specific precursor target cell subpopulations might initiate and propagate changes in cell states leading to breast cancer.

6. SUMMARY AND CONCLUSIONS

In general, this study was performed as an exploratory study with gestational-lactational exposure of Harlan mice to high-fat diets switched at 6-weeks postnatal. The results can be summarized as follows. First, maternal high-fat diet altered growth trajectories and gene expression in the mammary gland at 10 weeks. Second, flaxseed oil diet (high in omega-3) had a lasting effect on body weight compared to a diet with corn oil (high in omega-6). Third, the genomic response signatures based on a comprehensive microarray analysis suggested that the changes were associated with a decrease in expression of energy metabolism pathways and a reciprocal increase in cytokine and inflammatory signaling pathways. Fourth, the changes in gene expression may reflect imbalance of cellularity of parenchymal and adipose tissue that may influence the risk of breast cancer.

We may conclude from this analysis that prenatal and perinatal exposure to a high-fat (24%) diet had a long-term effect on postnatal body weights versus the regular low-fat (5%) diet and that the latent effect persisted longer when the high-fat diet was provided by flaxseed oil versus corn oil. Computational analysis demonstrated a decrease in expression of energy pathways with a reciprocal increase in cytokine and inflammatory signaling pathways. Computational pathway analysis using GeneGo software showed these changes were reflected in the adipocytokine pathway, where the protein kinase cascade and immune response pathways displayed upward programming, and metabolic

processes displayed downward programming. Janus kinase 1 (JAK-1) and carnitine O-palmitoyltransferase I (CPT-1) expression levels were particularly affected. Because of these changes, we may conclude that maternal dietary fat exposure significantly influenced gene expression associated with metabolic programming and signaling networks in the mammary gland of female offspring. In addition, the adipocytokine pathway may be a sensitive trigger to dietary changes and this may influence or enhance activation of an immune response, a key event in cancer development. Together, these data suggest that maternal intake of a high-fat diet may predispose their offspring to breast cancer.

More specifically, due to experimental design, it may be stated that omega-3 PUFAs causally affect mouse adolescent obesity and immune and metabolic reprogramming of the mammary gland. Potential clinical implications of a enriched omega-3 diet are epigenetic risk factors for breast cancer via dysregulatory inflammatory changes and for type 2 diabetes via insulin resistance stemming from adolescent obesity.

REFERENCES

Armitage, J. A., I. Y. Khan, et al. (2004). "Developmental programming of the metabolic syndrome by maternal nutritional imbalance: how strong is the evidence from experimental models in mammals?" <u>J Physiol</u> **561**(Pt 2): 355-77.

Bansode, R. R., W. Huang, et al. (2008). "Protein kinase C deficiency increases fatty acid oxidation and reduces fat storage." <u>J Biol Chem</u> **283**(1): 231-6.

Barker, D. J. (1995). "Fetal origins of coronary heart disease." <u>BMJ</u> **311**(6998): 171-4.

Barker, D. J., A. R. Bull, et al. (1990). "Fetal and placental size and risk of hypertension in adult life." <u>BMJ</u> **301**(6746): 259-62.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing."

Binukumar, B. and A. Mathew (2005). "Dietary fat and risk of breast cancer." <u>World J Surg Oncol</u> **3**: 45.

Boyd, N. F., J. Stone, et al. (2003). "Dietary fat and breast cancer risk revisited: a meta-analysis of the published literature." <u>Br J Cancer</u> **89**(9): 1672-85.

Bult CJ, E. J., Kadin JA, Richardson JE, Blake JA; and the members of the Mouse Genome Database Group (2008). "The Mouse Genome Database (MGD): mouse biology and model systems." <u> Nucleic Acids Res 36(Database issue):D724-8. </u> .

Choi, C. S., D. B. Savage, et al. (2007). "Continuous fat oxidation in acetyl-CoA carboxylase 2 knockout mice increases total energy expenditure, reduces fat mass, and improves insulin sensitivity." <u>Proc Natl Acad Sci U S A</u> **104**(42): 16480-5.

De Assis, S. and L. Hilakivi-Clarke (2006). "Timing of dietary estrogenic exposures and breast cancer risk." <u>Ann N Y Acad Sci</u> **1089**: 14-35.

Dennis, G., Jr., B. T. Sherman, et al. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." <u>Genome Biol</u> **4**(5): P3.

Dobbing, J. (1993). "Fetal nutrition and cardiovascular disease in adult life." <u>Lancet</u> **341**(8857): 1421-2.

Evans, M. K., C. R. Yu, et al. (2006). "Expression of SOCS1 and SOCS3 genes is differentially regulated in breast cancer cells in response to proinflammatory cytokine and growth factor signals." Oncogene 26(13): 1941-1948.

Hacker, H. and M. Karin (2006). "Regulation and function of IKK and IKK-related kinases." Sci STKE 2006(357): re13.

Hilakivi-Clarke, L., A. Stoica, et al. (1998). "Consumption of a high-fat diet alters estrogen receptor content, protein kinase C activity, and mammary gland morphology in virgin and pregnant mice and female offspring." Cancer Res 58(4): 654-60.

Hortobagyi, G. N., J. de la Garza Salazar, et al. (2005). "The global breast cancer burden: variations in epidemiology and survival." Clin Breast Cancer 6(5): 391-401.

Howe, G. R., T. Hirohata, et al. (1990). "Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies." J Natl Cancer Inst 82(7): 561-9.

Kadowaki, T. and T. Yamauchi (2005). "Adiponectin and adiponectin receptors." Endocr Rev 26(3): 439-51.

Kato, I., S. Tominaga, et al. (1987). "Relationship between westernization of dietary habits and mortality from breast and ovarian cancers in Japan." Jpn J Cancer Res 78(4): 349-57.

Kushi, A., H. Sasai, et al. (1998). "Obesity and mild hyperinsulinemia found in neuropeptide Y-Y1 receptor-deficient mice." Proc Natl Acad Sci U S A 95(26): 15659-64.

Lee, M. Y., Y. H. Joung, et al. (2006). "Phosphorylation and activation of STAT proteins by hypoxia in breast cancer cells." Breast 15(2): 187-95.

Lucas, A. (1998). "Programming by early nutrition: an experimental approach." J Nutr 128(2 Suppl): 401S-406S.

Mamalakis, G., E. Jansen, et al. (2006). "Depression and adipose and serum cholesteryl ester polyunsaturated fatty acids in the survivors of the seven countries study population of Crete." Eur J Clin Nutr 60(8): 1016-23.

Mattisson, I., E. Wirfalt, et al. (2004). "High fat and alcohol intakes are risk factors of postmenopausal breast cancer: a prospective study from the Malmo diet and cancer cohort." Int J Cancer 110(4): 589-97.

McMillen, I. C. and J. S. Robinson (2005). "Developmental origins of the metabolic syndrome: prediction, plasticity, and programming." Physiol Rev 85(2): 571-633.

Mihindukulasuriya, K. A., G. Zhou, et al. (2004). "Protein phosphatase 4 interacts with and down-regulates insulin receptor substrate 4 following tumor necrosis factor-alpha stimulation." J Biol Chem **279**(45): 46588-94.

Mootha, V. K., C. M. Lindgren, et al. (2003). "PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes." Nat Genet **34**(3): 267-73.

P. J. Brown, M. V. T. F. (1998). "Multivariate Bayesian variable selection and prediction." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **60**(3): 627-641.

Parkin, D. M., F. Bray, et al. (2005). "Global cancer statistics, 2002." CA Cancer J Clin **55**(2): 74-108.

Prentice, R. L. and L. Sheppard (1990). "Dietary fat and cancer: consistency of the epidemiologic data, and disease prevention that may follow from a practical reduction in fat consumption." Cancer Causes Control **1**(1): 81-97; discussion 99-109.

Qu, C. K. (2000). "The SHP-2 tyrosine phosphatase: signaling mechanisms and biological functions." Cell Res **10**(4): 279-88.

Qu, C. K., W. M. Yu, et al. (1999). "Genetic evidence that Shp-2 tyrosine phosphatase is a signal enhancer of the epidermal growth factor receptor in mammals." Proc Natl Acad Sci U S A **96**(15): 8528-33.

Rees, W. D., C. J. McNeil, et al. (2008). "The Roles of PPARs in the Fetal Origins of Metabolic Health and Disease." PPAR Res **2008**: 459030.

Sampath, H., M. Miyazaki, et al. (2007). "Stearoyl-CoA desaturase-1 mediates the pro-lipogenic effects of dietary saturated fat." J Biol Chem **282**(4): 2483-93.

Sha, N., M. Vannucci, et al. (2004). "Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage." Biometrics **60**(3): 812-9.

Sinae, K., G. T. Mahlet, et al. (2006). "Variable selection in clustering via Dirichlet process mixture models." Biometrika **93**(4): 877-893.

Singh, A. V., C.D. Bastian, E. C. Rouchka, et al. (2007). "Integrative database management for mouse development: systems and concepts." Birth Defects Res C Embryo Today **81**(1): 1-19.

Szopa, M., M. Malczewska-Malec, et al. (2004). "[Adiponectin--adipocytokine with a broad clinical spectrum]." Przegl Lek **61**(2): 109-14.

Tadesse, M. G., N. Sha, et al. (2005). "Bayesian Variable Selection in Clustering High-Dimensional Data." <u>Journal of the American Statistical Association</u> **100**: 602-617.

Takahata, C., Y. Miyoshi, et al. (2007). "Demonstration of adiponectin receptors 1 and 2 mRNA expression in human breast cancer cells." <u>Cancer Lett</u> **250**(2): 229-36.

Young, C., E. Nolte, et al. (2008). "Activated Akt1 accelerates MMTV-c-ErbB2 mammary tumourigenesis in mice without activation of ErbB3." <u>Breast Cancer Research</u> **10**(4): R70.

Zhou, X., J. Coad, et al. (2008). "SHP2 is up-regulated in breast cancer cells and in infiltrating ductal carcinoma of the breast, implying its involvement in breast oncogenesis." <u>Histopathology</u>.

CURRICULUM VITAE

<div align="right">November, 2008</div>

Caleb D. Bastian

Current Address	**Permanent Address**
2428 Eagles Eyrie Court #6	4350 Calista Rd.
Louisville, KY. 40206	Cross Plains, TN. 37049
e-mail: caleb.bastian@louisville.edu	
phone: (615) 218 - 3781	

Education

University of Louisville   *Louisville, KY.*

2010	D.M.D.	Doctor of Dental Medicine	School of Dentistry
2010	M.B.A.	Professional MBA	College of Business
2010	M.S.	Anatomical Sciences and Neurobiology	Graduate School
2008	M.S.	Oral Biology	Graduate School

University of Tennessee   *Knoxville, TN.*
Summa Cum Laude, University Honors Scholar

2006	M.S.	Nuclear Engineering, Mathematics Minor	Graduate School
2006	B.S.	Nuclear and Radiological Engineering	Dept. of Nuclear Engineering

Vanderbilt University   *Nashville, TN.*

1997 - 2002	Myra Jackson Blair Scholar, Pre-College	Blair School of Music

Research

University of Louisville   *Louisville, KY.*

2008 - Present	Oral and Maxillofacial Radiology	School of Dentistry
	Dept. of Surgical & Hospital Dentistry	
	Perceptor: Allan Farman, BDS, PhD, MBA, DSc	
2008 - Present	Finance	College of Business
	Graduate Research Assistant (2008)	
	Perceptor: David Dubofsky, PhD	
2006 - Present	Bioinformatics, Computational Genetics	School of Dentistry
	Systems Analysis Laboratory, Birth Defects Center	
	Perceptors: Thomas Knudsen, PhD, Gregory Rempala, PhD, DSc	

Environmental Protection Agency, Research Triangle Park, NC.

Summer 2008	Computational Toxicology
	National Center for Computational Toxicology
	Perceptor: Thomas Knudsen, PhD

<div align="center">1</div>

University of Tennessee   Knoxville, TN.

<table>
<tr><td>2005 - 2006</td><td>Graduate Research Assistant</td><td>Dept. of Nuclear Engineering</td></tr>
</table>

Perceptors: Wes Hines, PhD, Lawrence Townsend, PhD

Publications

1. GREGORY REMPALA, **CALEB D. BASTIAN** . "Modified CART with comparison of variable selection techniques in high-dimensional data." Publication Pending 2008.

2. THOMAS B. KNUDSEN, AMAR V. SINGH, **CALEB D. BASTIAN**, M LUIJTEN, A DE VRIES, AND AH PIERSMA. "Altered Developmental (Fetal) Programming of the Mouse Mammary Gland in Female Offspring Following Maternal Dietary Exposures." Publication Pending 2008.

3. AMAR V. SINGH, ERIC C. ROUCHKA, GREG A. REMPALA, **CALEB D. BASTIAN**, AND THOMAS B. KNUDSEN. "Integrative Database Management for Mouse Development: System and Concepts." *Birth Defects Research (Part C)*. 81:1-19, 2007.

4. J. WESLEY HINES, PETER GROER, BELLE UPADHYAYA, ALEXANDER USYNIN, AND **CALEB D. BASTIAN**. "Improved Probability of Failure Analysis Using On-Line Equipment Condition Monitoring Data." EPRI, Palo Alto, CA, The University of Tennessee : 2007.

5. **CALEB D. BASTIAN.** "Analysis of large solar particle events with extraction of doses per energy contribution with implications for space radiation shielding." The University of Tennessee, 2005.

Presentations

<table>
<tr><td>2008</td><td>Thesis defense (M.S.) - Oral Biology - November 21
<i>Altered Developmental Programming of the Mouse Mammary Gland in
Female Offspring Following Perinatal Dietary Exposures.</i></td></tr>
<tr><td>2008</td><td>Research! Louisville, University of Louisville, October 21 (Poster).
<i>Altered Developmental Programming of the Mouse Mammary Gland in
Female Offspring Following Perinatal Dietary Exposures.</i></td></tr>
<tr><td>2008</td><td>American Association of Dental Research (AADR), Dallas, TX, April 4.
<i>Systems-Based Analysis of Developmental Programming of the Mammary Gland using
Bayesian Variable Selection.</i></td></tr>
<tr><td>2008</td><td>ULSD Student Convention, Louisville, KY, February 6 (Poster).
<i>Microarray-derived Biomarkers: Identification and Diagnostic Implications</i></td></tr>
<tr><td>2007</td><td>Research! Louisville, University of Louisville, October 16 (Poster).
<i>Systems-Based Analysis of Developmental Programming of the Mammary Gland using
Bayesian Variable Selection.</i></td></tr>
<tr><td>2007</td><td>CGeMM Computational Biology Core, University of Louisville, September 17.</td></tr>
</table>

Systems-Based Analysis of Developmental Programming of the Mammary Gland using Bayesian Variable Selection.

2006   American Nuclear Society (student), Rensselaer Polytechnic Institute, March 31. *Improved Probability of Failure (POF) Analysis using On-Line Equipment Condition Monitoring Data.*

Honors

2008	Research! Louisville (health sciences research competition) - 2nd place *Altered Developmental Programming of the Mouse Mammary Gland in Female Offspring Following Perinatal Dietary Exposures.*
2008	National Cancer Institute Research Grant (Renewal)(R25 CA044789)
2008	Honored (DMD) - Pharmacology and Toxicology
2007	ULSD Student Convention Clinical Research Poster Competition - 3rd place *Microarray-derived Biomarkers: Identification and Diagnostic Implications*
2007	Research! Louisville (health sciences research competition) - 1st place *Systems-Based Analysis of Developmental Programming of Mammary Gland using Bayesian Variable Selection*
2007	National Cancer Institute Research Grant (R25 CA044789)
2007	Honored (DMD) - Biochemistry, Microbiology, Research Seminar
2006	Most Outstanding Undergraduate in College of Engineering
2006	Phi Mu Alpha School of Music Scholastic Award
2006	Speaker at Chancellor's Banquet
2005 - 2006	Principal Oboe - Symphonic Band
2004 - 2006	Tau Beta Pi - Member
2005	*Rhodes Scholarship Applicant*
2005	UT/Oak Ridge National Lab Chancellors Honors Research Internship
2005	Passed Engineer in Training/Fundamentals of Engineering exam in Tennessee, National Council of Examiners for Engineering & Surveying (NCEES)
2004	Junior Outstanding Academic Achievement Award in Nuclear Engineering
2004	Pianist for All-Sing Competition - Phi Mu Alpha & Phi Mu
2003	Head of programming team in building optical character recognition software in MATLAB
2003 - 2005	Principal Oboe - Wind Ensemble
2002 - 2003	Principal Oboe - Symphonic Band
2002	Valedictorian, White House High School, TN.

Scholarships

2008	Graduate Research Assistantship (UofL: College of Business)
2007	School of Dentistry Scholarship
2006	Graduate Research Assistantship (UT: Dept. of Nuclear Engineering)
2005 - 2006	National American Nuclear Society Scholarship
2004 - 2006	National Academy for Nuclear Training Scholarship
2002 - 2006	Bicentennial Scholarship

2002 - 2006	Ned Ray McWherter Scholarship
2002 - 2006	Henry A. Haenseler Engineering Scholarship
2002 - 2006	Elizabeth Buford Shepherd Scholarship
2002 - 2006	Band Scholarship
2004 - 2005	Nuclear Engineering/Health Physics Scholarship program
	through MUSC/SCUREF
2004	Department of Energy Fellowship
2002 - 2004	Pasqua Nuclear Engineering Scholarship
2002 - 2004	Alumni Valedictorian Scholarship

Extracurricular Activies and Organizations

2008 - Present	UofL Task Force for Tuition and Fees - health-science campus student representative
2007 - Present	Alpha Omega Dental Fraternity
2007 - Present	Strength Training
2006 - Present	Student American Dental Association
2006 - Present	Louisville American Student Dental Association
2007	Kosair Children's Hospital Pediatric Dentistry Integration Project
2007	Performer - "A Convocation of Thanks" - Cadaver Appreciation Ceremoney
2007	Microbiology & Immunology Class Representative
2006	Junior Recital - Oboe
2005 - 2006	Mills Music Mission
2005 - 2006	Symphonic Band
2004 - 2006	UTK American Nuclear Society - Member
2004 - 2006	National American Nuclear Society
2003 - 2006	Phi Mu Alpha - Member, Secretary 2005
2002 - 2006	Oboe lessons
2005	Academic Affairs Committee Member (Student Government Association)
May 2005	Study Abroad - Cambridge, UK (Emmanuel College)
	"From Medieval to Cathedral"
2003 - 2005	Wind Ensemble
2004	Pep Band
2003	Piano lessons
2003	Intramural Flag Football
2003	SGA Campaign - Engineering Senator
2002 - 2003	Symphonic Band

Work Experience

2007	DMD tutor - Physiology
2004	Private artists - sanding, grinding, and general shaping of metals
	for large statuary
2002	Smith Travel Research Internship (Hendersonville, TN)

Database Management - Full-time

Computing Experience

Operating Systems:	Windows, UNIX, OS-X
Programming Languages:	C, FORTRAN 77/90
Technical Software:	R, MATLAB, Maple, Mathematica, WinBUGS, SCALE, EViews

Other Skills

Musical Instruments:	Piano - 17 yrs, primarily compositions by Chopin
	Oboe - 14 yrs, 17th through 20th century compositions