A Multi-Comparative Systems Biology Approach Delineates Unique Time-course Differences in Host Immune Pathway Response Infected by *Brucella Melitensis*, *Salmonella Typhimurium*, and *Mycobacterium Avium ssp. Avium*

Kenneth L. Drake[1], Brian Kamery[1], Mark Tong[1], Kelly Stanton[1], Sara D. Lawhon[2], Sangeeta Khare[2], Carlos A. Rossetti[2], and L. Garry Adams[2*]

[1]Seralogix LLC, 3839 Bee Cave Rd, Ste 203, Austin, TX 78746; [2]Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843-4467

Abstract

Systems biology approach to comprehensively study molecular pathway gene expression profiles of microbial pathogens and their host cells holds great promise as a methodology to identity, quantify, analyze, model and predict the overall behavior and dynamics of the host-pathogen immune response. A novel computational approach is introduced based on dynamic Bayesian network (DBN) machine learning that performs comparative pathway analysis to identify the unique and common host response to three different infectious agents.

A bovine ligated ileal loop model in conjunction with a custom bovine and pathogen microarrays were used to study the early temporal-changes in the host response to wild type *S.* Typhimurium (STMWT), *B.* Melitensis (BMEL), and *Mycobacterium avium* subsp. *Avium* (MAA) at several time-points post inoculation. This poster presents preliminary research findings that provide a deeper understanding of the overall complexity of the host defensive processes as well as how the molecular events within the pathogen may lead to the initiation and progression of the infectious disease.

It is believed that cross-comparative analysis of different infectious agent conditions can lead to the identification of novel, unique, as well as common molecular biosignatures. These biosignatures may represent future targets for intervention and could aid in the identification of emerging pathogens and in the development of antimicrobial drugs, immunotherapeutics, vaccines and diagnostics for prevention and treatment of diseases caused by potential agents of bioterrorism and emerging/reemerging infectious diseases.

Introduction

Applying a systems biology approach to temporal multi-conditional comparative analysis is an important technique that enables the identification of unique and common modulations of groups of genes belonging to host defensive or pathogen invasive functions. Traditional methods of finding significantly changed genes and then mapping these genes to known pathways is limited by the chosen statistical test methods and techniques for selecting pathways for analysis based on the given set of significant genes. This traditional approach overlooks interrelated genes (with temporal dependencies) that as a group may have a significant perturbation trend over time, but individually do not meet a predetermined significance threshold. Hence, valuable information can be overlooked and in fact, the more subtle perturbations of interrelated genes may well be just as important, if not more important, then identifying a single gene having several fold change from the control state.

We introduce novel computational approach based on dynamic Bayesian network machine learning that performs comparative pathway analysis to identify the unique and common host response between three different infectious agents. This approach provides a powerful technique to discerning perturbations at the pathway, subnetwork, and individual gene levels. The multi-comparative approach provides a comprehensive systems perspective for both the host and pathogen leading to important insights that may lead to novel mechanistic discoveries important for vaccine and immunotherapeutic development.

Experimental Methods

Gene expression data was collected by our collaborators at Texas A&M College of Veterinary Medicine from a methodology they developed for simultaneously collecting total RNA from the host that included both pathogen and host RNA. For each pathogen condition, a bovine ligated ileal loop model was used to study host response. Under BSL3 condition, there were four non-survival surgeries performed on 3-week old male pathogen-free beef calves for each pathogen experiment. Host RNA was collected and co-hybridized against bovine reference RNA & 13K custom bovine arrays (UIUC) to allow for cross-comparison between experimental conditions. Non-infected ligated ileal loops where used as healthy state controls. Pathogen RNA was enriched & amplified from total RNA samples. The intracellular in vivo pathogen gene expression was compared to the gene expression in the inoculum (cultures of the pathogen at late-log growth phase).

Computational Methods

A novel computational approach is introduced based on dynamic Bayesian network (DBN) machine learning that performs comparative pathway analysis. The computational tools scored and ranked 206 known bovine metabolic and signaling pathways for each condition and 98 pathogen specific pathways. For our novel approach, we developed computational techniques for identifying groups of interrelated genes that as a whole represents the activation (perturbation) of a pathway over the time-course of these experiments (15min, 30min, 1hr, 2hr, 4hr, 8hr, 12hr). We termed this technique "Dynamic Bayesian Gene Group Activation" (DBGGA). DBGGA enables the scoring and ranking of groups of genes across all time points in lieu of just individual genes in a single time point to determine differences between infected and control conditions. The tools that implement DBGGA also automates the process of selecting which pathways within Seralogix's database have genes observed on the microarray sets and hence are included in the scoring process. Pathway descriptions were obtained from publically available resources including KEGG and Biocarta.

For each pathway, a dynamic Bayesian network (DBN) model is constructed to represent known genetic interactions. A DBN is a Bayesian network for dynamic processes and is instantiated via graphical models called directed acyclic graphs (DAG). A simple DBN is shown in Figure 1 to illustrate the causal Bayesian network (1a) and its rollout for three time points creating the DBN model (1b). A node represents a gene expression value and connecting edges within the casual network represent the relationships such as activation or deactivation while the edges connecting the same nodes between the rolled out networks represent the temporal relationships. For example, the Toll-like Receptor pathway will have a known network structure which defines the static causal relationships for our DBN model (1a). This model is trained with the control (non-infected) expression data which determines the DBN parameters (covariances, weights, means) through an Expectation Maximization algorithm. The training datasets are created from preprocessed experimental data that involves global scaling and universal/reference control merging. The infected host gene expression data is then used as evidence to test its "goodness of fit" within the trained control DBN model. By measuring the normalized log likelihood of this infected condition within the control non-infected model, the magnitude of perturbation of this pathway can be scored and ranked in comparison to all the other scored pathways within and between conditions. The ranking of perturbed pathways provides a method of data reduction that focuses attention to those gene groups that are uniquely perturbed pathways associated with the given experimental conditions and also those that are in common.

Host Pathway Results

Table 1 provides a list of the uniquely activated pathways or gene groups ranked in order of perturbation magnitude by time point. The pathway scoring identified significantly activated pathways as early as 30 minutes post exposure for BMEL and MAA while STMWT showed first pathway activations at 60 min post exposure. However, STMWT host response showed stronger trend in pathway activations in later time points in comparison to BMEL and MAA. There were a significant number of metabolic pathway activations along with important immune response signaling transduction pathways observed to be significant in all infectious conditions. As shown in the Venn Diagram of Figure 2, when compared across all time points, we found that STMWT had 31, MAA had 24 and BMEL had 8 uniquely activated pathways while STMWT had 7 in common with MAA and 6 with BMEL, MAA had 3 in common with BMEL.

Figure 3 shows the time course plot of pathway activation scores for a few selected pathways. The plots show the relative activation of pathways in comparison to the three infectious agent conditions within the host. The results show a remarkable uniqueness in the host response to three different pathogens.

Pathogen Pathway Results

The novelty of this experimental comparison is the inclusion of the pathogen gene expression profile extracted from the total host RNA. As with the host pathway analysis, we were able to score the pathogen pathways. At the time of producing this poster, pathogen data was only available for STMWT and BMEL. Each pathogen had 98 pathways in our database. Table 2 provides a list of the significantly activated pathways or gene groups ranked in order of perturbation magnitude by time point. The distinctive pathways/groups which stood out for STMWT were Type III Secretion System and Flagellar Assembly while the BMEL included ABC Transporters and the Two Component System.

Host Mechanistic Gene Discovery

The pathway analyses identifies which host or pathogen processes are being significantly perturbed. From these perturbed pathways we are able to determine which subnetworks are also being perturbed as well as which individual gene is contributing to the model perturbations over time which we define as "mechanistic genes". For each pathway or gene group, mechanistic genes were identified by employing a two state DBN model that is trained with data from both the infected and non-infected conditions. Each model node represents a gene whose model parameters for each condition are determined using a maximum-likelihood criterion using the Expectation Maximization algorithm. The model parameters, which include incoming edge weights, covariance, and means, are used in an algorithm to determine the divergence of the node parameters between study conditions and identifies which nodes (genes) are contributing most to the model divergence. Genes which have a 95% or better confidence level for having significantly different model parameters are labeled as candidate mechanistic genes. A mechanistic gene is identified in the context of its neighboring (parent genes) within a pathway and over the time course of the experiment.

As an example, STMWT modeling results for the MAPK Pathway is shown in Figure 4 where the discovered mechanistic genes are indicated by concentric rings around the node. The graphical model is used to view the state at any time point of all measured genes upstream and downstream of each mechanistic gene candidate. The products of mechanistic genes control key regulatory points in pathways through a variety of methods including altering gene transcription or translation or through post-translation processes such as protein phosphorylation. Subtle alteration of mechanistic genes can induce substantial effects on biological processes.

For this three way comparison, 44 unique mechanistic genes were discovered for STMWT, 94 for MAA, and 132 for the BMEL condition respectively. The Venn diagram of Figure 5, shows the distribution of unique and common candidate mechanistic genes across all conditions. Figure 6 plots a few selected mechanistic genes that were discovered through out systems biology pathway analysis approach.

Similar mechanistic gene analyses was performed on the pathogen pathways. The STMWT analysis discovered 144 candidate mechanistic genes, while for BMEL, there were 42 candidate mechanistic genes.

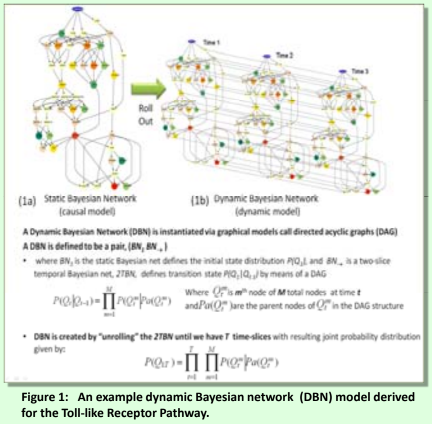

Figure 1: An example dynamic Bayesian network (DBN) model derived for the Toll-like Receptor Pathway.

Table 1: Host Significant Pathway Activation Comparisons

Pathogen	15 min	30 min	60 min	120 min	240 min	480 min
S. Typhimurium (STMWT)	*No Significant Pathways Found*	*No Significant Pathways Found*	• Pentose and glucuronate interconver • ATP synthesis • Purine metabolism • Pyrimidine metabolism • Taurine and hypotaurine metabolism • Glutathione metabolism • Keratan sulfate biosynthesis • Linoleic acid metabolism • DNA polymerase • MAPK signaling pathway • Cytokine-cytokine receptor interaction • Phosphatidylinositol signaling system • mTOR signaling pathway • ECM-receptor interaction • Adherens junction • Gap junction • Complement and coagulation cascades • Antigen processing and presentation • Natural killer cell mediated cytotoxicity • T cell receptor signaling pathway • Leukocyte transendothelial migration • Long-term depression • Taste transduction • GnRH signaling pathway • Lectin Induced Complement Pathway • BRC signaling pathway • B Lymphocyte Cell Surface Molecules	• Pentose and glucuronate interconversions • Linoleic acid metabolism • mTOR signaling pathway	• Oxidative phosphorylation • Linoleic acid metabolism • Cytokine-cytokine receptor interaction • Apoptosis • Toll-like receptor signaling pathway • B cell receptor signaling pathway • Leukocyte transendothelial migration • Circadian rhythm • BRC signaling pathway	• Oxidative phosphorylation • Cysteine metabolism • Linoleic acid metabolism • Ganglioside biosynthesis • Propanoate metabolism • Cell cycle • Regulation of autophagy • Apoptosis • TGF-beta signaling pathway • Toll-like receptor signaling pathway • Hematopoietic cell lineage • Natural killer cell mediated cytotoxicity • T cell receptor signaling pathway • B cell receptor signaling pathway • Leukocyte transendothelial migration • Circadian rhythm • Insulin signaling pathway • Adipocytokine signaling pathway • Epithelial cell signaling in Helicobacter pylori infection • CD40L Signaling Pathway • Trefoil Factors Initiate Mucosal Healing
Mycobacterium avium subsp. *Avium (MAA)*	*Not Sampled*	• Gap junction	• Folate biosynthesis • Protein export • Neuroactive ligand-receptor interaction • Axon guidance • VEGF signaling pathway • Blockade of Neurotransmitter Relase by Botulinum Toxin pathway	• Fructose and mannose metabolism • Purine metabolism • Arginine and proline metabolism • Selenoamino acid metabolism • Glycosylphosphatidylinositol(GPI) • Blood group glycolipid biosynthesis • Stilbene, coumarine and lignin biosynthesis • Alkaloid biosynthesis I • Aminoacyl-tRNA synthetases • MAPK signaling pathway • Neuroactive ligand-receptor interaction • Ubiquitin mediated proteolysis • Axon guidance • Integrin-mediated cell adhesion • Cell adhesion molecules (CAMs) • Complement and coagulation cascades • Antigen processing and presentation • Jak-STAT signaling pathway • Hematopoietic cell lineage • Leukocyte transendothelial migration • Long-term potentiation • Olfactory transduction • Regulation of actin cytoskeleton	• Alkaloid biosynthesis I	• Fatty acid biosynthesis (path 1) • Phenylalanine, tyrosine and tryptophan biosynthesis • Carbon fixation • Neuroactive ligand-receptor interaction • Tight junction • Gap junction • Pathogenic Escherichia coli infection • Cytokine Inflammatory Response
B. Melitensis (BMEL)	*No Significant Pathways Found*	• Oxidative phosphorylation • Ribosome • Wnt signaling pathway • Tight junction • Pathogenic Escherichia coli infection	• Selenoamino acid metabolism • Glycosphingolipid metabolism • Propanoate metabolism • Ribosome • Wnt signaling pathway • B cell receptor signaling pathway • Melanogenesis	• Citrate cycle (TCA cycle) • Fatty acid metabolism • Oxidative phosphorylation • Tryptophan metabolism • Ribosome • Wnt signaling pathway • ECM-receptor interaction • Renin-angiotensin system • Natural killer cell mediated cytotoxicity • B cell receptor signaling pathway • Trefoil Factors Initiate Mucosal Healing	*No Significant Pathways Found*	*Not Sampled*

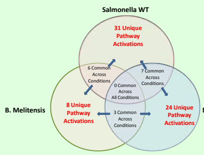

Figure 2: Host Unique and Common Pathway Venn Diagram.

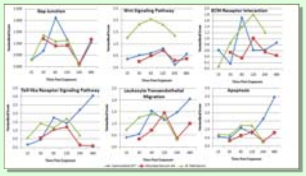

Figure 3: Time Course Pathway Activation Plots for Selected Pathways Significant to Host Response to Pathogen Invasion.

Table 2: Pathogen Significant Pathway Activation Comparisons

Pathogen	15 min	30 min	60 min	120 min	240 min	480 min
S. Typhimurium (STMWT)	• Glycolysis / Gluconeogenesis • Citrate cycle (TCA cycle) • Glycine, serine and threonine metabolism • Phenylalanine, tyrosine and tryptophan biosynthesis • Glycerolipid metabolism • Pyruvate metabolism • Flagellar assembly	*No Significant Pathways Found*	• Glycolysis / Gluconeogenesis • Citrate cycle (TCA cycle) • Arginine and proline metabolism • Phenylalanine, tyrosine and tryptophan biosynthesis • Cyanoamino acid metabolism • Pyruvate metabolism • Glyoxylate and dicarboxylate metabolism • Type III secretion system • Flagellar assembly	*No Significant Pathways Found*	• Glycolysis / Gluconeogenesis • Citrate cycle (TCA cycle) • Phenylalanine, tyrosine and tryptophan biosynthesis • Pyruvate metabolism • Type III secretion system	• Glycolysis / Gluconeogenesis • Citrate cycle (TCA cycle) • Glycine, serine and threonine metabolism • Phenylalanine, tyrosine and tryptophan biosynthesis • Cyanoamino acid metabolism • Glycerolipid metabolism • Sphingolipid metabolism • Bacterial chemotaxis • Type III secretion system • Flagellar assembly
B. Melitensis (BMEL)	• Glycolysis / Gluconeogenesis • Oxidative phosphorylation • Purine metabolism • Glycine, serine and threonine metabolism • ABC transporters • Two component system	• Oxidative phosphorylation • Purine metabolism • Glycine, serine and threonine metabolism • Folate biosynthesis • ABC transporters • Two component system	• Oxidative phosphorylation • Purine metabolism • ABC transporters • Two component system	• Oxidative phosphorylation • ATP synthesis • Glycine, serine and threonine metabolism • ABC transporters • Two component system	• Oxidative phosphorylation • Purine metabolism • Glycine, serine and threonine metabolism • Folate biosynthesis • ABC transporters • Two component system	*Not Sampled*
Mycobacterium avium subsp. *Avium (MAA)*	*Not Sampled at time of Poster Creation*					

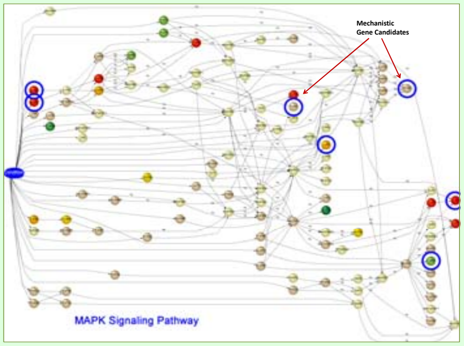

Figure 4: A Visualization Example of the MAPK Pathway Analysis for STMWT Condition at Time=480 minutes. The Blue Concentric Rings Indicate Candidate Mechanistic Genes Discovered Through the Pathway Bayesian Modeling Techniques.

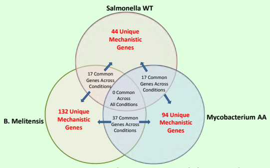

Figure 5: Host Unique and Common Candidate Mechanistic Gene Venn Diagram.

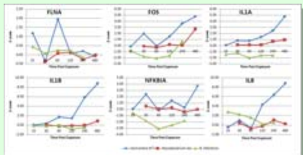

Figure 6: Time Course Plots of a Few Selected Candidate Mechanistic Genes Discovered from DBN Pathway Analysis.

Summary

Preliminary analysis is shown on work in progress in developing the computational techniques to perform multi-conditional comparative analysis. The techniques involve a systems biology approach using dynamic Bayesian network (DBN) modeling as a means to determine pathway or gene group activation. The DBN techniques can be used at various modeling levels that include the subnetworks within pathways and discovering the individual genes that contribute to their activation. We have termed this technique "Dynamic Bayesian Gene Groups Activation" (DBGGA).

The pathways and gene groups that are perturbed should be those that are attributed to the host response to the infectious agents. The cross-comparisons presented here suggest that each pathogen generates remarkably unique pathway activations and gene expression profiles. From these pathways, we plan to develop techniques to construct hypothesized disease models. We have already developed techniques to generate a disease model from pathway analysis from a one-way and two-way comparison as shown in figure 7. Our future work will require the development of techniques to merge data from three-way (or more) comparative studies.

These computational techniques have improved our ability to identify important pathways, genetic mechanisms, and novel genes previously not associated with T. Typhimurium, *Mycobacterium avium* subsp. *Avium* or B. Melitensis infections. The benefits of such analysis will be invaluable for: 1) selecting potential intervening gene targets; 2) guiding future experiments for validating host response models; and 3) identifying dynamic gene expression patterns for early diagnostics.

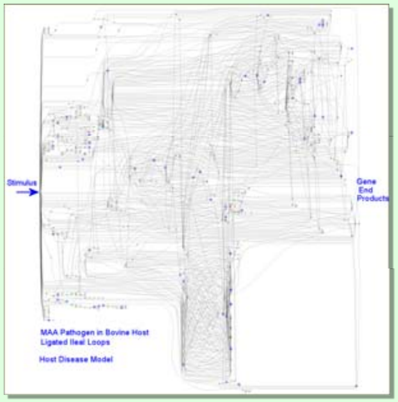

Figure 7 Systems Disease Model for MAA Infection in Bovine Host Derived from Activated Pathways and Candidate Mechanistic Genes.

Seralogix
Contact: Dr. Kenneth Drake (drake@seralogix.com)
3839 Bee Caves Rd, Ste 203, Austin, TX 78746
Phone: 512-828-7955 www.seralogix.com

Acknowledgement: Research was sponsored through a grant from the National Institute of Allergy and Infectious Diseases - R44 AI058362