



Pathway Analysis of Serological Metabolite Profiles in Rheumatoid Arthritis

Ally Perlina, Julie Bryant, Aalim Weljie, Jack Newton, Colin Vitols

March 2009

Metabolomic data sets obtained using Chenomx NMR Suite often require additional analysis to place them in proper context. The GeneGo pathway analysis platform can help establish this context. In this note, we present an example analysis using serological metabolic profiles from Chenomx NMR Suite, known inflammatory markers from prior studies of the K/BxN mouse model and the GeneGo platform to generate a mechanistic hypothesis for the K/BxN mouse model.

Introduction

Rheumatoid arthritis (RA) is a systematic inflammatory disease in which wide assortments of metabolic processes are perturbed. Animal models such as the K/BxN mouse model have proven useful surrogates to build an understanding of the molecular basis of the disease processes. In this application note, we detail a workflow for combining metabolic data obtained from the sera of this K/BxN mouse model with known inflammatory markers to enhance our understanding of the arthritic disease process.

The basic workflow includes the following steps:

1. Use Chenomx NMR Suite to obtain serological metabolite profiles using targeted profiling
2. Integrate known inflammatory markers from previous studies of the mouse model system
3. Use the GeneGo pathway analysis platform to generate a mechanistic hypothesis for the K/BxN mouse model

Study Design and Sample Processing

Serum was collected from transgenic mice with the K/BxN profile (arthritic animals) and control animals with the same genetic background. The study group included 15 arthritic (10 female, 5 male) and 19 control (11 female, 8 male) animals.

The serum samples were treated by ultrafiltration using a 3 kDa MW cutoff filter (Pall Life Sciences) to separate higher molecular weight components from the metabolites of interest [1]. Deuterium oxide (D_2O) was added to the filtrate to give a sample volume of 600 μL and buffer concentration of 100 mM Na_2HPO_4/NaH_2PO_4 at pH 7.0 ± 0.05 . DSS was added as an internal standard at concentrations of 0.25-1.7 mM. $1D^1H$ spectra were acquired on a Bruker AVANCE 600 MHz spectrometer equipped with a TXI probe using the standard noesypr1d pulse sequence.

Acquisition parameters were selected for compatibility with the Chenomx NMR Suite library, as described in [2], and spectra were acquired using between 512 and 2048 transients. 2D TOCSY and HSQC spectra were acquired for three representative samples of each type (arthritic and control) using standard methods. Metabolites were identified and quantified using Chenomx NMR Suite, with identities confirmed using 2D NMR. Quantification was by targeted profiling as implemented in Chenomx NMR Suite [3].

Custom library entries were created for unidentified resonances in order to carry them through the analysis for relative concentration comparison [4].

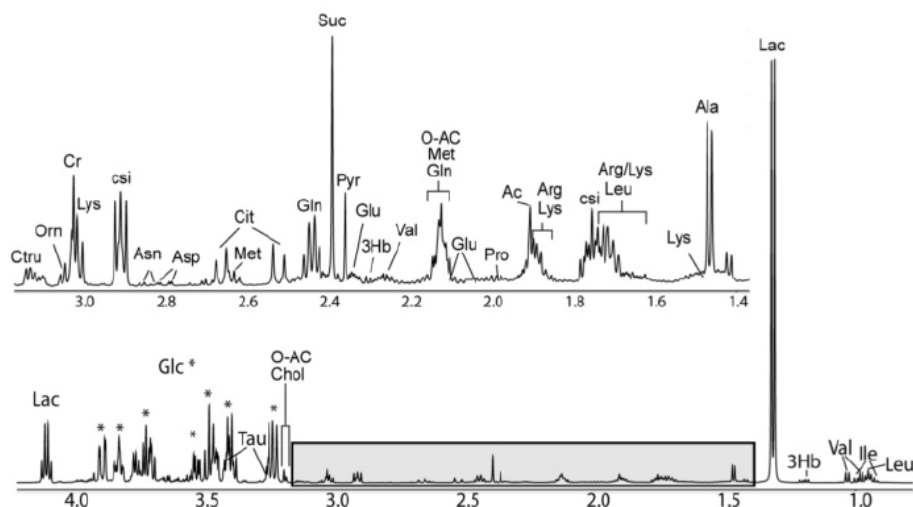


Figure 1. Representative regions of the 1D proton NMR spectrum of the ultrafiltrate from an arthritic mouse. Expansion of the aliphatic region from 0.80 to 4.2 ppm. The upper panel is an expansion of the highlighted region.

Integration of Known Protein Markers

Previous research on this mouse model system has uncovered a number of clues regarding the molecular interactions. T-cells from the arthritic animals recognize a peptide derived from glucose-6-phosphate isomerase (G6PI), a commonly occurring protein central to energy metabolism. In addition, molecular inflammation markers (cytokines) and kinases have been implicated in the response of the arthritic mice. The most important are summarized in Table 1.

Table 1. Protein markers implicated in RA disease process

interleukin-1 (α and β)	always implicated in the disease process
MKK3	always implicated in the disease process
TNF-alpha	moderately implicated, not required for disease onset but marker of increased severity
interleukin-6	not implicated in the disease process

Integration of Data Using GeneGo

Significant metabolites from serum profiling results were uploaded into MetaCore™, a systems biology pathway analysis tool. Ontology enrichment analysis was performed to assess overall functional character of the sample set. Enrichment analysis involves matching an "omics" data set to terms in GeneGo functional ontologies, providing a ranked representation of ontologies that are most saturated or "enriched" with the input data. For this sample set, we used a general enrichment category, GeneGo Biological Processes. This ontology represents prebuilt networks of manually curated protein-protein, protein-compound/metabolite, or protein-nucleic acid interactions, assembled by GeneGo scientific annotators based on proven literature evidence. Each GeneGo Process Network represents a comprehensive biological process with a specific functional theme. The histogram in Figure 2 shows the ten most enriched GeneGo processes for the sample set. The top two networks represent amino acid biosynthesis, while the third network (highlighted in yellow) represents insulin signaling, which may be particularly relevant for this data set. Of the metabolites uploaded into MetaCore, a maximum of 26 compounds were associated with any GeneGo process. Of these, eight mapped to Signal transduction_Insulin signaling, a network representation of which appears in Figure 3.

In the network representations appearing throughout this note, a number of conventions are used. Small molecules appear as purple hexagons, reactions as grey boxes, enzymes as orange shapes, transcription factors as red stars, ligands and extracellular peptides as green shapes, and transporters as purple "X" shapes. Metabolites with available data points are tagged with red or blue circles, indicating positive or negative values, respectively. Arrows denote mechanisms of interaction, green for activation and red for inhibition. Grey interactions represent transport or consumption/production of metabolic intermediates.

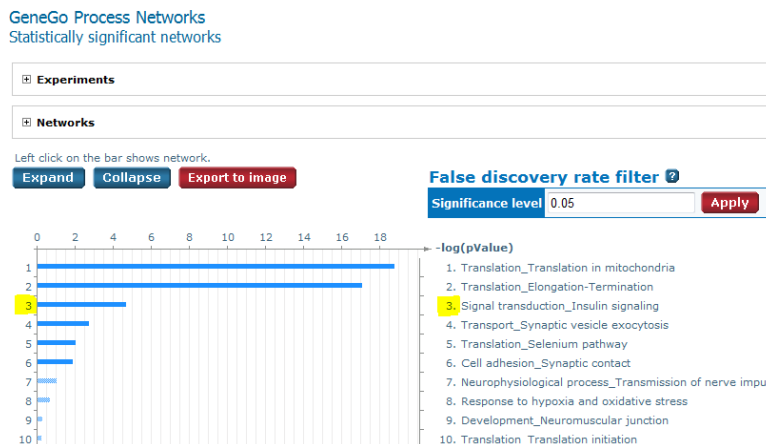


Figure 2. The ten most significantly enriched GeneGo biological processes for the metabolomic data set. The bars represent significance as $-\log(p\text{-value})$ for hypergeometric distribution. All ontology enrichments were filtered to allow no more than 5% false discovery rate.

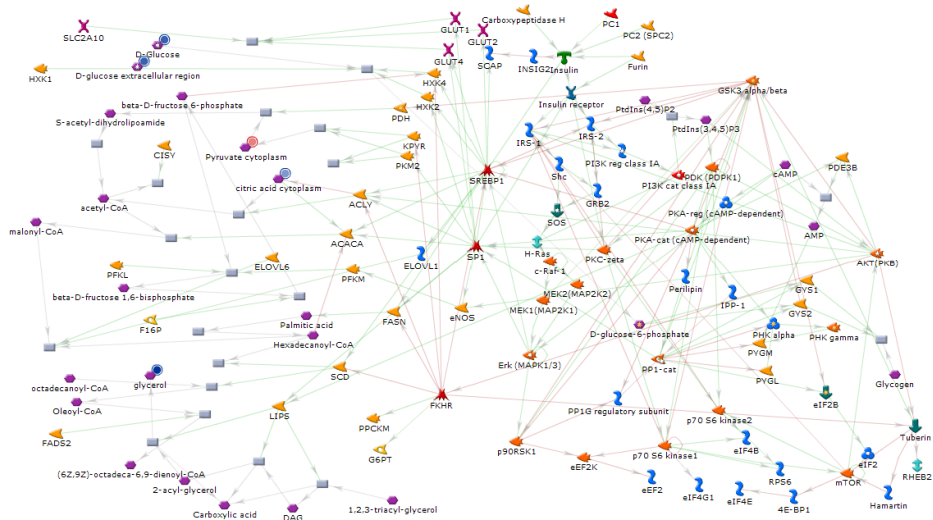


Figure 3. One of the significant GeneGo process networks identified during ontology enrichment analysis was **Signal transduction_Insulin signaling**.

Insulin, in addition to its role in glucose uptake, stimulates cell growth and differentiation. It also promotes the storage of substrates in fat, liver and muscle by stimulating lipogenesis, glycogen and protein synthesis, and inhibiting lipolysis, glycogenolysis and protein breakdown. The stimulatory effects of insulin are mediated through complex intracellular signaling pathways, beginning with ligand binding to insulin receptors and activation of intrinsic receptor tyrosine kinases. Insulin receptor tyrosine kinases phosphorylate a number of cytosolic molecules, including their major substrates, insulin receptor substrate (IRS) proteins.

Tyrosine phosphorylated IRS proteins transmit intracellular signals that mediate growth, metabolic functions, and viability by interacting with downstream molecules that contain SH2 domains, including the growth factor receptor-bound protein 2

(Grb2), SHPTP-2 protein tyrosine phosphatase, and the p85 regulatory subunit of phosphatidylinositol-3 kinase (PI3 kinase). The binding of IRS to Grb2 results in sequential activation of protein kinase ERK (extracellular signal-regulated kinase), which directly contributes to insulin-stimulated mitogenesis, neuritic sprouting, and gene expression.

The indirect implication of mitogen-activated protein kinase (MAPK) signaling for this data set could be inferred from interconnectivity of metabolites and their corresponding enzymes (mostly on the left of the network in Figure 3) with signaling and regulatory portion (mostly to the right in Figure 3) via Sp1 and SREBP1, which serve as key transcriptional hubs on this network. Insulin signaling cascades, Sp1 signaling, and SREBP1 have been implicated in inflammatory response also associated with various MAPK signaling routes, and may reveal further mechanistic details of pathogenesis in RA.

Another ontology, GeneGo Pathway Maps, also shows enrichment (Figure 4). Map folders here particularly demonstrate the higher level of enrichment on the map tree. Each map folder contains numerous maps. The data set contained 43 mappable metabolites, of which ten appeared in the protein degradation folder and six in atherosclerosis.

GeneGo Canonical Pathway Maps represent a prebuilt functional ontology of over 500 active maps that focus only on consensus multistep pathways (canonical pathways) of signaling and/or metabolism. Maps are organized into map folders, and enrichment shows map folder ranking according to the significance of the number of metabolites from the data set mapped to each folder of the provided folder size (total objects/entities).

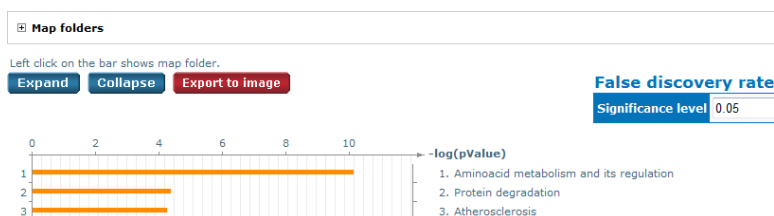


Figure 4. The ten most significantly enriched GeneGo map folders for the metabolomic data set. The bars represent significance as $-\log(p\text{-value})$ for hypergeometric distribution. All ontology enrichments were filtered to allow no more than 5% false discovery rate.

The map folders (Figure 4) tie the active data, which consists only of compounds, with major functional themes like protein degradation and atherosclerosis.

The shortest paths network option in MetaCore connects all given metabolites with the shortest possible path, no more than 3 steps (Figure 5). Shortest paths networks allow determining whether the significant serum metabolites can directly connect to inflammatory factors that may be related to defense response or other disease-related pathways. The network is built dynamically, connecting all objects via direct mechanistic interactions based on manually curated literature evidence. It can then be used for local functional enrichment of the network itself, to associate the interconnected cluster of metabolites and proteins with specific processes or diseases.

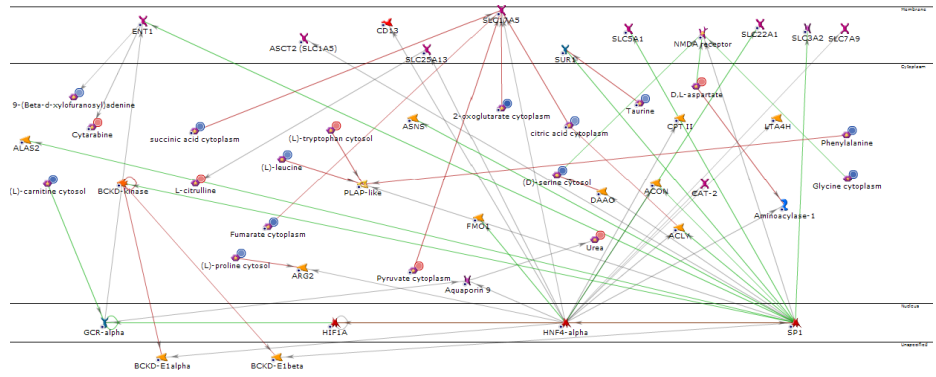


Figure 5. Network of metabolite interactions (unsupervised) related to HNF4-alpha or Sp1 regulation. The network is shown in subcellular localization layout mode with objects arranged from extracellular domain (top) to the nucleus (bottom).

The shortest path network incorporates Sp1 as the key transcription factor hub, followed by HNF4-alpha and HIF1A. This reinforces the implication of possible inflammatory or stress-response processes for the proteins that were in a close network proximity to the metabolites from the RA data set. HIF1A may suggest oxidative stress or general stress response and/or hypoxia, but it is also known to regulate G6PI.

To focus the network approach on the emerging genes of interest and of relevance to disease, we then pursued a more "supervised" network building option, which is often used for novel pathway modeling. This allows drawing networks with specified directionality from or through the genes/proteins of interest to all the significant metabolites. In this manner, more metabolite end points can be connected as potential biomarkers indicating regulatory perturbations of particular signaling mediators.

A key portion of the "supervised" network appears in Figure 6, interconnecting IL1 alpha and beta with HNF4-alpha and HIF1A signaling, which leads to regulation of G6PI and various enzymes via several protein kinases, and then to production of several key significant metabolites. All proteins on this network have been associated with stress response. Other proteins and metabolites were hidden.

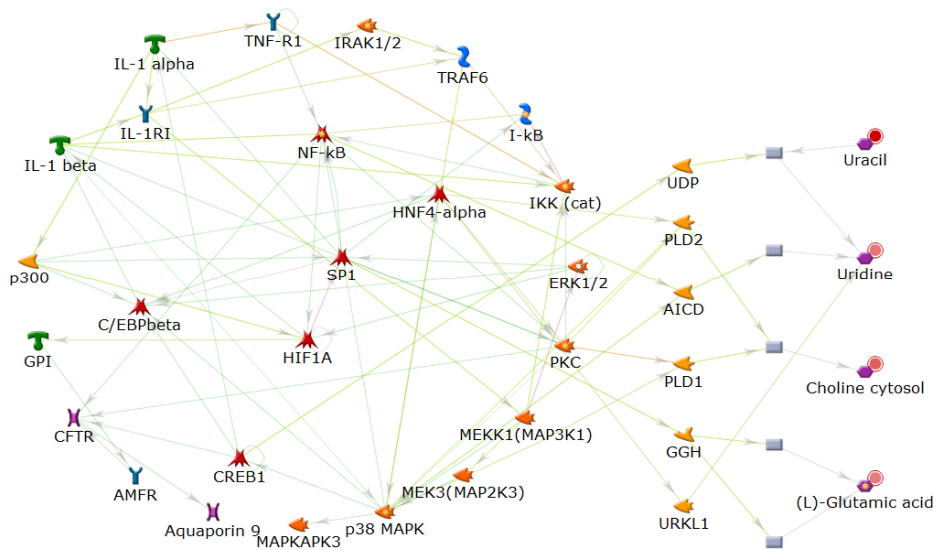


Figure 6. Network of directional (supervised) shortest paths leading to key metabolites.

The functional relevance of the network in Figure 6 is illustrated in Figure 7. The table shows enrichment of proteins for particular gene ontology (GO) terms. Only the checked terms were used to display the remaining objects shown in Figure 6.

#	Check	Process	%	p-Value
1	<input type="checkbox"/>	protein kinase cascade	23.68	3.255e-18
2	<input type="checkbox"/>	positive regulation of cellular metabolic process	26.32	7.515e-18
3	<input type="checkbox"/>	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	23.03	1.765e-17
4	<input type="checkbox"/>	positive regulation of metabolic process	26.32	1.838e-17
5	<input checked="" type="checkbox"/>	response to stimulus	53.95	1.550e-16
6	<input type="checkbox"/>	positive regulation of transcription	21.05	1.223e-15
7	<input type="checkbox"/>	positive regulation of cellular process	36.18	4.509e-15
8	<input type="checkbox"/>	positive regulation of biological process	36.84	1.859e-14
9	<input type="checkbox"/>	lipopolysaccharide-mediated signaling pathway	5.92	5.984e-14
10	<input checked="" type="checkbox"/>	response to stress	32.89	1.126e-13
11	<input type="checkbox"/>	positive regulation of transcription from RNA polymerase II promoter	15.79	1.636e-13
12	<input type="checkbox"/>	positive regulation of transcription, DNA-dependent	17.76	2.995e-13

Figure 7. Top gene ontology (GO) processes for the directional protein-metabolite network. The processes are adopted in MetaCore according to the GO public database, and represent enrichment with corresponding p-values for the network proteins. % refers to the proportion of objects in the network associated with the GO process term.

Novel Findings

Combining the metabolite data profiled by Chenomx with protein data in GeneGo provided a number of possible proteins and pathways of interest. Two particular proteins identified were Sp1 and HNF1. These proteins have been implicated in other inflammation processes [5]. There may also be a direct link between the G6PI peptide, which induces the arthritic processes, and HNF4-alpha action, as the latter protein is a pancreatic regulator of glucose. This relationship has not been previously observed in this system, and serves as a foundation for further experimental validation and possible therapeutic development.

References

- [1] AM Weljie and C Vitols. **June 2006**. *Identifying and Quantifying Metabolites in Blood Serum and Plasma*. Chenomx Inc.
- [2] C Vitols. **2008**. *Chenomx NMR Suite (User Guide)*. Chenomx Inc.
- [3] R Rosewell and C Vitols. **March 2006**. *Identifying Metabolites in Biofluids*. Chenomx Inc.
- [4] AM Weljie, J Newton, FR Jirik, and HJ Vogel. **2008**. *Evaluating Low-Intensity Unknown Signals in Quantitative Proton NMR Mixture Analysis*. *Anal Chem*. 80(23):8956-8965
- [5] R Kekuda, P Saha, and U Sundaram. **2008**. *Role of Sp1 and HNF1 transcription factors in SGLT1 regulation during chronic intestinal inflammation*. *Am J Physiol Gastrointest Liver Physiol*. 294(6):G1354-61