mView™ REPORT

Metabolic Analysis of LH Stimulation in Bovine Corpus Luteum

UNNE-01-10VW

CLIENT: University of Nebraska
John Davis, PhD

AUTHOR: Janice Jones, PhD
APPROVAL: Elizabeth Kensicki, PhD
DATE: November 30, 2012

METABOLON

Metabolon, Inc. • 617 Davis Drive, Suite 400, Durham, NC 277133 • (919) 572-1711 www.metabolon.com
• busdev@metabolon.com

DELIVERABLES

1. Data tables

Data are transferred electronically by Excel spreadsheet which includes the following:

a) Raw data
b) Imputed data
c) Results of statistical tests with associated heat maps, p-values, and q-values.

2. Graphics

Data are provided in box plot image format in a single Excel file. In addition, key findings are provided as PowerPoint slides.

3. Written report

Study description and results, QC information, general platform and statistical information, and specific statistical descriptions are provided in the present report and appendices, in MS Word format.

STUDY DESCRIPTION AND RESULTS

I. Purpose of Experiment

The goal of this study was to characterize the metabolic effects of luteinizing hormone treatment in bovine corpus luteum cells.

II. Experimental design

Global biochemical profiles were compared in bovine corpus luteum cell samples and corresponding media supernatant samples. Cells were treated with luteinizing hormone in a time course experiment, with samples obtained at baseline, 10 minutes, 30 minutes, 60 minutes, and four hours.

Group	*Cells (n)*	*Media (n)*
0m-CTRL	3	3
10m-LH	3	3
30m-LH	3	3
60m-LH	3	3
4h-CTRL	3	3
4h-LH	3	3

III. Summary of Procedure

Metabolon received 18 cell and 18 media samples on October 18, 2012. Following receipt, samples were inventoried, and immediately stored at -80^{o}C. At the time of analysis samples were extracted and prepared for analysis using Metabolon's standard solvent extraction method. The extracted samples were split into equal parts for analysis on the GC/MS and LC/MS/MS platforms. Also included were several technical replicate samples created from a homogeneous pool containing a small amount of all study samples ("Client Matrix"). General platform methods are described in APPENDIX A.

IV. Data Quality: Instrument and Process Variability

QC Sample	*Measurement*	*Median RSD Cells*	*Median RSD Media*
Internal Standards	Instrument Variability	8 %	8 %
Endogenous Biochemicals	Total Process Variability	16 %	16 %

Instrument variability was determined by calculating the median relative standard deviation (RSD) for the internal standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the Client Matrix samples, which are technical replicates of pooled client samples. Values for instrument and process variability meet Metabolon's acceptance criteria as shown in the table above.

V. Metabolite Summary and Significantly Altered Biochemicals

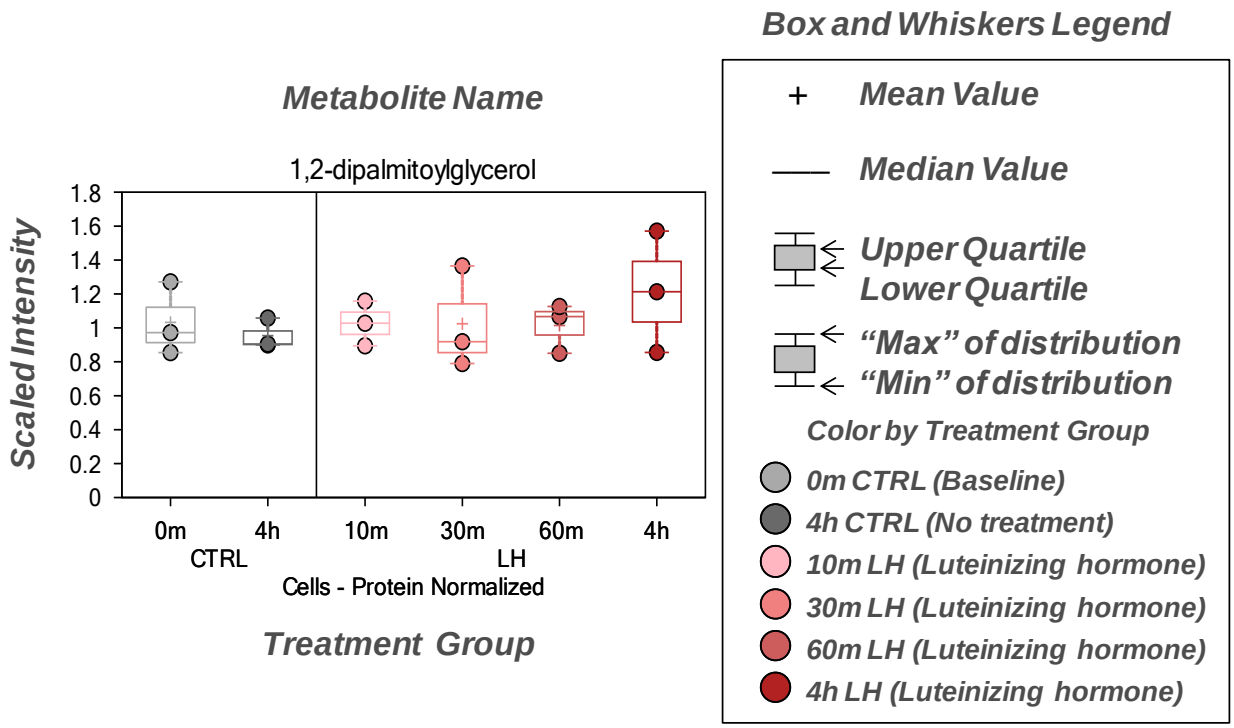
The mView product specification includes all detectable compounds of known identity (named biochemicals). The cells dataset comprises a total of 276 named biochemicals. The media dataset comprises a total of 117 named biochemcials. Following log transformation and imputation with minimum observed values for each compound, Welch's two-sample *t*-tests were used to identify biochemicals that differed significantly between experimental groups. A summary of the numbers of biochemicals that achieved statistical significance ($p \leq 0.05$), as well as those approaching significance ($0.05 < p < 0.1$), is shown below.

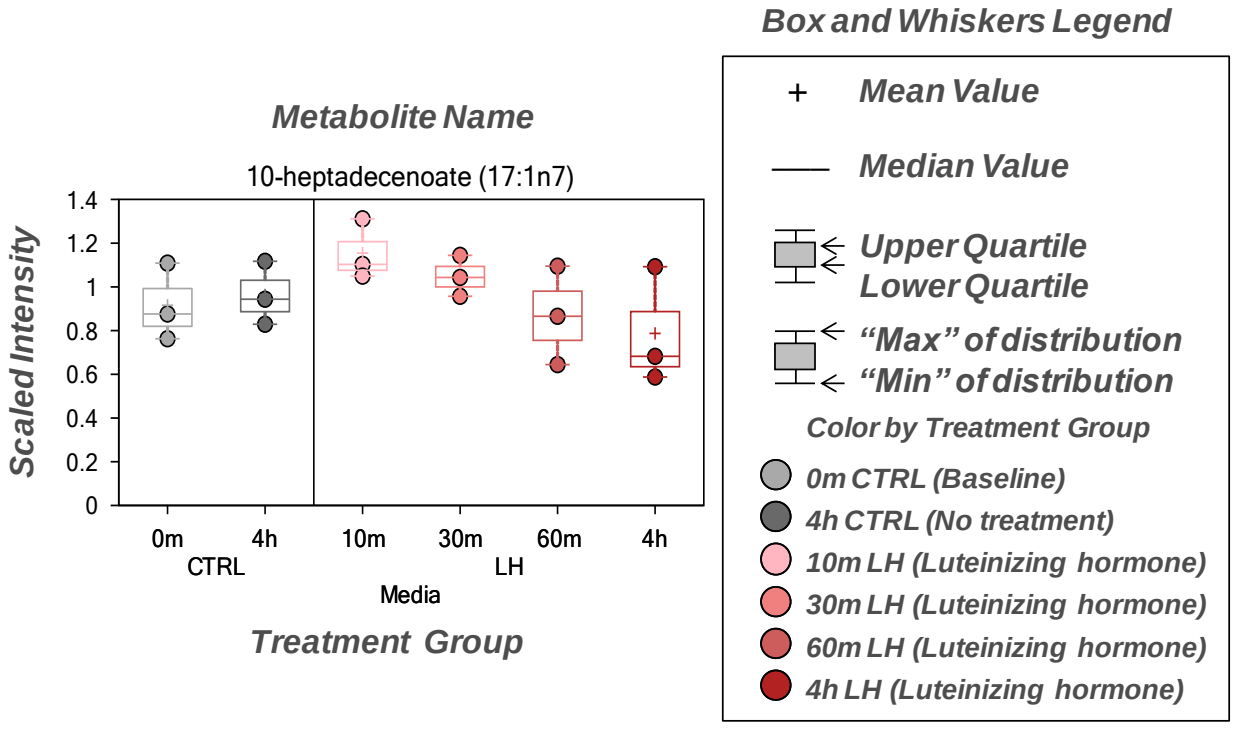
An estimate of the false discovery rate (*q*-value) is calculated to take into account the multiple comparisons that normally occur in metabolomic-based studies. For example, when analyzing 200 compounds, we would expect to see about 10 compounds meeting the $p \leq 0.05$ cut-off by random chance. The *q*-value describes the false discovery rate; a low *q*-value ($q < 0.10$) is an indication of high confidence in a result. While a higher *q*-value indicates diminished confidence, it does not necessarily rule out the significance of a result. Other lines of evidence may be taken into consideration when determining whether a result merits further scrutiny. Such evidence may include a) significance in another dimension of the study, b) inclusion in a common pathway with a highly significant compound, or c) residing in a similar functional biochemical family with other significant compounds. Refer to APPENDIX B for general definitions and further descriptions of false discovery rate and other statistical tests used at Metabolon.

Statistical Comparisons *Cells*				
Welch's Two Sample t-Tests	Total number of biochemicals with $p \leq 0.05$	Biochemicals (↑|↓) p≤0.05	Total number of biochemicals with $0.05 < p < 0.10$	Biochemicals (↑|↓) 0.05<p<0.10
10m-LH **0m-CTRL**	3	3 | 0	2	1 | 1
30m-LH **0m-CTRL**	16	15 | 1	5	2 | 3
60m-LH **0m-CTRL**	10	9 | 1	10	5 | 5
4h-LH **0m-CTRL**	42	27 | 15	20	16 | 4
30m-LH **0m-LH**	5	4 | 1	5	5 | 0
60m-LH **30m-LH**	3	3 | 0	6	2 | 4
4h-LH **60m-LH**	20	14 | 6	17	10 | 7
4h-LH **4h-CTRL**	18	13 | 5	15	12 | 3
4h-CTRL **0m-CTRL**	18	14 | 4	15	8 | 7

Statistical Comparisons *Media*				
Welch's Two Sample t-Tests	Total number of biochemicals with $p \leq 0.05$	Biochemicals (↑|↓) p≤0.05	Total number of biochemicals with $0.05 < p < 0.10$	Biochemicals (↑|↓) 0.05<p<0.10
10m-LH **0m-CTRL**	1	1 | 0	4	3 | 1
30m-LH **0m-CTRL**	7	6 | 1	6	6 | 0
60m-LH **0m-CTRL**	12	11 | 1	8	6 | 2
4h-LH **0m-CTRL**	47	39 | 8	11	5 | 6
30m-LH **10m-LH**	5	5 | 0	4	2 | 2
60m-LH **30m-LH**	1	1 | 0	3	1 | 2
4h-LH **60m-LH**	29	29 | 0	14	9 | 5
4h-LH **4h-CTRL**	19	5 | 14	3	0 | 3
4h-CTRL **0m-CTRL**	46	36 | 10	9	8 | 1

We have also included in the electronic deliverables, a file with data for each biochemical displayed as box plots like that shown in the example figure below.





VI. Biochemical Summary

In females, leutenizing hormone (LH) triggers ovulation and corpus luteum maturation. An acute surge of LH typically starts around day 12 of the human female menstrual cycle and lasts up to 2 days. The corpus luteum itself plays an important role in producing progesterone (and other hormones) to prepare the endometrium for pregnancy, and it triggers feedback loop that propagates the menstrual cycle if fertilization does not occur.

The purpose of this study was to measure metabolic changes in cultured bovine corpus luteum cells treated with LH over a 4-hour time course. Comparison of global biochemical profiles for cells and media identified several metabolic effects of short-term treatment with LH. A few themes from these data are highlighted below to provide initial focus for further investigation.

- **Hormones**:

LH is known to stimulate progesterone production and release from the corpus luteum. In this study, **progesterone** was more abundant in cells and media with LH treatment by the 10 minute time point, and levels continued to increase throughout the 4-hour treatment time course. A progesterone metabolite (**20α-dihydroprogesterone**) known to be made in the corpus luteum was also more abundant with LH and had a particularly large increase in abundance between the 1-hour and 4-hour time points. Hormones and bile acids are generated from cholesterol. **Isocaproic acid**, known to be elevated in cells that generate steroids from cholesterol, was also elevated in cells and media with LH treatment, with the peak observed at the 60 minute time point. Two cholesterol precursors (**squalene** and **lanosterol**) increased in abundance in cells throughout the LH treatment time course suggesting altered regulation of cholesterol biosynthesis with LH. In contrast, the observed decrease in **7β-hydroxycholesterol**, an intermediate between cholesterol and bile acids, suggests less cholesterol utilization for bile acid production or altered uptake from media.

- **LH signaling**:

LH activates the LH receptor, a G protein-coupled receptor that stimulates production of the second messenger cAMP. In this study, **cAMP** production was greatly stimulated with LH, even at the earliest (10 minute) time point.

Prostaglandins are autocrine and paracrine signaling molecules generated from membrane phospholipids in a series of phospholipase, desaturase and elongase reactions. LH is known to stimulate prostaglandin synthesis through increased cyclooxygenase (COX-2) activity. In this time course, two eicosanoid metabolites (**12,13-DHOME** and **14,15-DiHETE**) increased greatly over time in cells and media with control treatment, but increased at a slower rate with LH treatment. 12,13-DHOME is a PPARγ2 activator that has been reported to have toxic effects in tissues, and these results suggest a role for LH in decreasing production of these signaling lipids. LH also had a large effect on production of **lysolipids**, metabolites generated from phospholipase activity toward membrane phospholipids for membrane remodeling, signaling or

membrane catabolism for energy production. Many lysolipids were more abundant in cells (and 1 lysolipid in media) at the 4-hour time point with LH treatment compared with the control treatment at 4 hours. Together these results suggest an effect of LH on phospholipase activity and prostaglandin metabolism that may affect corpus luteum survival and development.

- **Lipid metabolites**:

3-hydroxyisobutyrate, generated from catabolism of branched chain amino acids or fatty acids, was more abundant with time of LH treatment in cells and media. Given that other measured branched chain amino acid metabolites did not differed with LH, altered abundance of this metabolite likely reflects altered lipid metabolism.

Membrane phospholipid catabolites (and osmolytes): Glycerol 3-phosphate and **glycerate**, each possibly generated from phospholipid catabolism, decreased over time in cells treated with LH. Glycerol 3-phosphate had a different pattern in media where it increased greatly between the 1-hour and 4-hour time point with LH treatment. **Glycerophosphocholine**, a cellular osmolyte generated from phosphatidylcholine catabolism, increased in abundance between the 1-hour and 4-hour time points, but this increase was much less apparent in media from LH-treated cells. Additional osmolytes that differed with LH treatment included the sugar alcohols **xylitol** and **sorbitol**, which decreased in abundance in cells at a more accelerated pace with LH treatment. These LH-dependent changes in metabolites used to maintain osmotic balance may relate to effects of LH on corpus luteum size.

- **Central energy metabolism**:

Glucose metabolism: Glucose derived from cellular uptake or glycogen catabolism can be used to generate ATP and reduced electron carriers in glycolysis. In this study, **glucose** and **fructose** were more rapidly depleted from media after LH treatment compared with the control treatment. Likewise, glycogen catabolites (**maltotetraose** and **maltotriose**), as well as **glucose**, were more rapidly depleted in cells after LH treatment. Downstream glycolysis intermediates and end-products of glycolysis (**pyruvate** and **lactate**), however, did not differ significantly with LH treatment, suggesting that the increased glucose taken in from media (and generated from glycogen catabolism) may have been consumed in non-glycolysis reactions.

Glutamine uptake: Glutamine can be catabolized to contribute carbon skeletons to the TCA cycle for energy production. Like glucose, abundance of **glutamine** also decreased in media over time with LH treatment. Also consistent with altered glutamine metabolism with LH treatment, gamma-glutamylglutamine greatly increased over time in control-treated media samples, but did not increase over time in LH-treated samples. Combined with the observed depletion of glucose from media, the depletion of glutamine from media may reflect increased energy expenditure with LH treatment. Additional amino acids depleted from media after LH treatment included **threonine**, **glutamate**, **histidine** and **proline**. These amino acids may be used for energy production, protein translation and/or osmotic balance.

- **Other notable observations**:

  - **Cofactor metabolites:** Coenzyme A (CoA) is a cofactor required for amino acid and lipid metabolism. In this study, **pantothenate**, the precursor to coenzyme A, was depleted from media after LH treatment, but it was not depleted with control treatment, possibly due to increased cellular utilization of this precursor in making CoA. Metabolites of the cofactor NAD+, critical for central energy metabolism, also differed with LH treatment and included NAD+ salvage metabolites **nicotinamide** and **nicotinamide riboside**. Together these differences in cofactor metabolites may reflect effects of LH on cofactor utilization in enzymatic reactions.

  - **Trans-4-hydroxyproline**, a catabolite of collagen, was more abundant in cells over time with LH, but did not increase with time in the control-treated cells. This increase could result from increased collagen synthesis or extracellular matrix remodeling with LH stimulation.

  - **GDP-fucose**, a substrate for glycosyltransferases that add the fucose moiety to proteins, increased more significantly in control cell samples relative to LH-treated samples, suggesting a possible role for LH in regulating protein glycosylation, which would be expected to affect cell-cell communication and adhesion.

Summary

The results of this study identified several metabolites that differed with LH treatment in corpus luteum cells and media. Many of these changes likely reflect altered metabolism in preparation to synthesize large quantities of hormones including progesterone. The hormones themselves increased with LH, as well as bi-products made during hormone biosynthesis. Increased uptake of glucose may be required to provide ATP and/or NADPH required for hormone biosynthesis. Metabolic profiling also identified effects of LH on signaling lipids (PPAR ligands) and (as expected) cAMP. PPAR is known to regulate cholesterol metabolism, and the effects of LH on PPAR ligands may underlie the connection between LH and changes in cholesterol metabolism. Finally, LH affected abundance of 2 metabolites related to extracellular matrix structure, suggesting a role for LH in regulating cell-cell adhesion and communication through modification of (glyco)protein and (glyco)lipid content.

Appendix A: Metabolon Platform

Sample Accessioning: Each sample received was accessioned into the Metabolon LIMS system and was assigned by the LIMS a unique identifier, which was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results *etc.* The samples (and all derived aliquots) were bar-coded and tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80 ºC until processed.

Sample Preparation: The sample preparation process was carried out using the automated MicroLab STAR® system from Hamilton Company. Recovery standards were added prior to the first step in the extraction process for QC purposes. Sample preparation was conducted using a proprietary series of organic and aqueous extractions to remove the protein fraction while allowing maximum recovery of small molecules. The resulting extract was divided into two fractions; one for analysis by LC and one for analysis by GC. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. Each sample was then frozen and dried under vacuum. Samples were then prepared for the appropriate instrument, either LC/MS or GC/MS.

QA/QC: For QA/QC purposes, a number of additional samples are included with each day's analysis. Furthermore, a selection of QC compounds is added to every sample, including those under test. These compounds are carefully chosen so as not to interfere with the measurement of the endogenous compounds. Tables 1 and 2 describe the QC samples and compounds. These QC samples are primarily used to evaluate the process control for each study as well as aiding in the data curation.

Table 1: Description of Metabolon QC Samples

Type	Description	Purpose
MTRX	Large pool of human plasma maintained by Metabolon that has been characterized extensively.	Assure that all aspects of Metabolon process are operating within specifications.
CMTRX	Pool created by taking a small aliquot from every customer sample.	Assess the effect of a non-plasma matrix on the Metabolon process and distinguish biological variability from process variability.
PRCS	Aliquot of ultra-pure water	Process Blank used to assess the contribution to compound signals from the process.
SOLV	Aliquot of solvents used in extraction.	Solvent blank used to segregate contamination sources in the extraction.

Table 2: Metabolon QC Standards

Type	Description	Purpose
DS	Derivatization Standard	Assess variability of derivatization for GC/MS samples.
IS	Internal Standard	Assess variability and performance of instrument.
RS	Recovery Standard	Assess variability and verify performance of extraction and instrumentation.

Liquid chromatography/Mass Spectrometry (LC/MS, LC/MS^2): The LC/MS portion of the platform was based on a Waters ACQUITY UPLC and a Thermo-Finnigan LTQ mass spectrometer, which consisted of an electrospray ionization (ESI) source and linear ion-trap (LIT) mass analyzer. The sample extract was split into two aliquots, dried, then reconstituted in acidic or basic LC-compatible solvents, each of which contained 11 or more injection standards at fixed concentrations. One aliquot was analyzed using acidic positive ion optimized conditions and the other using basic negative ion optimized conditions in two independent injections using separate dedicated columns. Extracts reconstituted in acidic conditions were gradient eluted using water and methanol both containing 0.1% Formic acid, while the basic extracts, which also used water/methanol, contained 6.5mM Ammonium Bicarbonate. The MS analysis alternated between MS and data-dependent MS^2 scans using dynamic exclusion.

Gas chromatography/Mass Spectrometry (GC/MS): The samples destined for GC/MS analysis were re-dried under vacuum desiccation for a minimum of 24 hours prior to being derivatized under dried nitrogen using bistrimethyl-silyl-triflouroacetamide (BSTFA). The GC column was 5% phenyl and the temperature ramp is from 40° to 300° C in a 16 minute period. Samples were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization. The instrument was tuned and calibrated for mass resolution and mass accuracy on a daily basis. The information output from the raw data files was automatically extracted as discussed below.

Accurate Mass Determination and MS/MS fragmentation (LC/MS), (LC/MS/MS): The LC/MS portion of the platform was based on a Waters ACQUITY UPLC and a Thermo-Finnigan LTQ-FT mass spectrometer, which had a linear ion-trap (LIT) front end and a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer backend. For ions with counts greater than 2 million, an accurate mass measurement could be performed. Accurate mass measurements could be made on the parent ion as well as fragments. The typical mass error was less than 5 ppm. Ions with less than two million counts require a greater amount of effort to characterize. Fragmentation spectra (MS/MS) were typically generated in data dependent manner, but if necessary, targeted MS/MS could be employed, such as in the case of lower level signals.

Bioinformatics: The informatics system consisted of four major components, the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software

foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition.

LIMS: The purpose of the Metabolon LIMS system was to enable fully auditable laboratory automation through a secure, easy to use, and highly specialized system. The scope of the Metabolon LIMS system encompasses sample accessioning, sample preparation and instrumental analysis and reporting and advanced data analysis. All of the subsequent software systems are grounded in the LIMS data structures. It has been modified to leverage and interface with the in-house information extraction and data visualization systems, as well as third party instrumentation and data analysis software.

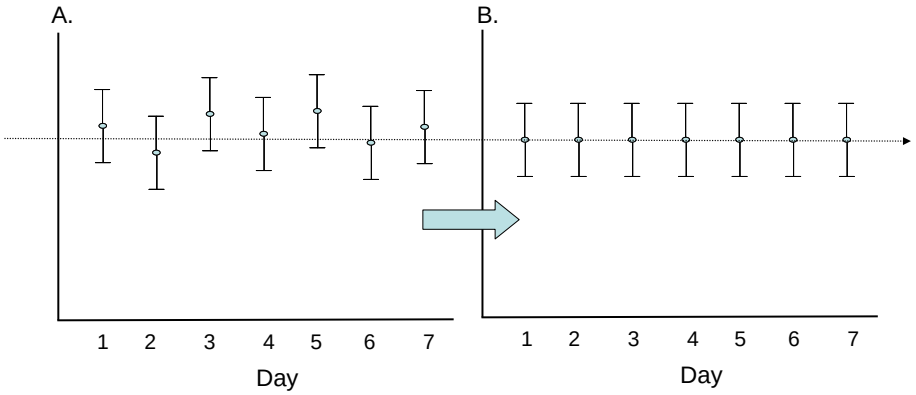
Data Extraction and Quality Assurance: The data extraction of the raw mass spec data files yielded information that could loaded into a relational database and manipulated without resorting to BLOB manipulation. Once in the database the information was examined and appropriate QC limits were imposed. Peaks were identified using Metabolon's proprietary peak integration software, and component parts were stored in a separate and specifically designed complex data structure.

Compound identification: Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Identification of known chemical entities was based on comparison to metabolomic library entries of purified standards. As of this writing, more than 1000 commercially available purified standard compounds had been acquired registered into LIMS for distribution to both the LC and GC platforms for determination of their analytical characteristics. The combination of chromatographic properties and mass spectra gave an indication of a match to the specific compound or an isobaric entity. Additional entities could be identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Curation: A variety of curation procedures were carried out to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise.
Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Normalization: For studies spanning multiple days, a data normalization step was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one (1.00) and normalizing each data point proportionately (termed the "block correction"; Figure 1). For studies that did not require more than one day of analysis, no normalization is necessary, other than for purposes of data visualization.

Figure 1: Visualization of Data Normalization



Statistical Calculation: For many studies, two types of statistical analysis are usually performed: (1) significance tests and (2) classification analysis. (1) For pair-wise comparisons we typically perform Welch's t-tests and/or Wilcoxon's rank sum tests. For other statistical designs we may perform various ANOVA procedures (e.g., repeated measures ANOVA). (2) For classification we mainly use random forest analyses. Random forests give an estimate of how well we can classify *individuals* in a *new* data set into each group, in contrast to a t-test, which tests whether the unknown means for two populations are different or not. Random forests create a set of classification trees based on continual sampling of the experimental units and compounds. Then each observation is classified based on the majority votes from all the classification trees. Statistical analyses are performed with the program "R" http://cran.r-project.org/.

Appendix B: Statistical Terminology

***t*-tests:** *t*-tests test whether the unknown means for two populations are different or not. The *p*-value gives the amount evidence that the population means are different based on the data (through the *t*-statistic). The smaller the *p*-value, the more evidence we have that the population means are different. Often, a significance level of 0.05 is used. When the *p*-value is less than 0.05, we have enough evidence to conclude that the population means are different ("statistical significance"). The level of 0.05 is the false positive rate. This means that 5% of the time, the *t*-test would incorrectly conclude the population means are different when they are the actually the same.

***q*-values**: The level of 0.05 is the false positive rate when there is one test. However, for a large number of tests we need to account for false positives. If the data were simply random noise, approximately 5% of the *p*-values would be less than 0.05, 10% of the *p*-values would be less than 0.10, etc. Thus, even if the data were only random noise, we would get approximately 10 "significant" results out of 200 compounds when the false positive rate is 0.05.

There are different methods to correct for multiple testing. The oldest methods are family-wise error rate adjustments (Bonferroni, Tukey, etc.), but these tend to be extremely conservative for a very large number of tests. With gene arrays, using the False Discovery Rate (FDR) is more common. The family-wise error rate adjustments give one a high degree of confidence that there are *zero* false discoveries. However, with FDR methods, one can allow for a small number of false discoveries. The FDR for a given set of compounds can be estimated using the *q*-value (see Storey J and Tibshirani R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**: 9440-9445).

In order to interpret the *q*-value, first sort the data by the *p*-value then choose the cutoff for significance (typically $p<0.05$). The *q*-value gives the false discovery rate for the selected list (i.e., an estimate of the proportion of false discoveries for the list of compounds whose *p*-value is below the cutoff for significance). For Table 1 below, if the whole list is declared significant, then the false discovery rate is approximately 10%. If everything from Compound 079 and above is declared significant, then the false discovery rate is approximately 2.5%.

Table 1: Example of *q*-value interpretation

Compound	*p*-value	*q*-value
Compound 103	0.0002	0.0122
Compound 212	0.0004	0.0122
Compound 076	0.0004	0.0122
Compound 002	0.0005	0.0122
Compound 168	0.0006	0.0122
Compound 079	0.0016	0.0258
Compound 113	0.0052	0.0631
Compound 050	0.0053	0.0631
Compound 098	0.0061	0.0647
Compound 267	0.0098	0.0939

Random Forest: Random forest is a supervised classification technique based on an ensemble of decision trees (see Breiman. 2001. *Machine Learning*. 45:5, for the original description; Goldstein et al. 2010. *BMC Genetics*. 11:49, for additional information). For a given decision tree, a random subset of the data with identifying true class information is selected to build the tree ("bootstrap sample" or "training set"), and then the remaining data, the "out-of-bag" (OOB) variables, are passed down the tree to obtain a class prediction for each sample. This process is repeated thousands of times to produce the forest. The final classification of each sample is determined by computing the class prediction frequency ("votes") for the OOB variables over the whole forest. For example, suppose the random forest consists of 50,000 trees and that 25,000 trees had a prediction for sample 1. Of these 25,000, suppose 15,000 trees classified the sample as belonging to Group A and the remaining 10,000 classified it as belonging to Group B. Then the votes are 0.6 for Group A and 0.4 for Group B, and hence the final classification is Group A. This method is unbiased since the prediction for each sample is based on trees built from a subset of samples that do not include that sample. When the full forest is grown, the class predictions are compared to the true classes, generating the "OOB error rate" as a measure of prediction accuracy. Thus, the prediction accuracy is an unbiased estimate of how well one can predict sample class in a new data set.

To determine which variables (biochemicals) make the largest contribution to the classification, a "variable importance" measure is computed. We use the "Mean Decrease Accuracy" (MDA) as this metric. The MDA is determined by randomly permuting a variable, running the observed values through the trees, and then reassessing the prediction accuracy. If a variable is not important, then this procedure will have little change in the accuracy of the class prediction (permuting random noise will give random noise). By contrast, if a variable is important to the classification, the prediction accuracy will drop after such a permutation, which we record as the MDA. Thus, the random forest analysis provides an "importance" rank ordering of biochemicals; we typically output the top 30 biochemicals in the list as potentially worthy of further investigation.