|  |
| --- |
|  |
| mViewTM **REPORT** |
| **C. difficile in mouse cecal contents**  **MICH-02-16VW** |
| CLIENT: University of Michigan  Matthew Jenior  AUTHOR: Gregory R. Wagner, PhD  APPROVAL: Edward D. Karoly, PhD  DATE: December 12, 2016 |
|  |





Metabolon, Inc. • 617 Davis Drive, Suite 400, Durham, NC 27713 • (919) 572-1711 www.metabolon.com • Contact: [info@metabolon.com](mailto:info@metabolon.com)

Table of Contents

[**Objective** 3](#_Toc448326797)

[**Experimental Procedures** 3](#_Toc448326798)

[**Results and Biological Interpretation** 4](#_Toc448326799)

[Metabolite Summary and Significantly Altered Biochemicals 4](#_Toc448326800)

[Biological Interpretation 6](#_Toc448326801)

[**Conclusions** 9](#_Toc448326802)

[**Study Parameters** 10](#_Toc448326803)

[Data Quality: Instrument and Process Variability 10](#_Toc448326804)

[**Appendix** 11](#_Toc448326805)

[Metabolon Platform 11](#_Toc448326806)

[Statistical Methods and Terminology 15](#_Toc448326808)

# **Objective**

***Purpose of Experiment***

The goal of this study was to compare the metabolomic profiles in cecal contents from non-infected (NI) mice and from mice infected with *Clostridium difficile* (C. diff), both with and without various antibiotic treatments.

# **Experimental Procedures**

***Experimental design***

This 81 sample study examined global metabolic profiles in mouse cecal contents from non-infected mice and mice infected with C. difficile. Mice were received no treatment or antibiotic treatment with Cefoperazone, Clindamycin, or Streptomycin. Included as controls are germ-free mice, either uninfected or infected (no antibiotic). Global metabolic profiles were determined from the experimental groups outlined in the table below.

|  |  |  |
| --- | --- | --- |
| **Group** | **Description** | **n** |
| Germ-Free NI, No TX | Cecal contents from germ-free mice, with no infection or treatment | 9 |
| Germ-Free C. diff, No TX | Cecal contents from germ-free mice, with C. diff infection and no treatment | 9 |
| Conv NI, No TX | Cecal contents from conventional mice, with no infection or treatment | 9 |
| Conv NI, Cef | Cecal contents from conventional mice, with no infection and treated with Cefoperazone | 9 |
| Conv C. diff, Cef | Cecal contents from conventional mice, with C. diff infection and treated with Cefoperazone | 9 |
| Conv NI, Clinda | Cecal contents from conventional mice, with no infection and treated with Clindamycin | 9 |
| Conv C. diff, Clinda | Cecal contents from conventional mice, with C. diff infection and treated with Clindamycin | 9 |
| Conv NI, Strep | Cecal contents from conventional mice, with no infection and treated with Streptomycin | 9 |
| Conv C. diff, Strep | Cecal contents from conventional mice, with C. diff infection and treated with Streptomycin | 9 |

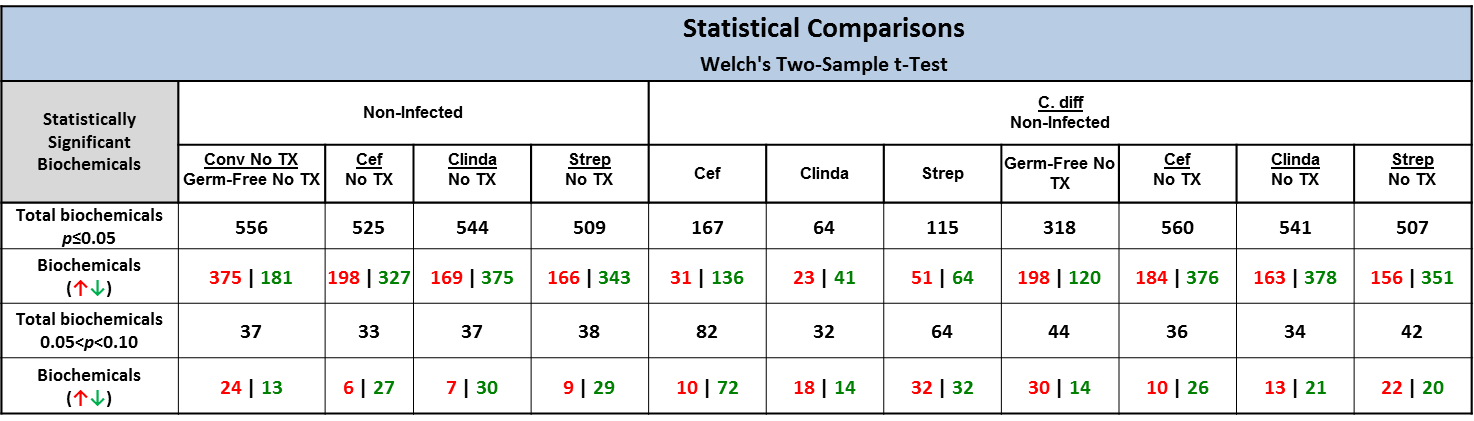
# **Results and Biological Interpretation**

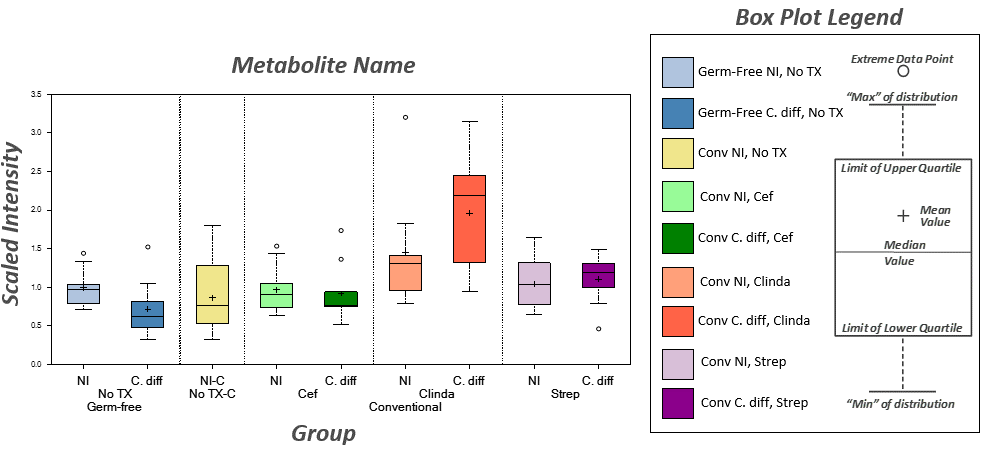
## 

## Metabolite Summary and Significantly Altered Biochemicals

The present dataset comprises a total of 727 compounds of known identity (named biochemicals). Following log transformation and imputation of missing values, if any, with the minimum observed value for each compound, Welch’s two-sample *t*-test was used to identify biochemicals that differed significantly between experimental groups. A summary of the numbers of biochemicals that achieved statistical significance (*p*≤0.05), as well as those approaching significance (0.05<*p*<0.10), 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*≤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 the Appendix for general definitions and further descriptions of false discovery rate and other statistical tests used at Metabolon.



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.

## 

## Biological Interpretation

*Clostridium difficile* is a bacterium that can infect the human gastrointestinal tract and cause diarrhea, abdominal pain and, sometimes, life-threatening complications including dilation or perforation of the colon and sepsis. *C. difficile* infection (CDI) often occurs following the clearance of competing gut microbial communities with antibiotic (Abx) treatment. Recent evidence suggests that Abx-induced alterations in the gut microbiome generate a metabolic environment favoring the establishment and growth of *C. difficile* in mice (PMID: 24445449). However, different Abx may produce different microbial ecosystems that are more or less susceptible to colonization by *C. difficile*. The purpose of this study was to identify metabolic indicators and pathway shifts in mouse cecal contents during CDI and treatment with three different Abx. The provided Excel data table (CDT) includes the raw data for each sample and the accompanying statistical analysis results.

This report focuses primarily on changes in carbohydrates and sugar alchohols as directed by the PI, but additional changes of interest are discussed. Datasets provided in the mView product can be large and contain a great deal of information. To provide an initial focus for further consideration, a few observations are offered below from a cursory view of the data. *These are not presented as a comprehensive analysis; the PI, with a much greater knowledge of the experimental system, is encouraged to make a detailed study of the data for additional or alternative interpretations.*  For the discussion below, please refer to graphical illustrations in the accompanying PowerPoint file, or to the statistical tables, box plots, and other graphics found in the accompanying Excel workbook. *Note that an identical mass-equivalent of cecal contents was extracted for the platform; no additional normalization was applied prior to statistical analyses.*

**Principle Component and Statistical Overview (slides 5):** Principal component analysis (PCA) is a mathematical procedure that can be used to obtain a high-level view of the structure of a dataset. Briefly, PCA permits visualization of how individual samples in a dataset differ from each other. Samples with similar biochemical profiles cluster together whereas samples with different biochemical profiles segregate from one another. As such, this analysis tool aids in determining if the different experimental groups can be segregated based on differences in their overall metabolic signature. In this study, samples within experimental groups clustered closely together, suggesting little variation in their global metabolomic signatures (**slide 5**). The conventionally-housed, no Abx, and no infection group of samples was well-separated from the remaining sample groups, suggesting this group displays a distinct global metabolic profile relative to the remaining groups (orange circle for comparison). The Abx-treated samples clustered closely together by PCA, irrespective of the particular Abx used and CDI status, suggesting Abx treatment induced similar changes in the global cecal metabolic profile (black circle for comparison) (**slide 5**). Consistently, the germ-free groups without Abx treatment clustered closely together and were well-separated from the Abx-treated and conventionally housed groups, indicating that germ-free and Abx treatment strongly differentiate cecal metabolomic signatures.

**Abx-related Changes in Microbiome-derived Metabolites (slide 6)**: Abx can have profound effects on the composition of the animal microbiome as well as the levels of biochemicals generated by the microbiome. In validation of the Abx treatment, the Abx-treated groups (both infected and non-infected with *C. diff*) samples showed low to undetectable levels of multiple secondary bile acids including **7-ketodeoxycholate**, **deoxycholate**, and **6-beta-hydroxylithocholate**, among many others. Furthermore, the tyrosine/benzoate-derived microbial metabolites **2-(4-hydroxyphenyl)propionate, 3-(4-hydroxyphenyl)propionate,** and3-phenylpropionatewere significantly reduced in Abx-treated groups. Conversely, many of these same biochemicals were significantly increased in conventionally-housed mice relative to germ-free mice, consistent with an intact microbiota in conventionally-housed mice. As expected, these data suggest that Abx-treated animals show significant changes in microbiome-derived biochemicals relative to untreated animals.

**Abx-induced Changes in Carbohydrates and Sugar Alcohols (slide 7):** Abx-treatment induced decreases in several carbohydrates relative to untreated mice including **glucose**, **ribose**, **arabinose**, **fucose**, and **mannose**. In contrast, many other carbohydrates, polysaccharides, and sugar alcohols were significantly increased by Abx-treatment. These included **ribitol**, **arabitol/xylitol**, **arabinate/xylonate**, **maltotetraose**, **maltotriose**, **lactobionate**, **sucrose**, **galactitol**, **raffinose**, and **galactonate**. These Abx-mediated changes in major carbon sources may provide a more growth-permissive cecal environment for *C. difficile.*

**Abx-Specific Differences in Carbohydrates and Sugar Alcohols (slide 8):** Previous work has shown that treating mice with the antibiotic cefoperazone leads to increases in mannitol/sorbitol in the mouse gut and that *C. difficile* can grow on this defined carbon source (PMID: 24445449). Consistent with these earlier findings, cefoperazone treatment was associated with a ~74-fold increase in **mannitol/sorbitol** in cecal contents relative to untreated mice, but showed no change in the clindamycin or streptomycin treated mice. This may suggest that cefoperazone induces a more growth-permissive substrate/environment for *C. difficile* relative to the other Abx, however there were no changes in mannitol/sorbitol levels during CDI relative to mice treated with Abx alone, suggesting *C. difficile* may prefer a different carbon source. A similar trend was observed for **galactitol** with cefoperazone treatment. **Sucrose** was elevated more so in the clindamycin-treated group relative to the cefoperazone- and streptomycin-treated groups, but underwent further increases during CDI, possibly suggesting *C. difficile* does not utilize sucrose for growth in these conditions. *C. difficile* can also grow on **stachyose** (PMID: 24445449), which was dramatically elevated in the clindamycin-treated group as compared to the other Abx. However, there was little change in stachyose levels during CDI relative to mice treated with Abx alone, suggesting *C. difficile* may prefer a different carbon source in these conditions.

**CDI-induced changes in Carbohydrates and Sugar Alcohols (slide 9):** Upon CDI, several carbohydrate and sugar alcohols were significantly decreased relative to Abx treatment alone. Importantly, some of these biochemicals were also significantly increased by Abx treatment relative to untreated animals (see earlier section), suggesting that these carbohydrates and sugar alcohols are elevated in cecal contents by Abx treatment and may be preferentially consumed by *C. difficile* during infection. The metabolites decreasing during CDI included **ribitol**, **maltopentaose**, and **verbascose** in the cefoperazone-treated group and diacetylchitobiose in the streptomycin group. Strikingly, the pentose sugar **arabonate/xylonate** was significantly reduced during CDI regardless of Abx treatment after undergoing ~100-fold increases with Abx alone. The similar metabolite **arabitol/xylitol** was significantly reduced during CDI with clindamycin and streptomycin after ~25-fold increases with Abx alone. These data indicate that arabonate/xylonate and arabitol/xylitol are increased with Abx *and* suggest that these metabolites may be preferentially utilized by *C. difficile.*  Ribulose/xylulose was not substantially altered by Abx, but was also significantly decreased during CDI in the cefoperazone and clindamycin groups, suggesting *C. difficile* may also preferentially assimilate this carbon source.

**Additional Changes of Interest (slide 10):** Many dipeptides exhibited significant decreases with Abx treatment. One notable exception to this trend was **pro-hydroxy-pro**, which was dramatically increased with Abx and, furthermore, was significantly decreased during CDI, suggesting that *C. difficile* may influence the levels of this metabolite. Several N-acetyl-amino acids were significantly elevated with Abx, but decreased with CDI including N-acetylthreonine, N-acetyl-aspartyl-glutamate, N-acetylhistidine, N-acetylphenylalanine, N-acetyltryptophan, and N-acetylleucine among others, again suggesting that Abx and CDI have opposing influences on the levels of these metabolites and that they may be an important source of amino acids for the growth of *C. difficile*. Other metabolites showing similar trends included the nucleotide metabolite N2, N2-demethylguanine, N-acetylglucosaminylasparagine and the cholesterol synthesis intermediates **mevalonate** and **mevalonolactone** (not with streptomycin)which are typically utilized in anabolic/growth processes. In contrast, the bacterial lysine degradation produce 5-aminovalerate was significantly decreased with Abx, and then dramatically increased upon CDI, which may simply reflect the degradation of lysine by *C. difficile* as it becomes established in the gut*.* Abx induced significant decreases in **fatty acids** as a class, some of which were further altered (primarily decreased) during CDI. Importantly, it has been reported that short-chain fatty acids in particular inhibit the growth and decrease toxin production in *C. difficile* (PMID: 7839098)and based on this, it was hypothesized that the production of SCFAs by the indigenous gut microbiota is a major contributor to innate resistance to *C. difficile*. Notably, the microbiota-derived SCFA valerate was ~40-fold higher in the animals without Abx.

# **Conclusions**

## A high-level view of the cecal contents data through principle component analysis indicates a high degree of sample clustering within experimental groups and strong segregation between groups according to antibiotic and germ-free status, generally suggesting that antibiotics and housing conditions elicit differences in global cecal metabolomic signatures. In validation of the experimental system, antibiotic treatment led to a marked reduction in microbiome-derived biochemicals in cecal contents. Notably, antibiotics induced significant shifts in the cecal composition of carbohydrates and sugar alcohols, the levels of which were previously implicated in creating a metabolic ecosystem that is more conducive to the establishment and growth of C. difficile. There were several antibiotic-specific differences in carbohydrates and sugar alcohols including large increases in mannitol/sorbitol and galactitol with cefoperazone treatment. In addition, C. difficile infection led to significant decreases in arabitol/xylitol, arabonate/xylonate, and ribulose/xylulose, suggesting C. difficile may preferentially utilize these biochemicals for growth. Additional changes of interest include dramatic decreases in fatty acids with antibiotic treatment and biochemicals responsive to C. difficile infection including pro-hydroxy-pro, N-acetyl amino acids and cholesterol intermediates.

As a path forward, it may be useful to test the growth of *C. difficile* in defined medium in which the metabolites changing with antibiotics or *C. difficile* infection are manipulated (e.g., medium without arabonate/xylonate). Testing the metabolomic changes after treating with additional antibiotics and correlating these changes with the growth of *C. difficile* may provide further insight into the particular antibiotic classes and metabolic conditions that affect the establishment and growth of this opportunistic pathogen.

## 

# **Study Parameters**

## 

## Data Quality: Instrument and Process Variability

|  |  |  |
| --- | --- | --- |
| ***QC Sample*** | ***Measurement*** | ***Median RSD*** |
| Internal Standards | Instrument Variability | 3% |
| Endogenous Biochemicals | Total Process Variability | 14% |

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.

# **Appendix**

## 

## Metabolon Platform

**Sample Accessioning:** Following receipt, samples were inventoried and immediately stored at -80oC. Each sample received was accessioned into the Metabolon LIMS system and was assigned by the LIMS a unique identifier that 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 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 -80oC until processed.

**Sample Preparation:**Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

**QA/QC:** Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Tables 1 and 2 describe these QC samples and standards. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the 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 pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections, as outlined in Figure 1.

**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 the 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** |
| RS | Recovery Standard | Assess variability and verify performance of extraction and instrumentation. |
| IS | Internal Standard | Assess variability and performance of instrument. |

****

**Figure 1.**  **Preparation of client-specific technical replicates.** A small aliquot of each client sample (colored cylinders) is pooled to create a CMTRX technical replicate sample (multi-colored cylinder), which is then injected periodically throughout the platform run. Variability among consistently detected biochemicals can be used to calculate an estimate of overall process and platform variability.

**Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS):** All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSn scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

**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 Compound Identification:** Raw data was extracted, peak-identified and QC processed using Metabolon’s hardware and software. These systems are built on a web-service platform utilizing Microsoft’s .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z)*, and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been 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.

**Metabolite Quantification and Data Normalization:** Peaks were quantified using area-under-the-curve. 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 2). For studies that did not require more than one day of analysis, no normalization is necessary, other than for purposes of data visualization. In certain instances, biochemical data may have been normalized to an additional factor (e.g., cell counts, total protein as determined by Bradford assay, osmolality, etc.) to account for differences in metabolite levels due to differences in the amount of material present in each sample.

**Figure 2: Visualization of data normalization steps for a multiday platform run.**

## Statistical Methods and Terminology

**Statistical Calculations:** For many studies, two types of statistical analysis are usually performed: (1) significance tests and (2) classification analysis. Standard statistical analyses are performed in ArrayStudio on log transformed data. For those analyses not standard in ArrayStudio, the programs R (<http://cran.r-project.org/>) or JMP are used. Below are examples of frequently employed significance tests and classification methods followed by a discussion of p- and q-value significance thresholds.

1. **Welch’s two-sample *t*-test**

Welch’s two-sample *t*-test is used to test whether two unknown means are different from two independent populations.

This version of the two-sample *t*-test allows for unequal variances (variance is the square of the standard deviation) and has an *approximate* *t*-distribution with degrees of freedom estimated using Satterthwaite’s approximation. The test statistic is given by *t* , and the degrees of freedom is given by , where , are the sample means, s1, s2, are the sample standard deviations, and n1, n2 are the samples sizes from groups 1 and 2, respectively. We typically use a two-sided test (tests whether the means are different) as opposed to a one-sided test (tests whether one mean is greater than the other).

1. **Matched pairs *t*-test**

The matched pairs *t*-test is used to test whether two unknown means are different from paired observations taken on the same subjects.

The matched pairs *t*-test is equivalent to the one-sample *t*-test performed on the differences of the observations taken on each subject (i.e., calculate (x1 – x2) for each subject; test whether the mean difference is zero or not). The test statistic is given by , with *n* – 1 degrees of freedom, where , are the sample means for groups 1 and 2, respectively, sd is the standard deviation of the differences, *n* is the number of *subjects* (so there are 2*n* observations).

1. **One-way ANOVA**

ANOVA stands for analysis of variance. For ANOVA, it is assumed that all populations have the same variances. One-way ANOVA is used to test whether at least two unknown means are all equal or whether at least one pair of means is different. For the case of two means, ANOVA gives the same result as a two-sided *t*-test with a pooled estimate of the variance.

An ANOVA uses an F-test which has two parameters – the numerator degrees of freedom and the denominator degrees of freedom. The degrees of freedom in the numerator are equal to *g* – 1, where *g* is the number of groups. If *n* is the total number of observations (*n*1 + *n*2), then, the denominator degrees of freedom is equal to *n* – *g*. The F-statistic is the ratio of the between-groups variance to the within-groups variance, hence the higher the F-statistic the more evidence we have that the means are different.

Often within ANOVA, one performs linear contrasts for specific comparisons of interest. For example, suppose we have three groups A, B, C, then examples of some contrasts are A vs. B, the average of A and B vs. C, etc. For single-degree of freedom contrasts, these give the same result as a two-sided *t*-test with the pooled estimate of the variance from the ANOVA and degrees of freedom *n* – *g*. Below, we show the three formulas for A vs. B from a three group design as shown above. The numerator is same in each case, but the denominator differs by the estimates of the variances, and the degrees of freedom are different for each (if the theoretical assumptions hold, then the contrast has the most power, as it has the largest degrees of freedom).

Welch’s two-sample *t*-test

By *t* , and the degrees of freedom is given by

Two-sample *t*-test with pooled estimate of variance from A and B

where , where the degrees of freedom is *nA* + *nB* – 2.

The contrast from the ANOVA,

where , where the degrees of freedom is given by where the degrees of freedom is *nA* + *nB* + *nC* – 3.

1. **Two-way ANOVA**

ANOVA stands for analysis of variance. For ANOVA, it is assumed that all populations have the same variances. For a two-way ANOVA, three statistical tests are typically performed: the main effect of each factor and the interaction. Suppose we have two factors A and B, where A represent the genotype and B represent the diet in a mouse study. Suppose each of these factors has two levels (A: wild type, knock out; B: standard diet, high fat diet). For this example, there are 4 combinations (“treatments”): A1B1, A1B2, A2B1, A2B2. The overall ANOVA F-test gives the p-value for testing whether all four of these means are equal or whether at least one pair is different. However, we are also interested in the effect of the genotype and diet. A main effect is a contrast that tests one factor across the levels of the other factor. Hence the A main effect compares (A1B1 + A1B2)/2 vs. (A2B1 + A2B2)/2, and the B-main effect compares (A1B1 + A2B2)/2 vs. (A1B2 + A2B2)/2. The interaction is a contrast that tests whether the mean difference for one factor depends on the level of the other factor, which is (A1B2 + A2B1)/2 vs. (A1B1 + A2B2)/2.

Some sample plots follow. For the first plot, there is a B main effect, but no A main effect and no interaction, as the effect of B does not depend on the level of A. For the second plot, notice how the mean difference for B is the same at each level of A and the difference in A is the same for each level of B, hence there is no statistical interaction. The final plot also has main effects for A and B, but here also has an interaction: we see the effect of B depends on the level of A (0 for A1 but 2 for A2), i.e., the effect of the diet depends on the genotype. We also see here the interpretation of the main effects depends on whether there is an interaction or not.







1. **Two-way Repeated Measures ANOVA**

This is typically an ANOVA where one factor is applied to each subject and the second factor is a time point. See two-way ANOVA as many of the details are similar except that the model takes into account the repeated measures, i.e., the treatments are given to the same subject over time. The two main effects and the interaction are assessed, with particular interest to the interaction, as this shows where the time profiles are parallel or not for the treatments (parallel mean no interaction).

One additional note, the standard analysis assumes a condition referred to as compound symmetry, which assumes the correlation between each pair of levels of the repeated-measures factor is the same. Thus, for the case of time, it assumes the correlation is the same between time points 1 and 2, 1 and 3, and 2 and 3.

1. **Correlation**

Correlation measures the strength and direction of a *linear* association between two variables. The statistical test for correlation tests whether the true correlation is zero or not.

The square of the correlation is the percentage of the total variation explained by a linear relationship between the two variables. Thus, with large sample sizes there may be a sample correlation of 0.1 that is statistically significant. This means we have high confidence that the true correlation is zero, however, only 100\*(0.1\*0.1)% = 1% of the variation of one variable is explained by a linear relationship with the other variable, so while there is an association, it has little predictive ability.

1. **Hotelling’s T2 test**

The Hotelling’s T2 test is a multivariate generalization of the *t*-test, but here we are testing whether the mean vectors are different or not (the vector consists of multiple metabolites).

The Hotelling statistic is: , where *nx* and *n*y are the numbers of samples in each group, is the mean vector of the variables from group 1, is the mean vector of variables from group 2 and **S** is the pooled estimate of the variance-covariance matrix of the variables. This analysis assumes the underlying variance-covariance matrix is the same for each group. Notice that in the case of uncorrelated variables, this is simply a weighted average of the squared mean differences with weights inversely proportional to the sample variances (i.e., the metabolites less variable within a group are given higher weights).

1. **p- values**

For statistical significance testing, p-values are given. The lower the p-value, the more evidence we have that the null hypothesis (typically that two population means are equal) is not true. If “statistical significance” is declared for p-values less than 0.05, then 5% of the time we incorrectly conclude the means are different, when actually they are the same.

The p-value is the probability that the test statistic is at least as extreme as observed in this experiment given that the null hypothesis is true. Hence, the more extreme the statistic, the lower the p-value and the more evidence the data gives against the null hypothesis.

1. **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. 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; PMID: 12883005).

In order to interpret the q-value, the data must first be sorted 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



1. **Random Forest**

Random forest is a supervised classification technique based on an ensemble of decision trees (see Breiman L. (2001) Random Forests.  Machine Learning.  45: 5-32; http://link.springer.com/article/10.1023%2FA%3A1010933404324).  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.  Random forest has several advantages – it makes no parametric assumptions, variable selection is not needed, it does not overfit, it is invariant to transformation, and it is fairly easy to implement with R.

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.

1. **Hierarchical Clustering**

Hierarchical clustering is an unsupervised method for clustering the data, and can show large-scale differences. There are several types of hierarchical clustering and many distance metrics that can be used. A common method is complete clustering using the Euclidean distance, where each sample is a vector with all of the metabolite values. The differences seen in the cluster may be unrelated to the treatment groups or study design.

1. **Principal Components Analysis (PCA)**

Principal components analysis is an unsupervised analysis that reduces the dimension of the data. Each principal component is a linear combination of every metabolite and the principal components are uncorrelated. The number of principal components is equal to the number of observations.

The first principal component is computed by determining the coefficients of the metabolites that maximizes the variance of the linear combination. The second component finds the coefficients that maximize the variance with the condition that the second component is orthogonal to the first. The third component is orthogonal to the first two components and so on. The total variance is defined as the sum of the variances of the predicted values of each component (the variance is the square of the standard deviation), and for each component, the proportion of the total variance is computed. For example, if the standard deviation of the predicted values of the first principal component is 0.4 and the total variance = 1, then 100\*0.4\*0.4/1 = 16% of the total variance is explained by the first component. Since this is an unsupervised method, the main components may be unrelated to the treatment groups, and the “separation” does not give an estimate of the true predictive ability.

1. **Z-scores**

An intensity measurement for a metabolite by itself does not tell much. If for example a patient contains a blood glucose level of 300, this could be very good news if most people have blood glucose levels around 300, but less so if most people have levels around 100. In other words a measurement is meaningful only relative to the means of the sample or the population. This can be achieved by transforming the measurements into Z-scores which are expressed as standard deviations from the mean.

The Z-score, also called the standard score or normal score, is a dimensionless quantity derived by subtracting the control population mean from an individual raw score and then dividing the difference by the control population standard deviation. The Z-score indicates how many standard deviations an observation is above or below the mean of the control group. The Z-score is negative when the raw score is below the mean, positive when above. Since knowing the true mean and standard deviation of a control population is often unrealistic, the mean and standard deviation of the control population may be estimated using a random control sample.

 z = \frac{x - \mu}{\sigma}

Z-score =

where: x is a raw score to be standardized, μ is the mean of the control population, σ is the standard deviation of the control population

Subtracting the mean *centers* the distribution, and dividing by the standard deviation *standardizes* the distribution. The interesting properties of Z-scores are that they have a zero mean (effect of “centering”) and a variance and standard deviation of 1 (effect of “standardizing”). This is because all distributions expressed in Z-scores have the same mean (0) and the same variance (1), so we can use Z-scores to compare observations coming from different distributions. When a distribution is normal most of the Z-scores (more than 99%) lay between the values of -3 and +3.