|  |
| --- |
|  |
| mViewTM **FINAL REPORT** |
| **Microbial depletion and ozone exposure;  murine lung tissue and serum**  **HARV-04-14VW** |
| CLIENT: Harvard School of Public Health  CLIENT: Helen Youngji Cho  AUTHOR: Nicki Bell, PhD  APPROVAL: Edward D. Karoly, PhD  DATE: March 24, 2015 |
|  |





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](#_Toc414973046)

[**Experimental Procedures** 3](#_Toc414973047)

[**Results and Biological Interpretation** 3](#_Toc414973048)

[Metabolite Summary and Significantly Altered Biochemicals 3](#_Toc414973049)

[Biological Interpretation 6](#_Toc414973050)

[**Conclusions** 9](#_Toc414973051)

[**Study Parameters** 10](#_Toc414973052)

[Data Quality: Instrument and Process Variability 10](#_Toc414973053)

[**Appendix** 11](#_Toc414973054)

[Metabolon Platform 11](#_Toc414973055)

[Statistical Methods and Terminology 16](#_Toc414973056)

# **Objective**

***Purpose of Experiment***

The goal of this study was to interrogate biochemical profiles in murine lung tissue and serum originating from mice treated with an antibiotic cocktail and untreated control mice prior to exposure to either ambient air or ozone, with the aim of investigating the metabolic consequence of ozone exposure following attenuation of native microbiota.

# **Experimental Procedures**

***Experimental design***

Global biochemical profiles were determined in lung tissue and serum collected from untreated control mice and mice treated for two weeks with untreated drinking water or water containing an antibiotic cocktail (ampicillin, neomycin, metronidazole, and vancomycin), followed by a three hour exposure to ambient air or ozone (2ppm). Sample collection occurred 24 hours post-ozone exposure.

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **n** | | **Description** |
| **Lung Tissue** | **Serum** |
| H2O Air | 8 | 8 | Water treated, air exposed |
| Abx Air | 8 | 8 | Antibiotic treated, air exposed |
| H2O Ozone | 8 | 8 | Water treated, ozone exposed |
| Abx Ozone | 8 | 8 | Antibiotic treated, ozone exposed |

# **Results and Biological Interpretation**

## Metabolite Summary and Significantly Altered Biochemicals

The present datasets comprise a total of 499 of named biochemicals (compounds of known identity) in lung tissue and 562 named biochemicals in serum. Following log transformation and imputation of missing values, if any, with the minimum observed value for each compound, ANOVA contrasts were 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. Analysis by two-way ANOVA identified biochemicals exhibiting significant interaction and main effects for experimental parameters of ozone exposure and antibiotic treatment.

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 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

The purpose of this study was to examine metabolic interactions between microbial depletion with antibiotics and exposure to ozone in lung tissue and serum from mice. In particular, mice were treated with a cocktail of antibiotics (ampicillin, neomycin, metronidazole, and vancomycin) in their drinking water for two weeks, and were subsequently exposed to room air or ozone (2 ppm for 3 hours). After 24 hours, lung tissue and serum from mice given H2O (control; n=8) or antibiotics (Abx; n=8) and subjected to air (control; n=8) or ozone (n=8) was collected for metabolomic study. Individual lung tissue and serum samples were loaded by equivalent weights/volumes across the platform with no additional normalization performed prior to statistical analyses.

Datasets provided in the mView product can be quite large and contain a great deal of information. To provide an initial focus for further consideration, a few observations are provided below from a cursory view of the data. Separation of samples by principal component analysis (PCA) is presented first, followed by a discussion of metabolic pathways that were differentially altered in lung tissue and/or serum upon exposure to ozone in mice treated with H2O or antibiotics.

* **Clear metabolic shifts associated with exposure to Abx and/or ozone were apparent in both lung tissue and serum samples.** Application of PCA to determine separation of study groups demonstrated that the lung tissue and serum metabolic profiles of mice exposed to Abx and/or ozone were quite distinct. In this analysis, a large number of metabolic variables were transformed into a smaller number of orthogonal variables (i.e., Comp. 1, Comp. 2) in order to analyze variation between the groups and populations that differ are expected to group separately. In particular, a metabolic shift reflective of Abx intake was clearly present in lung tissue and serum regardless of exposure to ozone, and treatment with ozone induced further metabolic changes that resulted in clear separation of groups in lung tissue and serum samples. Results from the PCA are in agreement with the large numbers of statistically significant biochemical changes that were observed when comparing across study groups. It is important to note that one lung tissue sample (client identifier 27) exhibited distinct separation from all other samples; however, this sample was not deemed a true outlier according to the criteria (must show extreme values for at least 35-40% of metabolites) set forth by Metabolon statisticians and was therefore retained within the dataset for the purposes of statistical analysis.
* **Bile acids and potential hepatotoxicity.** One of the most consistent and marked changes in the dataset involved accumulation of numerous **primary** and **secondary bile acids** in lung tissue and serum from mice treated with Abx and exposed to ozone. Primary bile acids are synthesized and may be conjugated to glycine or taurine in the liver and are secreted into the small intestine upon food intake where they assist with emulsion and uptake of dietary lipids and fat-soluble vitamins. A large portion of bile acids are generally reabsorbed from the small intestine and returned to the liver via enterohepatic circulation, where they undergo hepatic extraction from portal blood. Primary bile acids that remain in the intestinal tract are typically deconjugated and decarboxylated by intestinal bacteria, resulting in production of secondary bile acids that may be reabsorbed or excreted in the feces. Due to their potentially toxic properties, bile acids are removed from the circulation by the liver with high efficiency; however, in the setting of hepatic dysfunction this process is disrupted and serum bile acids increase. General elevations in primary and secondary bile acids in mice treated with Abx and exposed to ozone may suggest that these two factors have a synergistic effect on hepatic dysfunction, resulting in accumulation of these metabolites in serum. Accumulation of bile acids in lung tissue may simply result from elevated levels in the circulation, with the alternative possibility that gastroesophageal reflux (GER) and/or increased expression of the enzyme machinery responsible for *de novo* bile acid synthesis within pulmonary vascular endothelial cells may also contribute to elevated bile acids in lung tissue from mice treated with Abx and exposed to ozone. These findings highlight an important interaction between treatment with Abx and ozone and suggest that hepatic dysfunction may be a consequence of these two factors.
* **Oxidative stress and the gamma-glutamyl cycle.** Gamma-glutamyl amino acids may be formed as part of the gamma-glutamyl cycle associated with activity of gamma-glutamyl transpeptidase (GGT), a metabolic pathway that is important for recycling and regeneration of the potent antioxidant glutathione. Consistent elevations in nearly all **gamma-glutamyl amino acids** measured in lung tissue were observed in mice treated with Abx and exposed to ozone, suggestive of increased activity of the gamma-glutamyl cycle in these animals. Indeed, GGT activity has been characterized in lung epithelium and surfactant as an important contributor to glutathione metabolism. While increased glutathione demand is expected to support detoxification and elimination of antibiotics, pronounced elevations in gamma-glutamyl amino acids were only observed in mice treated with Abx that were also exposed to ozone. These findings are suggestive of a synergistic effect of these two treatments on gamma-glutamyl cycle activity and may also be related to oxidative stress induced by ozone treatment in lung tissue. Although changes in additional metabolic indicators of oxidative stress were generally mixed in lung tissue from mice treated with Abx and exposed to ozone, an interesting and consistent reduction in several markers of an oxidative environment were observed in serum from these animals. In particular, reductions in **dimethylarginine (SDMA + ADMA), oxidized glutathione (GSSG),** the lipid peroxidation marker **13-HODE + 9-HODE,** the autoxidation product of docosahexaenoic acid (DHA) **17-HDoHe**, and the diol formed from hydrolysis of the 9,10 epoxide of linoleic acid **9,10-diHOME** were limited to mice treated with Abx and exposed to ozone. These changes were accompanied by a reduction in **alpha-tocopherol (vitamin E)**, which may suggest that increased utilization of this antioxidant plays an important role in the reduction in oxidative stress observed in these mice. While inflammation resulting from oxidative stress is thought to contribute to negative effects associated with exposure to ozone, various reports in the literature have also demonstrated that ozone therapy may act in an anti-oxidative and anti-inflammatory manner under certain conditions. Thus, findings from the current study suggest that reductions in systemic oxidative stress may result from concurrent exposure to Abx and ozone in mice, whereas the status of the oxidative environment in lung tissue as reflected by the metabolome in these animals is less clear.
* **Differential ozone-induced alterations in the serum free fatty acid profile.** One of the most striking differential changes observed when comparing study groups involved pronounced reductions in numerous **medium-chain** and **long-chain saturated, monounsaturated,** and **polyunsaturated free fatty acids** in mice treated with Abx and exposed to ozone, while these metabolites were generally unchanged or increased in mice treated with Abx but not exposed to ozone. Changes in lipid-related metabolites in lung tissue were generally consistent across groups and therefore do not appear to play a major role in the free fatty acid perturbations observed in serum. Despite consistent reductions in free fatty acids, including dicarboxylic and hydroxylated free fatty acids, in mice treated with Abx and exposed to ozone, circulating levels of metabolites reflective of lipid metabolism [i.e., long-chain acylcarnitines such as **palmitoylcarnitine** and **stearoylcarnitine** and the ketone body **3-hydroxybutyrate (BHBA)**] did not differ substantially across study groups. Long-chain free fatty acids must be conjugated with carnitine to facilitate crossing of mitochondrial membranes and entry into the mitochondrial matrix in order to undergo subsequent β-oxidation, and ketones such as BHBA are synthesized from excess acetyl-CoA that generally results from increased lipid metabolism in a process that typically occurs in the liver. Thus, despite the pronounced reductions in circulating free fatty acids induced by treatment with Abx and exposure to ozone, “downstream” pathways of lipid metabolism are not affected by this treatment. An alternative explanation for the changes in free fatty acids may involve alterations in lipolysis, as the biochemical marker of this process, **glycerol**, was concomitantly reduced in mice treated with Abx and exposed to ozone. Thus, the interaction between Abx treatment and ozone exposure may result in decreased release of free fatty acids from storage depots such as adipose tissue, giving rise to the pronounced reductions in circulating free fatty acids observed in mice subjected to these treatment conditions. Although ozone has been reported to induce changes in lipid metabolism through pathways related to induction of pro-oxidative and pro-inflammatory lipid mediators (i.e., prostaglandins and eicosanoids), these metabolites were not different when comparing across study groups in the current study. Taken together, these findings provide novel insight into changes in whole-body lipid handling induced by Abx treatment and ozone exposure and suggest that further study will be required to determine the precise pathways related to lipid metabolism that are affected in these animals.
* **Polyamine metabolism.** Polyamines are small nitrogen-containing compounds that play an integral yet poorly understood role in cellular physiology. Among other functions, these metabolites bind to and stabilize RNA and DNA to control gene expression programs, and are essential for cellular proliferation. Importantly, alterations in polyamine metabolism following exposure to ozone have been documented in the literature, and published studies indicate that these biochemicals may be protective against detrimental effects of ozone in plants and heart tissue. In the current study, elevations in the circulating polyamines **spermine** and **spermidine** and the related molecule **acisoga** were observed in serum samples from mice exposed to ozone in the absence of Abx, but increased polyamine levels were attenuated in mice treated with Abx and exposed to ozone. Although polyamines were measured in lung tissue, a differential pattern of change was not readily apparent as a function of treatment with Abx and/or exposure to ozone. Therefore, results from this study suggest that a potentially protective elevation in polyamines induced by exposure to ozone may be absent in mice treated with Abx, although this effect appears to be systemic and is not specific to lung tissue.

# **Conclusions**

In conclusion, results from this global profiling study demonstrated pronounced metabolic changes in lung tissue and serum resulting from treatment with Abx and/or exposure to ozone in mice. In particular, treatment with both Abx and ozone appears to have a synergistic effect with regard to increased hepatotoxicity and changes in the oxidative environment, induction of differential changes in lipid metabolism (possibly through altered lipolysis), and alterations in systemic polyamine metabolism that may decrease protection from negative effects of ozone exposure. Importantly, select results from the current study suggest that ozone treatment may not consistently induce pro-oxidative and pro-inflammatory effects as may be expected from this strong oxidizing agent. Follow-up studies investigating long-term exposure to ozone subsequent to treatment with Abx may reveal differential changes in lung tissue and serum metabolic profiles as compared to the short-term exposure paradigm tested in the current study.

# **Study Parameters**

## Data Quality: Instrument and Process Variability

|  |  |  |  |
| --- | --- | --- | --- |
| ***QC Sample*** | ***Measurement*** | ***Median RSD*** | |
| ***Lung*** | ***Serum*** |
| Internal Standards | Instrument Variability | 5 % | 6 % |
| Endogenous Biochemicals | Total Process Variability | 9 % | 8 % |

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 as shown in the table above meet Metabolon’s acceptance criteria.

# **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. A recovery standard was 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: one for analysis by UPLC-MS/MS with positive ion mode electrospray ionization, one for analysis by UPLC-MS/MS with negative ion mode electrospray ionization, one for LC polar platform, one for analysis by GC-MS, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. For LC, the samples were stored overnight under nitrogen before preparation for analysis. For GC, each sample was dried under vacuum overnight 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. |
| DS | Derivatization Standard | Assess variability of derivatization for GC-MS samples. |
| 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):** The LC/MS portion of the platform was based on 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 acidic or basic LC-compatible solvents, each of which contained 8 or more injection standards at fixed concentrations to ensure injection and chromatographic consistency. 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 (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm). Extracts reconstituted in acidic conditions were gradient eluted from a C18 column using water and methanol containing 0.1% formic acid. The basic extracts were similarly eluted from C18 using methanol and water, however with 6.5mM Ammonium Bicarbonate. The third 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. The MS analysis alternated between MS and data-dependent MS2 scans using dynamic exclusion, and the scan range was from 80-1000 m/z. Raw data files are archived and extracted as described below.

**Gas Chromatography-Mass Spectroscopy (GC-MS):**  The samples destined for analysis by GC-MS were dried under vacuum for a minimum of 18 h prior to being derivatized under dried nitrogen using bistrimethyl-silyltrifluoroacetamide.  Derivatized samples were separated on a 5% diphenyl / 95% dimethyl polysiloxane fused silica column (20 m x 0.18 mm ID; 0.18 um film thickness) with helium as carrier gas and a temperature ramp from 60° to 340°C in a 17.5 min period.  Samples were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization (EI) and operated at unit mass resolving power. The scan range was from 50–750 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 +/- 0.005 amu, 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 distribution to both the LC-MS and GC-MS 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.