mList+™ REPORT

Diabetes study

WCQA-02-12ML+

CLIENT: Weill Cornell Medical College in Qatar
Karsten Suhre, PhD

AUTHOR: Edward D. Karoly , PhD
APPROVAL: Lining Guo, PhD
DATE: August 31, 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 that includes the following:

a) **Raw Data**: Raw ion counts.

b) **Original Scale Data**: Raw data normalized to account for inter-day differences.

c) **Scaled Imputed Data**: Original Scale Data for each biochemical is rescaled to set the median equal to 1. Missing values, generally due to sample falling below the limit of detection, if present, are then imputed with the minimum observed value.

d) **Variable Normalized Imputed Data**: Original Scale Data normalized by another variable such as osmolality, protein or DNA concentration.

2. Written report

Study description and results, QC information, general platform 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 biochemically profile human plasma, urine and saliva samples in a blinded study. These results will be analyzed by the client with the objective of identifying metabolic markers of diabetes in each of the three matrices.

II. Experimental design / Summary of Procedure

Metabolon received 1052 samples on July 19, 2012.The total number of samples submitted for analysis under this study for each matrix is presented below.

Matrix	Sample n	Description
Plasma	359	Blinded plasma samples
Urine	360	Blinded urine samples
Saliva	333	Blinded saliva samples

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.

III. Data Quality: Instrument and Process Variability

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 below.

QC Sample	*Measurement*	*Median RSD*		
		Plasma	*Urine*	*Saliva*
Internal Standards	Instrument Variability	6 %	7 %	5 %
Endogenous Biochemicals	Total Process Variability	13 %	9 %	12 %

IV. Metabolite Summary and Significantly Altered Biochemicals

Numbers of compounds of known structural identity (named biochemicals) as well as compounds of unknown structural identity (unnamed biochemicals) detected in the present plasma, urine and saliva datasets are presented in the table below. Information related to biochemical pathways, chemical properties, and public databases are included in the electronic data files that accompany this report.

Matrix	*Platform Days*	*Named Biochemicals*	*Unnamed Biochemicals*	*Total Biochemicals*
Plasma	11	404	354	758
Urine	11	436	455	894
Saliva	10	374	228	602

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.

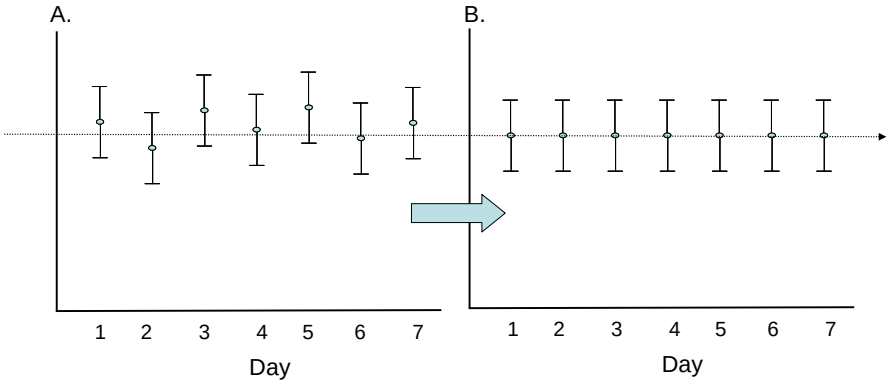
Compound identification: Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Identification of known chemical entities was based on comparison to metabolomic library entries of purified standards. As of this writing, more than 1000 commercially 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



References:

1. Evans, A.M., DeHaven, C.D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* 81, 6656-6667 (2009).
2. Dehaven, C.D., Evans, A.M., Dai, H. & Lawton, K.A. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J Cheminform* 2, 9 (2010).
3. DeHaven, C.D., Evans, A., Dai, H. & Lawton, K.A. in Metabolomics. (ed. U. Roessner) (InTech, 2012).