Metabolomic Profiling in Drug Discovery: Understanding the Factors that Influence a Metabolomics Study and Strategies to Reduce Biochemical and Chemical Noise

Mark Sanders[1], Serhiy Hnatyshyn[2], Don Robertson[2], Michael Reily[2], Thomas McClure[1], Michael Athanas[3], Jessica Wang[1], Pengxiang Yang[1], Yingying Huang[1] and David Peake[1];
[1]Thermo Fisher Scientific, San Jose, CA; [2]Bristol-Myers Squibb, Princeton, NJ; [3]Vast Scientific, Boston, MA

Key Words
Q Exactive Focus, SIEVE Software, Biomarker, Discovery, Metabolomics

Goal
To develop a new automated workflow and instrumentation for metabolomics biomarker discovery.

Introduction

Metabolomics is used within the pharmaceutical industry to investigate biochemical changes resulting from pharmacological responses to potential drug candidates. The ability to identify markers of toxicity/efficacy can significantly accelerate drug discovery and help define the appropriate clinical plan. Data from liquid chromatography-mass spectrometry (LC-MS) metabolomic profiling experiments contains large amounts of chemical background that often confounds biomarker discovery. New mass spectrometer technology and data processing software were utilized here to reduce chemical background in animal experiments investigating the relation of animal age and nutrition to discerning drug-induced changes.

In typical LC-MS metabolomics studies, much of the data is redundant (multiple ions per component) and irrelevant (chemical noise). External factors that influence metabolic profiles (age, nutrition) increase biological variation. Because many of the chemical entities are unknowns, it is especially important to filter false positives before implementing structure elucidation. Ultrahigh resolution instruments combined with ultra-high performance LC (UHPLC) separations address the issues of chemical noise and redundancy by providing sufficient resolution to distinguish metabolites from chemical background. Accurate mass data allows sophisticated processing needed to recognize related signals. This leads to significant reduction in data size and providing improved quantitation of targeted metabolites. Biological factors have profound impact on metabolic profiles and even modest metabolic changes can obscure drug-induced metabolic effects. Understanding normal metabolic changes in rats helps to minimize "biological noise" and provides more confidence in assigning specific drug-related metabolic changes.

Experimental

Sample Preparation

Blood samples were taken from groups of male rats (fully satiated, acute and chronic fasting, different ages) and analyzed using LC-MS. Protein was removed from serum samples (50 µL) by the addition of 100 µL of cold methanol with 0.1% formic acid. Samples were dried down and reconstituted in 200 µL of H_2O/methanol 90:10. N-benzoyl-D_5-glycine internal standard (tR = 4.27 min, m/z 185.0969) was spiked into every sample.

Liquid Chromatography

Chromatographic separation was achieved using an Thermo Scientific™ Open Accela™ 1250 UHPLC system and a Thermo Scientific™ Hypersil GOLD™ aQ column (150 × 2.1 mm, 1.9 µm particle size). The injection volume was 3 µL. The chromatographic conditions were as follows:

Flow Rate:	600 µL/min		
Column Temperature:	50 °C		
Solvent A:	0.1% formic acid in H_2O		
Solvent B:	0.1% formic acid in acetonitrile		
Gradient:	Time (min)	%A	%B
	0	100	0
	6	80	20
	8	40	60
	12	5	95
	14	5	95
	17	100	0

Mass Spectrometry

High resolution accurate mass (HRAM) data was acquired in both positive and negative ion mode using a Thermo Scientific™ Q Exactive™ Focus Hybrid Quadru-pole-Orbitrap mass spectrometer (Figure 1) operated at 70,000 resolution (FWHM).

Data Processing

The data was analyzed using Component Extraction (CE) data processing algorithms in Thermo Scientific™ SIEVE™ software to determine the metabolic effects of food deprivation on the rats.

Results and Discussion

Figure 2 illustrates the high quality LC-MS data obtained for the N-benzoyl-D_5-glycine internal standard. The positive ion data for serum QC replicates was obtained between 25 to 35 hours after mass calibration and demonstrates excellent peak area and mass measurement stability on a UHPLC timescale. The chromatographic peak width was 3.6 s at the base, and 15 scans were acquired across the peak.

Figure 3 illustrates the value of obtaining 70,000 resolution for determining elemental composition of endogenous metabolites. The expanded view around the A+2 isotope (*m/z* 313) shows a single ^{34}S is present. This assignment is not possible at 35,000 resolution (simulation) since the $^{13}C_2$ isotope is unresolved from ^{34}S at the lower resolution.
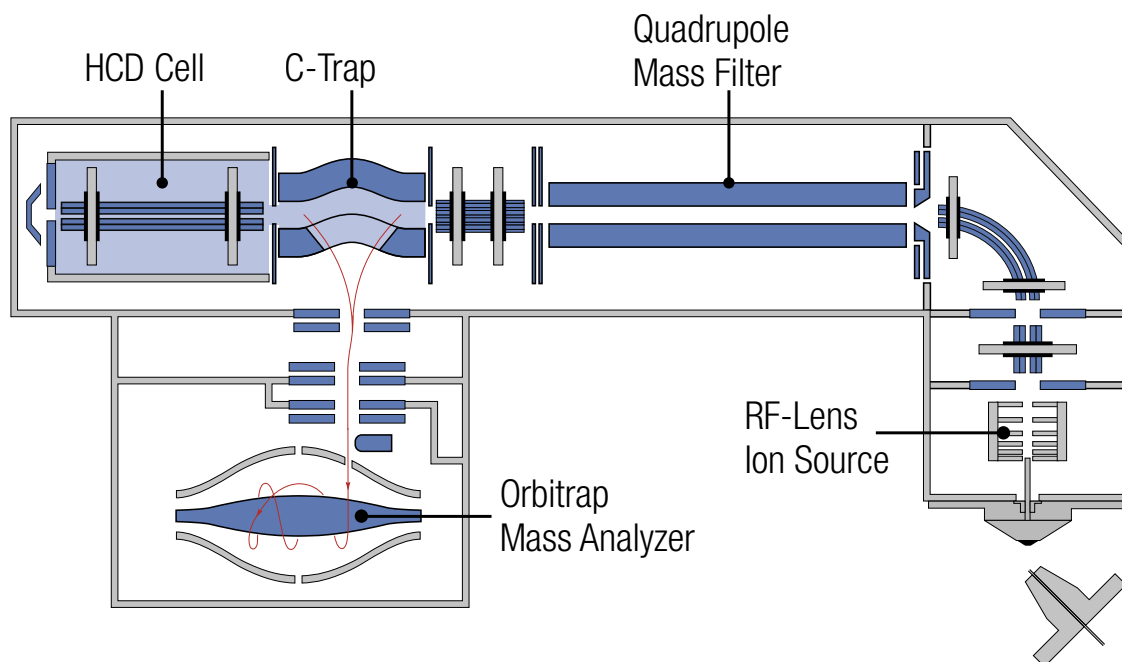

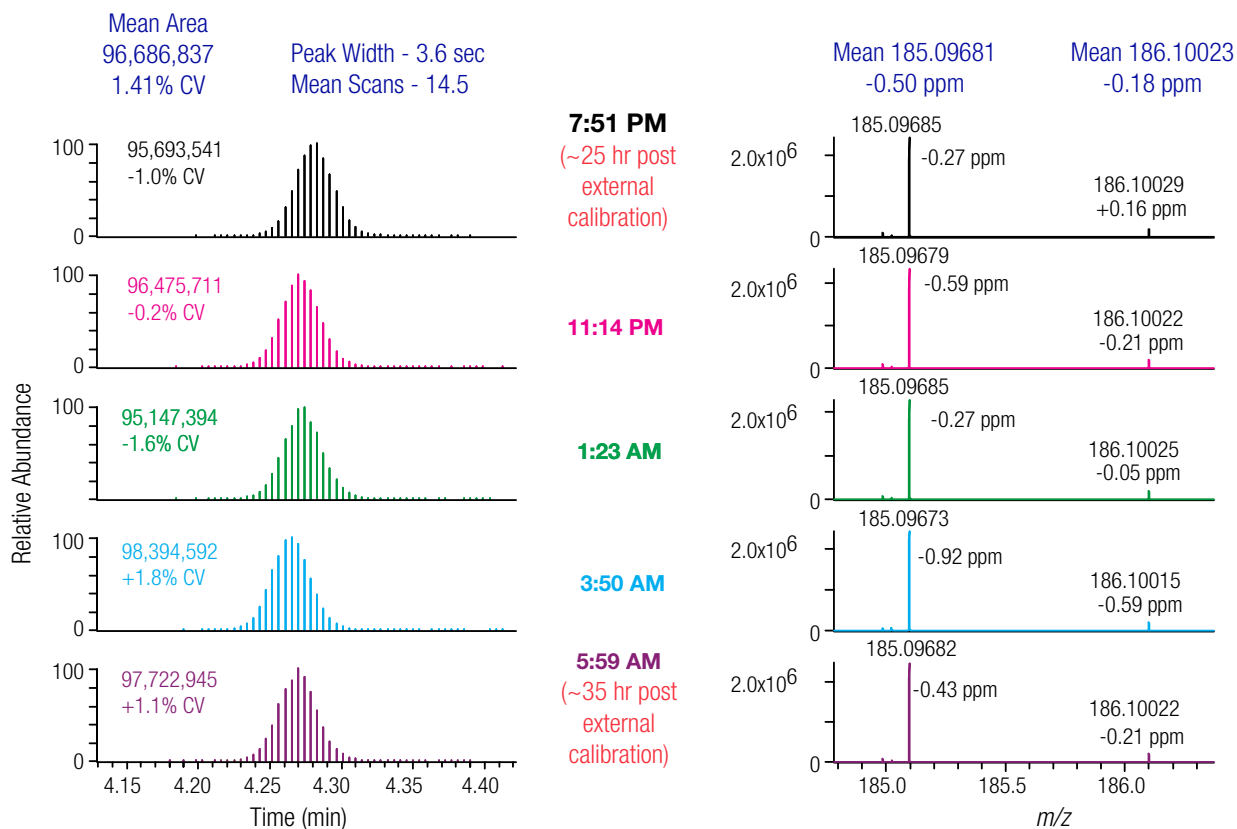Figure 1. Schematic diagram of the Q Exactive Focus mass spectrometer.

Figure 2. Mass and response stability of N-benzoyl-D$_5$-glycine, with external calibration, and resolution 82,000.
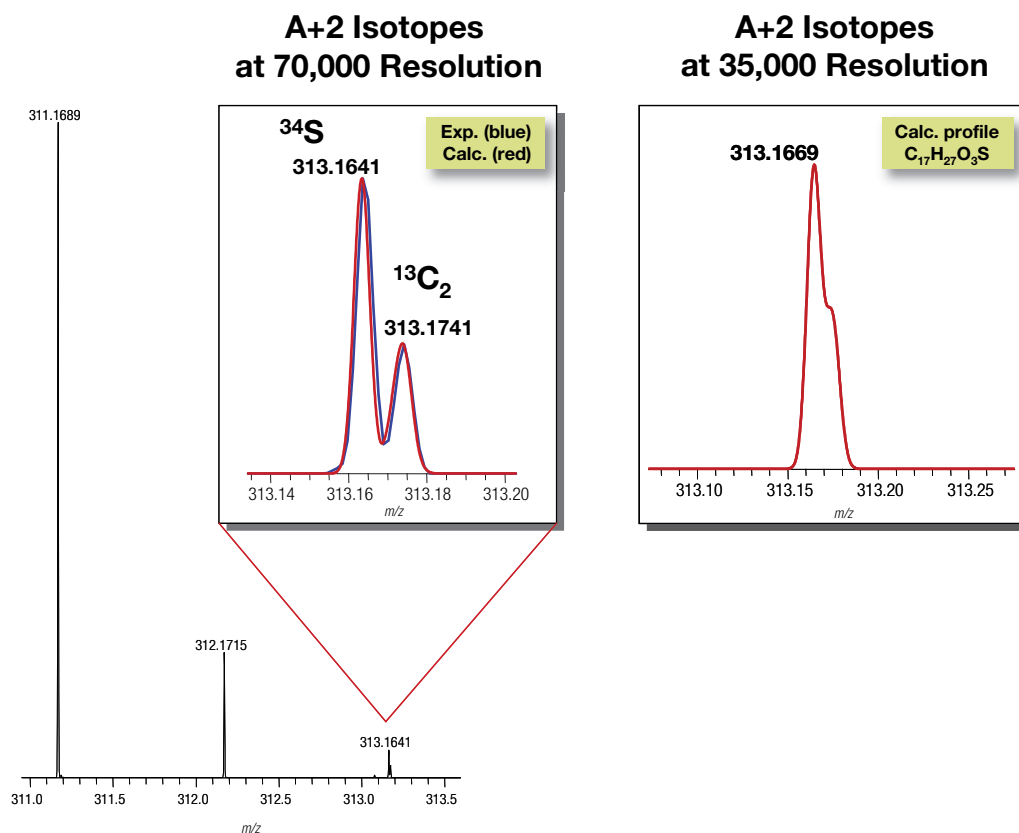

Figure 3. Isotopic fine structure combined with accurate mass measurement allows unambiguous assignments of elemental composition, e.g., 70,000 resolution is capable of separating the ^{34}S isotope from the ^{13}C$_2$ isotope, while 35,000 resolution is not sufficient.

The data processing workflow for component Extraction is shown in Figure 4. The software interprets the data like an analyst does. Instead of treating each data file as a separate entity, the data is processed in a batch and information gained in one run is used to verify information gained in the next run. In this way data gaps are minimized because each component is defined at its maximum concentration in the dataset. In samples where the concentration is much lower, the same component is identified and quantified using a more targeted approach.

A high degree of data reduction was achieved. The processing removed much of the noise from the system, leading to tighter statistical groupings and more confidence in the differential analysis and putative assignments.

Table 1 describes the rat study designed to monitor the effect of fasting on metabolic profiles. Figure 5 shows that the principal component analysis (PCA) nicely clusters the control group of fed rats, the pooled QCs, and 4, 12, and 16 hour fasted serum. There is clearly a difference between fed versus fasted serum and time of fasting. Figure 6 shows metabolites that are increasing (Met, 20:4 FA) and decreasing (Pro, 18:2 LPC) with fasting time. As shown in Figure 6, for each metabolite, excellent reproducibility was achieved in the pooled serum QC replicates. Hence technical replicates are not necessary. Figure 7 illustrates that the same patterns of uric acid are observed in serum analyzed in both positive and negative ion mode despite the 30 hours between the actual LC-MS run. The LCMS platform and method were demonstrated to be very robust. It is concluded that biological variability is the primary source of noise in these data.

Conclusion

The Q Exactive Focus mass spectrometer provides a precise and robust platform for untargeted metabolomics studies. The platform has fast scan speeds compatible with UHPLC, and can be used with external calibration in both positive and negative ion modes for an extended period of time while keeping excellent mass accuracy and response. Technical replicates are not necessary. The 70,000 resolving power allows fine isotope pattern to be obtained which can aid unambiguous elemental composition. To deal with the numerous sources of noise inherent to these studies, intelligent data reduction tools found in SIEVE software can be used to significantly reduce the chemical noise. In addition, the use of systematic studies help to characterize biological noise, while metabolomic prescreening can help identify biological outliers to ensure homogeneity within an entire study.

As demonstrated in this study, fasting is a significant variable in model design, and fasting data can help contextualize drug-induced changes in many metabolites. Fasting in rats was found to have a profound impact on metabolomic profiles. Although most metabolic changes were modest in extent, fasting exacerbated or obscured some drug-induced metabolic effects.
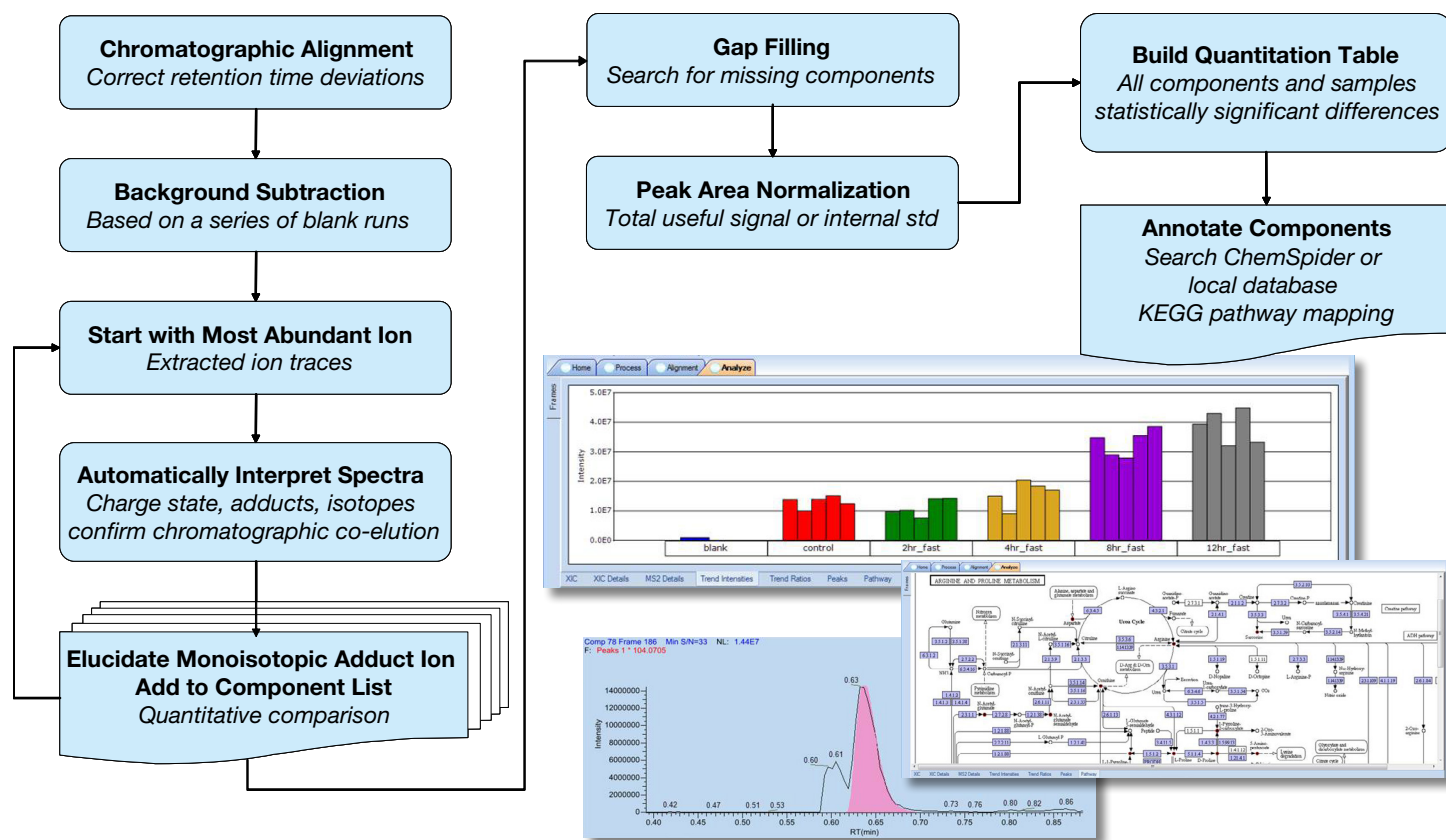
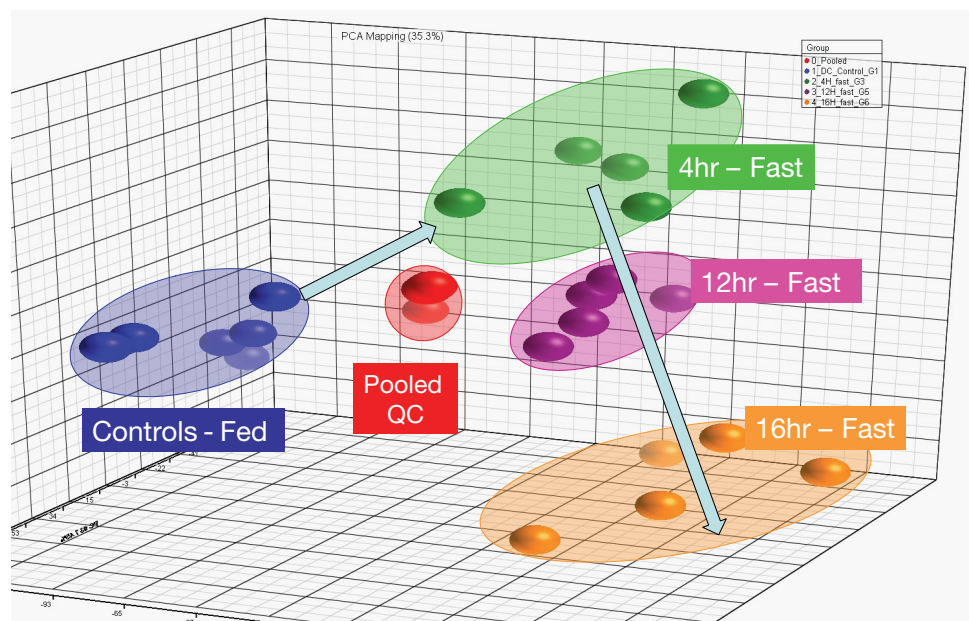Figure 4. Data processing workflow in SIEVE software with component elucidation algorithms.

Figure 5. Principal component analysis (PCA) of rat serum negative ion LC-MS sata.

Table 1. Rat fasting study design.

Group	Male, n	Fasting Time*
1: 1101-1105	5	Dark Cycle Control (No Fast)
2: 2101-2105	5	2 hr Fast
3: 3101-3105	5	4 hr Fast
4: 4101-4105	5	8 hr Fast
5: 5101-5105	5	12 hr Fast
6: 6101-6105	5	16 hr Fast

* Rats fasted during 6:00 PM to 6:00 AM dark cycle to capture peak feeding time.
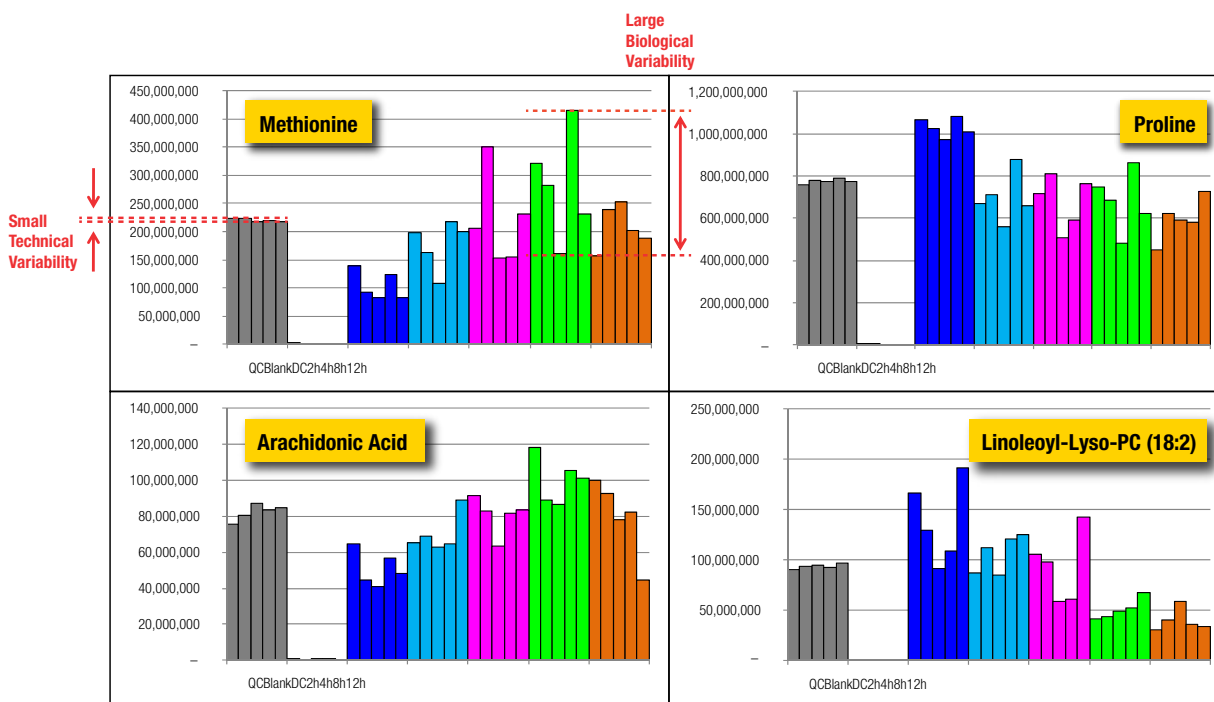
Figure 6. Examples of metabolite changes upon fasting. Excellent reproducibility was achieved for all metabolites in the pooled serum QC replicates. Hence technical replicates are not necessary.
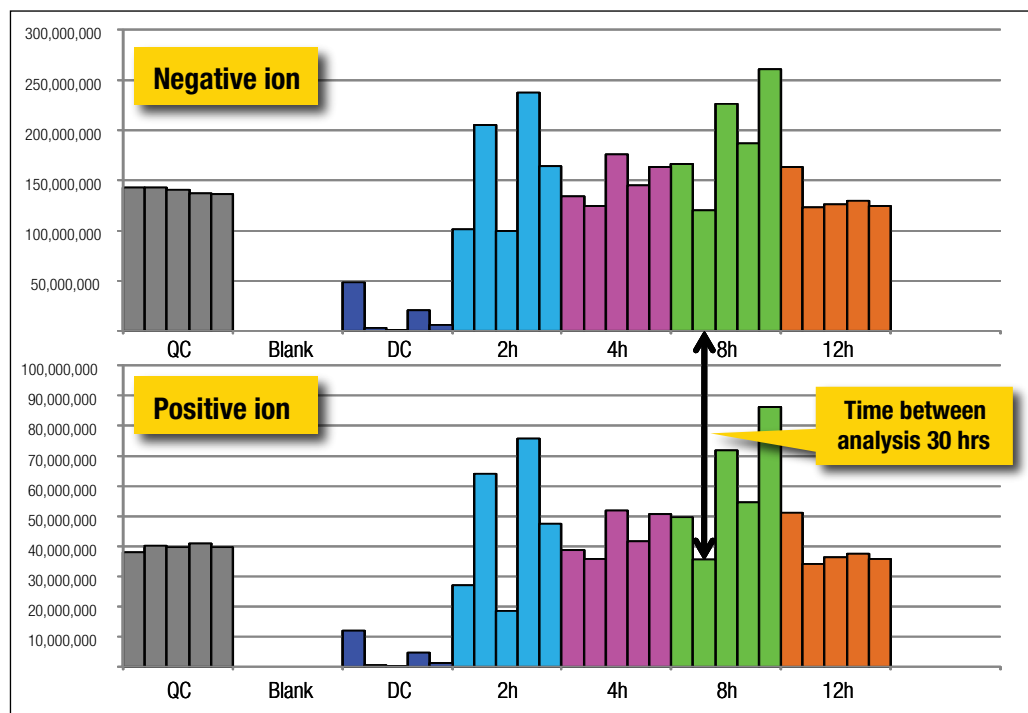
Figure 7. Uric acid positive and negative ion data showed perfect correlation between the measurements between biological samples in different polarity runs, in spit of the 30 hours between the actual LC-MS runs.

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand