Published in final edited form as:
Anal Chem. 2017 January 03; 89(1): 656–665. doi:10.1021/acs.analchem.6b02930.

Inter-laboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma

Alexandros P. Siskos[#1], **Pooja Jain**[#1], **Werner Römisch-Margl**[2], **Mark Bennett**[3], **David Achaintre**[4], **Yasmin Asad**[5], **Luke Marney**[6], **Larissa Richardson**[6], **Albert Koulman**[6], **Julian L. Griffin**[6], **Florence Raynaud**[5], **Augustin Scalbert**[4], **Jerzy Adamski**[7,8,9], **Cornelia Prehn**[7], and **Hector C. Keun**[1,*]

[1]Department of Surgery and Cancer, Imperial College London, W12 0NN, UK

[2]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany

[3]Department of Life Sciences, Imperial College London, SW7 2AZ, UK

[4]International Agency for Research on Cancer (IARC), Biomarkers Group, F-69372 Lyon, France

[5]The Institute of Cancer Research, ICR, Sutton, SM2 5NG, UK

[6]MRC Human Nutrition Research, Cambridge, CB1 9NL, UK

[7]Genome Analysis Center, Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany

[8]Chair of Experimental Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, 85354 Freising-Weihenstephan, Germany

[9]German Center for Diabetes Research (DZD), 85764 Neuherberg, Germany

[#] These authors contributed equally to this work.

Abstract

A critical question facing the field of metabolomics is whether data obtained from different centres can be effectively compared and combined. An important aspect of this is the inter-laboratory precision (reproducibility) of the analytical protocols used. We analysed human samples in six laboratories using different instrumentation but a common protocol (the Absolute*IDQ*™ p180 Kit) for the measurement of 189 metabolites via liquid chromatography (LC) or flow-injection analysis (FIA) coupled to tandem mass spectrometry (MS/MS). In spiked quality control (QC) samples 82% metabolite measurements had an inter-laboratory precision of <20%, while 83% of averaged individual laboratory measurements were accurate to within 20%. For 20 typical biological samples (serum and plasma from healthy individuals) the median inter-laboratory CV was 7.6%, with 85% of metabolites exhibiting a median inter-laboratory CV of <20%. Precision was largely independent of the type of sample (serum or plasma) or the anticoagulant used but was reduced in

[*]Corresponding author. Hector C. Keun, h.keun@imperial.ac.uk.

Conflict of Interest Disclosure
The authors declare no competing financial interest.

Europe PMC Funders Author Manuscripts

a sample from a patient with dyslipidaemia. The median inter-laboratory accuracy and precision of the assay for standard reference plasma (NIST SRM 1950) were 107% and 6.7%, respectively. Likely sources of irreproducibility were the near-LOD typical abundance of some metabolites and the degree of manual review and optimisation of peak integration in the LC-MS/MS data post-acquisition. Normalisation to a reference material was crucial for the semi-quantitative FIA measurements. This is the first inter-laboratory assessment of a widely-used, targeted metabolomics assay illustrating the reproducibility of the protocol and how data generated on different instruments could be directly integrated in large-scale epidemiological studies.

Introduction

Metabolic profiling (often referred to as metabonomics or metabolomics)[1,2] of body fluids provides a unique view of the metabolic status of an individual and their exposure to dietary or environmental factors, and can inform as to how these interact with genotype or are modulated by drugs. Liquid chromatography mass spectrometry (LC-MS) based platforms are most widely used for metabolomics studies and often exhibit greater sensitivity and metabolite coverage compared to other techniques. Many studies use high-mass resolution detectors, such as quadrupole time-of-flight (Q-ToF) and Orbitrap instruments, to screen biomolecules in an untargeted manner. Alternatively targeted approaches, that preselect species to measure, typically use multiple reaction monitoring (MRM) on tandem MS (MS/MS) low-mass resolution detectors to enhance sensitivity and selectivity.

The Biocrates Absolute*IDQ*™ p180 kit is a commercially available targeted metabolomics assay that can be used on a variety of LC-MS/MS triple quadrupole instruments. The kit has already been applied to many studies of human serum and plasma, including clinical studies for disease biomarker discovery,[3–5] biomarker of target engagement[6] and several large-scale prospective cohort studies (with number of participants n>1000) such as EPIC (European Prospective Investigation into Cancer and Nutrition)[3,7–10] and KORA (Cooperative Health Research in the Region of Augsburg).[5,11,12] Data are now available regarding many critical aspects of the use of this platform, including the influence of pre-analytical factors,[13,14] the differences between human plasma and serum metabolite profiles,[15] the evaluation of between- and within- person metabolite variation[7,16] and the influence of common confounders such as age,[17] anthropometry,[18] smoking,[19] the effect of sleep restriction and circadian clock disruption[20] as well as assessment of heredity[21] and genome-wide perspectives of variation in human metabolism.[11]

Despite the rapid progress in the use of metabolomics platforms a key limitation is currently the lack of methodological standardisation and testing of the comparability of data between laboratories. Very few multicentre metabolomics studies of reproducibility have been reported to date and those that exist have primarily used untargeted methods, and so have been unable to define metabolite specific data on reproducibility or accuracy. Both high intra-laboratory precision (across many samples over an extended period of time) and inter-laboratory precision (across different centres, instruments and studies) are critical for pooling of data and conducting meta-analyses on clinical and molecular epidemiological studies. In a conventional evaluation of precision and accuracy, the background matrix is

Europe PMC Funders Author Manuscripts

rarely varied. It is also very difficult, for blood metabolomics studies, to obtain a representative analyte-free matrix. This is inadequate when we are often trying to interpret modest concentration differences in a few metabolites in the context of large changes in the background composition, e.g. in plasma the lipoprotein and protein composition can vary dramatically affecting the sample matrix greatly. Therefore, we sought to evaluate the inter-laboratory reproducibility of this targeted metabolomics assay in the context of a panel of different test human plasma and serum samples, and also compared the effects of different anticoagulants on reproducibility. Across six participating laboratories several different instrumental platforms were used, allowing assessment of the comparability of data derived from different platforms. Finally, we assessed the pros and cons of data normalisation to improve the comparability of data generated in different laboratories. Our current work represents the first inter-laboratory assessment of the reproducibility of the widely-used, targeted LC-MS metabolomics Absolute*IDQ*™ p180 kit on human plasma and serum, and provides critical knowledge for the successful integration and validation of large-scale metabolomics biomarker datasets generated from different laboratories, on epidemiological and drug development research.

Material and Methods

Twenty-six test materials, including spiked quality control (QC) samples with known concentrations of metabolites, the NIST (National Institute of Standards and Technology) reference human plasma (SRM 1950) and serum/plasma collected from volunteers were distributed to six laboratories and independently analysed using the Absolute*IDQ*™ p180 kit following the manufacturer's protocol but using different combinations of MS instrument and HPLC/UHPLC. The kit allows the targeted analysis of up to 189 metabolites (see Table S-5) in the metabolite classes of amino acids, biogenic amines, acylcarnitines, glycerophospholipids, sphingolipids and sum of hexoses, covering a wide range of analytes and metabolic pathways in one targeted assay. The Kit consists of a single sample processing procedure, although two separate MS analytical runs, a combination of liquid chromatography (LC) and flow-injection analysis (FIA) coupled to tandem mass spectrometry (MS/MS). Isotope-labelled and chemically homologous internal standards are used for quantification, and in total 56 analytes are fully validated as absolutely quantitative.

Of the total 189 metabolites measured, 43 metabolites are measured by LC-MS/MS and 146 metabolites by FIA-MS/MS. The amino acids (21) and biogenic amines (22) are analysed quantitatively by LC-ESI-MS/MS, with the use of external calibration standards in seven different concentrations and isotope labelled internal standards for most analytes. All amino acids and amines are fully validated as absolutely quantitative. The acylcarnitines (40), glycerophospholipids (90), sphingolipids (15) and sum of hexoses (1) are analysed by FIA-ESI-MS/MS, using a one point internal standard calibration with representative internal standards (9 isotope-labelled acylcarnitines, 1 isotope-labelled hexose, 1 non-labelled lyso-PC, 2 non-labelled PCs, 1 non-labelled SM, a total of 14 internal standards). In terms of quantification, the lipids and a subset of acylcarnitines are called "semi-quantitative" since specific standards were not commercially available and a verification of the accuracy was not possible by the manufacturer. 12 acylcarnitines and the sum of hexoses are fully validated as absolutely quantitative. In addition many of the FIA-detected, semi-quantitative

Europe PMC Funders Author Manuscripts

lipid concentrations represent total concentrations of possible isobars and structural isomers. The results for the metabolites are displayed with a corresponding short name with the total length of side chains and the total number of double bonds. The kit utilises a patented 96-well plate design which allows simultaneous efficient sample derivatisation and reproducible analyte extraction. The kit is suitable for manual or automated high throughput operation, and it requires only a very small sample volume of 10 μL and comes with human plasma based quality controls in 3 concentration levels (low, medium, high) which can be used for quality control purposes but also potentially for batch normalisation. The baseline analytical validation and performance of the kit has been described by the manufacturer22 and a summary is given in Tables S-14 and S-16. Moreover three of the participating laboratories provided further data on the inter-plate and intra-plate variability (Tables, S-11, S-12 and S-13).

Study population/test materials

A total of 26 test materials were used, replicated 3-6 times, utilising a total of 85 positions on a single kit 96-well plate, in each participating laboratory (Table S-1):

- 3 QCs provided by the manufacturer with three concentration levels of 59 standards spiked into a plasma background: *p180-MetaDis QC levels* 1-3; low (x5 replicates per plate), medium (x5 replicates per plate), and high (x3 replicates per plate). Total 13 positions per plate.
- 1 NIST standard reference material (**SRM 1950**, lithium heparin plasma) (x3 replicates per plate).
- 1 EDTA plasma sample from an individual with dyslipidaemia (x3 replicates per plate).
- **20 test materials** (x3 replicates each per plate) collected from 8 healthy individuals. For 4 individuals EDTA plasma only was collected (Test samples 02 to 05). For the other 4 individuals repeated collections with three different anticoagulants (EDTA, citrate and heparin) and serum were conducted Test Samples 07-1 to 10-4).
- 1 pooled QC Sample (QCP), prepared by each individual laboratory by pooling 10 μL of each of 16 test materials (4 individuals x 4 collections, Test Samples 07-1 to 10-4) (x6 replicates per plate).

All samples were collated, recoded, aliquoted and distributed by laboratory E to the other laboratories with those participants blinded to the specific identify of each test material until data acquisition was completed. All experimental procedures were approved by the local Ethics Committee.

Metabolomics measurements

Basic common guidance (Protocol S-1) was agreed on the cleaning and benchmarking of instruments prior to analysis, and also the run order and the position of samples in 96 well plates, with vertical pipetting and run order mode (Figure S-1). Plasma and serum metabolite concentrations were determined using the targeted metabolomics kit Absolute*IDQ*™ p180

Europe PMC Funders Author Manuscripts

kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). The samples were prepared, by all participating laboratories, according to the manufactures's protocol22 (details in Protocol S-2).

Laboratory instrumentation – MS analysis

Each laboratory followed the manufacturer's protocol but used different UHPLC/HPLC or MS/MS platforms (details in Protocol S-2). Laboratories A, B, C and E used HPLC with SCIEX mass spectrometers. Lab F used Waters UHPLC with a Waters mass spectrometer. Lab D used a combination of Waters UHPLC with a SCIEX mass spectrometer. Two examples of typical MS analytical procedures are described in Protocol S-2.

Data transformation - Statistical analysis

For the LC-MS/MS assay, the metabolites were quantified by stable isotope dilution and seven point calibration curves. For the FIA-MS/MS assay, metabolite concentrations were calculated using a one point internal standard calibration, and are also isotope corrected. Metabolites were quantified (results shown in μM concentration units) according to the manufacturer's protocol using the Met*IDQ*™ Boron software for targeted metabolomic data processing and management. Blank PBS (phosphate-buffered saline) samples (3 replicates) were used for the calculation of the limits of detection (LOD). The median values of all PBS samples on the plate were calculated as approximation of the background noise per metabolite, and 3 times this value was calculated by each laboratory as the LOD (Table S-15 includes the LODs for all the metabolites as calculated by each laboratory). Raw data from each participating laboratory were exported as .xls format and then collated by lab B for further statistical analysis and inter-laboratory comparison. For each test sample, concentration means, accuracies and intra-lab CVs were calculated for the metabolites having at least two valid replicates (Table S-17). For each test sample, an inter-lab CV was then calculated for the metabolites that had valid intra-lab CV from at least three laboratories. Data are available on request and will also become publicly available.

Missing data, exclusions and outliers

In total, across all laboratories 91,025 individual metabolite measurements out of a theoretical total of 96,390 (189 metabolites x 85 samples x 6 laboratories) (94.4 %), were included in our study. Principal reasons for missing data and exclusions were: the manufacturer's protocol indicated not to acquire data for some metabolites for specific instrumental platforms due to known selectivity issues ('not acquired' or 'NQ', 850 measurements 0.88%); no peak was detected or the peak could not fit to the calibration curve ('N/A' reported, 842 measurements 0.87%); or the integral value gave a negative or zero concentration according to the calibration curve (zero value reported, 2182 measurements 2.26%). Laboratories that used HPLC did not acquire data for sarcosine, whereas laboratories that used UHPLC did not acquire data for total-DMA. Laboratories that used Waters mass spectrometry instruments did not acquire data for four lipids (PC aa C30:2, PC aa C32:2, PC aa C38:1, SM C22:3). Other exclusions included two cases of human error, specifically the omission of the addition of internal standards for four samples for Lab F and an error in the preparation of the six QCP samples for lab D. Very few measurements (0.4%) were considered to be unexplained outliers, as assessed by Principal

Components Analysis (PCA; Figure S-2 just 2 samples of 510) or by visual inspection (1 measurement). To avoid imputation during subsequent statistical analysis recorded values below the limits of detection (BLD) were not excluded from the analysis. Table S-2 gives a detailed breakdown of why certain data were missing or excluded.

Results

Assay performance: Accuracy, intra- and inter-laboratory reproducibility of spiked QC samples

In our analyses we first considered the reproducibility of the quality control (QC) samples supplied routinely by the manufacturer. These consist of mixtures of human plasma spiked with 59 metabolites (42 and 17 measured by LC and FIA respectively) to 3 nominal concentrations ('low' – QC1, 'medium' – QC2, 'high' – QC3) allowing assessment of intra- and inter-laboratory accuracy and precision for these metabolites (Table S-3). Two of the concentrations values were not provided (PEA and sarcosine for 'low' – QC1), and therefore data for these are not presented. Data for sarcosine for laboratories A, B, C and E were not acquired. Only one laboratory (D) reported further missing data (for histamine, putrescine and PEA).

Overall, 144 (82%) of the 175 repeated metabolite measurements (59 metabolites x 3 concentration levels – 2 missing), had an inter-laboratory precision of <20%. For the majority of the metabolites assessed (41/59), the reproducibility was <20% at all three concentrations, and for 18 metabolites (16 and 2 measured by LC and FIA respectively) the reproducibility was >20% for at least one concentration level. In terms of precision at the level of individual laboratories, of the 1042 repeated measurements (59 metabolites x 3 concentrations x 6 laboratories – 20 missing), 1008 (96.7%) had an intra-laboratory CV<20% and 771 (74%) had a CV<10% (for a full breakdown of QC sample data by laboratory see Table S-3). Moreover, 35/59 metabolites had intra-laboratory precision <20% for all three concentration levels and all six laboratories. For 48/59 metabolites the accuracy of the assay as determined by the averaged value across laboratories, was within 20% of the nominal concentration at all three concentrations, while 83% (863 of 1042) of averaged intra-laboratory measurements were within 20% of nominal concentration (Table S-3). It was observed that 7/59 metabolites had all accuracies within 20% for all three concentration levels and all six laboratories; at threshold of 30% this reached 32/59 metabolites.

Of the 18 metabolites exceeding an inter-laboratory precision of 20% for at least one concentration level, 10 metabolites did so for only one concentration level. 6 metabolites (glutamate, acetylornithine and methioninesulfoxide, *cis*-4-hydroxyproline (c4-OH-Pro), DOPA and dopamine) did so for the lowest concentration, while 1 metabolite (PEA) did so for the medium level, and 3 (kynurenine, lysine, ornithine) for the highest concentration only. Of the remaining 8 metabolites (sphingomyelin C18:0, dodecanoylcarnitine (C12), symmetric-dimethylarginine (SDMA), spermine, spermidine, *trans*-4-hydroxyproline (t-4-OH-Pro), carnosine and nitro-Tyr), with poor reproducibility for at least two concentration levels, dodecanoylcarnitine (C12) was measured at consistently higher concentrations (˜150%) for just Lab C and high accuracy and precision were observed for the other laboratories. Measurements of sphingomyelin C18:0 were precise but inaccurate by a fixed

percentage within Lab A (74%), Lab B (49%) and Lab F (185%), indicating that normalisation by a single multiplicative factor could improve inter-laboratory agreement for this and potentially other FIA-detected metabolites. For SDMA, only lab B reported acceptable analytical performance; one laboratory (F) reported values ˜10-fold higher than the nominal values while for four labs the accuracy ranged between 39.2% and 230%. For DOPA all laboratories reported poor accuracy results for the low concentration, whereas for dopamine two laboratories reported poor accuracy results again for the low concentration. The poor reproducibility of the polyamines spermine, spermidine and carnosine were primarily the result of anomalously low quantities reported by laboratory D (<3% of nominal value) while for t-4-OH-Pro laboratory D reported anomalously high quantities (up to 454%) and also reported poor accuracy for nitro-Tyr (12-208%). Notably for 8 metabolites (spermine, spermidine, t-4-OH-Pro, carnosine, nitro-Tyr, Glu, c-4-OH-Pro and PEA) an apparently low reproducibility could be attributed to the poor accuracy of a single outlying laboratory (see Table S-3). Excluding data from laboratory D for the analysis of these metabolites brought inter-laboratory precision to within 20% in each case.

Normalisation

We next analysed the reproducibility of the 20 test materials representing typical samples taken from healthy individuals (serum and plasma samples: Test Samples 02 to 05 and Test Samples 07-1 to 10-4) which formed our primary test set. Initial PCA indicated that systematic differences were present between the laboratories in the metabolic profiles (Figure S-3), particularly the FIA-detected lipid profile. Therefore, we investigated methods for normalising the data to correct for batch and/or instrumental platform effects.

In order to define a single normalisation factor for each metabolite per laboratory we used a single reference sample from each batch and calculated the fold change for each metabolite relative to the reference sample value for a specific reference laboratory. Thus, the normalised value X_{ij} for a metabolite i, for a laboratory j was given by:

$$Normalised\ X_{ij} = X_{ij} * \frac{\text{mean value (metabolite i) in reference material from reference Lab}}{\text{mean value (metabolite i) in reference material from Lab j}}$$

where, i= 1-189 are measured metabolites, and j=1-5 were the participating laboratories.

As all laboratories had some missing or zero values (Tables S-5 and S-6), the laboratory with the least overall missing data (Lab B) was selected as the reference laboratory. For the reference material we tested both TS 06 (NIST SRM 1950 plasma) and the spiked QC2 (Biocrates *p180-MetaDis QC level* 2), and also we investigated normalising to the mean or median of each reference material. For both reference materials it appears that normalising to the mean was marginally more effective than normalising to the median. Figure 1 and Figure 2 (Table S-4) show the effect of normalisation on our primary test set. Normalisation using the NIST SRM 1950 produced a substantial improvement in the CV distribution: of a total of 3780 inter-laboratory CV values, the proportion of CV values <20% increased from 54% to 84%. Normalisation using the QC2 samples also made improvements but to a lesser extent (72% of CVs <20%). For the metabolites quantified by LC-MS/MS the improvement in reproducibility of the data resulting from normalisation was only marginal (from 59% to

Europe PMC Funders Author Manuscripts

64% of CV<20% using the NIST SRM 1950), whereas for metabolites quantified via FIA-MS/MS the inter-laboratory precision was greatly improved (from 53% to 90% of CV<20%). For sphingolipids in particular the effect was dramatic: of a total of 300 sphingolipid inter-laboratory CVs calculated, 270 CVs were >30% before normalisation and 272 CVs <20% post normalisation. In light of these findings normalisation was applied only to the FIA part of the assay for subsequent analysis, and the mean values of NIST SRM 1950 used for adjustment. Unless explicitly stated, further analyses of our primary test set are presented using these 'FIA only normalised' data.

Inter-laboratory reproducibility (% CV) in 20 typical test materials

A full, per metabolite, breakdown of reproducibility is given in Table S-5, a per metabolite class in Table S-6, with summary values presented in Table 1 Across all 189 metabolites analysed with the Absolute*IDQ*™ p180 Kit, following normalisation of the FIA-quantified metabolites, the median inter-laboratory precision across 20 test materials was 7.6%. A high proportion (~85%) of the total measured metabolites (160 metabolites) had a median inter-laboratory precision (<20%), with 123 metabolites exhibiting median inter-laboratory precision <10% (Figure 3). For 24 metabolites the median inter-laboratory precision across 20 test materials was >20% and for a further 5 metabolites the calculation of inter-laboratory CV was not possible, due to missing data (with a minimum requirement of two replicates measurement for a given sample per lab, and for at least three laboratories per metabolite). The average ratio of intra-laboratory CV to inter-laboratory CV across all metabolites ranged from 1:1.4-1.5 for laboratories A, B, C & E while for laboratories D and F the ratio was ˜1:1, indicating that overall reproducibility was lower within these last two laboratories (Table S-7).

Breakdown of the reproducibility into metabolite classes revealed that with the exception of the 'biogenic amines', the median inter-laboratory precision for each class was also <20% (Table 1). FIA-normalisation made a dramatic improvement in the reproducibility of sphingolipids measurements (median CV from 68% to 6.7%) and also made a critical difference in the reproducibility of carnitines (median CV from 34% to 12%) which were frequently below the limits of detection (BLD) or missing (48%). For the remaining FIA part of the assay we observed a very high reproducibility (after normalisation) across the overwhelming majority of the metabolites across the 20 test materials: the median inter-laboratory precision was 6.4 % for the di-acyl PCs (PC aa), 5.9 % for the alkyl-acyl PCs (PC ae), 8.1 % for the lyso PCs and 6.7 % for the sphingolipids. Laboratories that used Waters instruments did not acquire data for 4 of the lipids (PC aa C30:2, SM C22:3, PC aa C32:2, PC aa C38:1). Of these 4 lipids PC aa C30:2 and SM C22:3 also showed poor performance for some of the other laboratories and poor inter-laboratory precision. With the exceptions of aspartic acid (Asp) and glutamate (Glu) the precision of the amino acid analysis was very high (median 7.1%), with almost all of the concentration values being above the limits of detection.

Of the 45 metabolites reporting at least 30% of their data missing or BLD, 26 were carnitines indicating that the typical abundances of many of these metabolites in blood serum or plasma were below the operating limits of the assay (Table S-8). However,

mitigating this was the observation that after normalisation only 3 carnitines (C3:1, C5:1-DC, C9) produced inter-laboratory CVs>20%. Of the 5 metabolites for which missing data precluded reproducibility analysis, sarcosine was not part of the assay for 4 laboratories due to the requirement of UHPLC for measurement. Between the two laboratories that did measure sarcosine the agreement was within ~10% (data not shown). For the other four metabolites (PEA, nitro-tyrosine, cis-4-OH-Pro and dopamine) all laboratories consistently reported extensive missing data (>80% in total) indicating that in normal serum/plasma these metabolites could not be detected reliably by the assay.

Of the 24 metabolites that produced a CV>20% (Table S-8), 12 also had a high proportion of missing values or BLD, although these were not always evenly distributed between laboratories (Table S-5). 11 of the 24 were of the 'biogenic amine' class and 13 were measured during the LC-MS step of the assay. For biogenic amines the average intra:inter-laboratory CV ratio was also high (~1:3-6.9) compared to all other metabolites (Table S-7). For carnosine, DOPA and histamine the majority of values across laboratories were reported as missing or BLD indicating that in normal serum/plasma these metabolites were not detected reliably by the assay (Table S-5). Of the biogenic amines with highly variable detection across laboratories Ac-Orn was not detected by laboratory A and both laboratory D and F generated a high proportion of missing values (53% and 97%, respectively). Methionine sulphoxide (Met-SO) was not detected in 90% of samples by Laboratory D while laboratories E & F reported high numbers of missing values. Laboratory D also reported a high proportion of BLD values (95%) for alpha amino-adipic acid (Alpha-AAA). Laboratories B, D and F did not detect one or more of the polyamines putrescine, spermidine and spermine as reliably as other laboratories. SDMA was poorly detected by Laboratory A only (55% missing values) but reported with good precision in other laboratories. Of the amino acids with low reproducibility aspartic acid (Asp) was measured with an intra-laboratory precision>20% for laboratories C and D. For the LC-MS detected metabolites only glutamic acid (Glu) and ADMA returned a high inter-laboratory CV despite typically good intra-laboratory precision and low numbers of missing samples. Overall this analysis indicated that the reliability and reproducibility of biogenic amine quantification in particular, was variable across laboratories as a result of both the relatively low abundance of several metabolites of this class in blood samples (compared to the LODs for the assay) and also differences in the level of individual operator review of the raw LC-MS data.

The effect of anti-coagulants, high lipids and manual sample pooling on inter-laboratory reproducibility

Table S-9 shows a comparison of inter-laboratory precision for sera and plasma samples obtained with different types of anti-coagulant from the same individuals. The overall reproducibility of serum data across all metabolites (median CV 6.9%) was marginally lower than for matched data across all laboratories generated on plasma (median CV 7.2-8.2%) with heparin plasma producing the highest values. This trend was observed for all metabolite classes detected by FIA but was less clear for LC-MS detected metabolites. Reproducibility measurements made on a sample from an individual with pathologically high blood lipids (Table S-9) revealed that inter-laboratory precision was substantially worse for phospholipids including sphingomyelins, with CVs increasing by ~4-fold (to 24-26%).

Increased CVs were also observed for hexoses and lyso-phosphatidylcholines (to 12.2 and 13%, respectively). Metabolites detected by LC-MS appeared to be generally unaffected. We also considered the reproducibility of a pooled QC sample (QCP) generated from the 20 typical biological test materials (Table S-9). Inter-laboratory precision for all classes of metabolites, with the exception of biogenic amines, was observed to be higher for the QCP than for the original test materials (data not shown). This suggests that the manual preparation of the pooled sample added discernible variability when comparing data from the assay between laboratories. This seemed to be the case with laboratory D which reported experimental issues and errors with the preparation of the QCP. With the exclusion of laboratory D, the median inter-laboratory precision for the QCP samples was comparable to the rest of the test material (Table S-9).

Inter-laboratory precision and accuracy of metabolite quantification compared to reference values for the NIST 1950 human plasma standard reference material

Reference values are currently available for 60 metabolic measurements on the NIST SRM 1950 of which 19 overlap with the Absolute*IDQ*™ p180 Kit panel.23–25 Comparison of the estimated concentration of these metabolites from this assay to reference values revealed that the overall accuracy of the assay (the mean accuracy across laboratories) was within 20% for 18/19 metabolites (Figure 4, Table S-10). The average accuracy of the assay across metabolites was 108% with accuracy>100% reported for 15/19 metabolites; this was consistent with a slight upward bias in the assay for most measurements. For the amino acid serine the bias was clearly present in all 6 laboratories (accuracy ranging from 133-155%). For laboratories D, E and F the upward bias was>20% in at least 4/19 metabolites. In terms of inter-laboratory reproducibility the assay performed very well, with CVs ranging from 1.2% (alanine) to 12.1% (lysine).

Discussion

Very few studies have compared the inter-laboratory reproducibility of metabolomics data. Of those that have, the predominant focus has been on untargeted profiling methods. In an early study, a high reproducibility was reported for NMR spectroscopy analysis of urine in rodents, with >95% correlation and 4-8% inter-laboratory variability for measurements of three selected metabolites between two laboratories.26 Similarly high reproducibility was reported for NMR spectroscopy-based plant metabolomics in a 5-laboratory study,27 and a 7-laboratory NMR environmental metabolomics study.28 Other studies have focused on GC-MS plant metabolomics,29 NMR-based metabolomics for olive oil,30 and yeast metabolite profiling.31 Recently the Metabolomics Research Group of the Association of Biomolecular Resource Facilities (ABRF) conducted a “round-robin” study across 14 laboratories and 25 different metabolomics platforms including GC-MS, LC-MS and NMR spectroscopy.32 In this study two groups of samples were generated by spiking 17 compounds at different levels into the NIST SRM 1950. The ABRF reported an 88.2% agreement in metabolite identification between platforms and a 33% quantitative agreement in terms of defining the fold-change in each metabolite between the two groups. Another study, the ‘metabo-ring’ initiative, also tested across several platforms (5 NMR and 11 LC-MS), using spiked samples and also real biological materials from a rodent study of vitamin D exposure. This

reported that, despite large differences in the number of spectral features produced and the heterogeneity of analytical conditions, the NMR spectral information between all platforms was very similar (average agreement of 64 to 91 %).33 The integration of semi-quantitative LC-MS lipidomic data generated in samples from three different large biobanks acquired in the time course of 3 years has been described.34 Moreover, in another study35 a high intra- and inter-laboratory reproducibility was reported of UHPLC-TOF-MS for urinary metabolic profiling. A total of 14 stable isotope labeled standard compounds were spiked into a pooled human urine sample, in dilution series and analytical features such as retention time drift, mass accuracy, signal intensity and adduct formation were evaluated. Recently, the inter-laboratory robustness of the Biocrates® Bile Acids Kit in human and mouse plasma, involving 12 laboratories was reported36. Our study is currently unique in that we report full quantitative reproducibility of a widely-used targeted metabolomics platform, tested using both spiked QCs and a varied set of normal human plasma/serum.

The protocol used here consisted of a single sample processing procedure, however it requires two separate MS analytical runs, the LC-MS/MS and the FIA-MS/MS analysis of metabolites. LC offers the advantage of greater selectivity and lower susceptibility to matrix effects. However, for low intensity peaks (low concentration samples) visual inspection, optimisation of integration and manual integration may be necessary to get accurate results, as automatic peak integration software may select the incorrect integral. Ambiguous integration is also an issue for peaks with imperfect shapes e.g. split peaks or tailing peaks. One of the conclusions of our study is that a thorough visual inspection of the integration of the LC-MS/MS peaks is required for optimum precision. This suggests that specific guidance should be provided for targeted LC-MS protocols to support correct peak identification as was the case for Asn, Thr, Ac-Orn, trans- and cis-OH-Pro, carnosine and sarcosine in the manufacturer's protocol used here.22 However, anecdotal evidence from the participating laboratories suggests that peak ambiguity was a problem for other metabolites for which guidance was not provided (Asp and Glu, and several biogenic amines) and that there was variability in how this was addressed at each centre. This may be reflected in the observation that 27/59 metabolites in the spiked QC samples showed an inaccuracy >30% in at least one sample for at least one laboratory. Unavoidably therefore, the overall quality of any targeted LC-MS metabolomics platform depends in part on the time spent on post-acquisition review of the data and the experience and skills of the analysts in this regard.

Many of metabolites that were difficult to integrate reliably from LC-MS/MS (e.g. histamine, carnosine, DOPA, Dopamine, Nitro-Tyr, c4-OH-Pro and PEA) were also present at around or just below the limits of detection. In general for any assay, this will make analysis more prone to minor instrument and laboratory variations, errors with peak integration, data processing and fit to calibration curves. These metabolites are likely to be excluded from most studies of healthy individuals, although for some pathological conditions or different tissues, the levels of some could be elevated and become more readily detectable. A further source of variability is the definition of LOD itself. LODs can be calculated from repeated injection of a blank sample within each run (in our case phosphate-buffered saline) or a historical value can be applied for a given instrument. In our study LOD was defined by the former approach, but this may lead to problems where the blank signal is improperly integrated, which could be the case for LC data. An alternative is to use repeated

injection of a blank sample for the FIA assay, while for LC data the analyst could make an informed decision based on in-house data. However, this is clearly also prone to inter-laboratory variation and further work is required to resolve this issue.

While the integration of peaks is unambiguous for any FIA detection method, variation in background, selectivity and matrix effects do remain and are likely to be a major source of inter-laboratory variability. In our dataset it was necessary to normalise FIA-metabolite data in order to get acceptable inter-laboratory precision, particularly for sphingolipids. Our study indicates how normalisation to measurements made in each laboratory on a common standard reference material is a vital step making data comparable between laboratories where the output is considered largely ‘semi-quantitative’. While a single, commercially-available reference material was used here (the NIST SRM 1950 plasma)24 using multiple QCs and reference materials could improve the robustness of this approach and comparability of data. Also QCs in 3 different levels are provided with the kit and we have demonstrated that using them for normalisation can improve the comparability of data obtained from different laboratories. Most metabolomics laboratories will use alternative pooled QC samples for assessing long-term platform stability. Exchange of these between metabolomics researchers and public dissemination of measurements made on these at multiple sites would make a significant impact on the inter-laboratory comparability of metabolomics data, benefitting the entire field.

Some important limitations to the simple normalisation approach we have used should be stated. Firstly it does not account for matrix effects between samples. A clear illustration of this was the effect of the high lipid sample, which significantly increased the inter-laboratory CV of several lipids. Secondly it does not take into account the underlying causes of error. For example corrections can be made for overlapping isotope peaks; while this is implemented in the protocol used, improvements can be continually made as the understanding of a metabolomics assay improves. It is interesting to note also that the FIA part of the present method uses a single point calibration curve while the LC part of the assay uses a seven point isotope dilution calibration curve for quantification. This is likely to have large influence on the susceptibility to matrix effects. Our study had several other general limitations. On average each sample was replicated 3-5 times; ideally more repeats would have minimized the impact of missing data. In our analyses we have focused on direct measurements and have not considered metabolite ratios, which are an important set of biomarkers for researchers using this platform.11

Conclusions

This is the first inter-laboratory assessment of a widely used targeted metabolomics platform for human serum and plasma, illustrating the reproducibility of the protocol and providing critical information for users to interpret such data appropriately. Data generated on a range of biological test materials using this targeted metabolomics kit are highly reproducible across multiple laboratories. Sets of metabolites likely to require manual integration review and/or likely to fall below limits of detection were identified, and these were the major sources of irreproducibility between laboratories. Normalisation to a reference material substantially improved the reproducibility of FIA-based, largely ‘semi-quantiative’

measurements. We recommend the routine use of common, well-characterised reference materials within laboratories, such as the NIST 1950 SRM, and the public exchange of these data to facilitate comparability and integration of metabolomics datasets. High lipids affect the precision of the assay and further work should be carried out to assess the impact on studies of patients with dyslipidaemia or other conditions that could alter lipid levels. The specific instrumentation used, notably the use of UHPLC or HPLC, had a minor effect on comparability of data although we were limited to just two main MS platforms. Our work demonstrates that human metabolomics data generated in different laboratories using this platform can be directly combined with minimal pre-treatment facilitating large-scale integrated studies. Such meta-analyses of existing cohorts enhance the return on investment for often very laborious and costly biomarker studies, and will provide unprecedented power to detect novel associations between the human serum/plasma metabolome and disease, the exposome or genome in the years to come.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the six participating laboratories: Imperial College London - UK, The Institute of Cancer Research - UK, MRC Human Nutrition Research – Cambridge – UK, Helmholtz Zentrum München - Germany, International Agency for Research on Cancer - France, BIOCRATES Life Sciences AG – Austria. Work in the HK lab (APS/PJ) is supported by the FP7 projects, EuroMotor (grant agreement no. 259867), HECATOS (grant agreement no. 602156), HELIX (grant agreement no. 603864), DETECTIVE (grant agreement no. 266838). We also acknowledge valuable discussions with Dr Timothy M.D Ebbels. Work in the JLG lab is supported by grants from the Medical Research Council (MC_UP_A090_1006, MC_PC_13030 and MRC Omics call (MC_PC_13046)). We thank Julia Scarpa and Katharina Faschinger for metabolomics measurements performed at the Helmholtz Zentrum München, Genome Analysis Center, Metabolomics Core Facility. Kits for the analyses were provided by Biocrates Life Sciences AG to all laboratories participating to the study with the exception of IARC which purchased its own kit. We thank Dr Kristaps Klavins, Bettina Burkard, Dr Therese Koal, Dr Manuel Kratzke, and Dr Markus Langsdorf at BIOCRATES Life Sciences AG. All statistical analyses and the writing of the publication were conducted independently from the manufacturer of the kit, with whom the manuscript has only been shared after submission for publication in this journal.

References

(1). Nicholson JK, Lindon JC, Holmes E. Xenobiotica. 1999; 29:1181–1189. [PubMed: 10598751]

(2). Fiehn O. Plant Mol Biol. 2002; 48:155–171. [PubMed: 11860207]

(3). Kühn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W, von Bergen M, Boeing H, Kaaks R. BMC Med. 2016; 14:13. [PubMed: 26817443]

(4). Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, et al. Mol Syst Biol. 2012; 8:615. [PubMed: 23010998]

(5). Suhre K, Meisinger C, Döring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, Mewes HW, et al. PLoS One. 2010; 5

(6). Ang JE, Pandher R, Ang JC, Asad YJ, Henley AT, Valenti M, Box G, de Haven Brandon A, Baird RD, Friedman L, Derynck M, et al. Mol Cancer Ther. 2016; 15:1412–1424. [PubMed: 27048952]

(7). Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, Onland Moret NC, Scalbert A, Rinaldi S, Ferrari P. PLoS One. 2015; 10

(8). Schmidt JA, Rinaldi S, Scalbert A, Ferrari P, Achaintre D, Gunter MJ, Appleby PN, Key TJ, Travis RC. Eur J Clin Nutr. 2016; 70:306–312. [PubMed: 26395436]

(9). Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, Achaintre D, Assi N, Tjønneland A, Overvad K, Bastide N, et al. Int J Cancer. 2016; 138:348–360. [PubMed: 26238458]

(10). Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG, Fritsche A, Häring HU, Hrabě de Angelis M, Peters A, Roden M, et al. Diabetes. 2013; 62:639–648. [PubMed: 23043162]

(11). Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes HW, Meitinger T, et al. Nat Genet. 2010; 42:137–141. [PubMed: 20037589]

(12). Jourdan C, Petersen AK, Gieger C, Döring A, Illig T, Wang-Sattler R, Meisinger C, Peters A, Adamski J, Prehn C, Suhre K, et al. PLoS One. 2012; 7

(13). Anton G, Wilson R, Yu ZH, Prehn C, Zukunft S, Adamski J, Heier M, Meisinger C, Römisch-Margl W, Wang-Sattler R, Hveem K, et al. PLoS One. 2015; 10

(14). Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, Banning F, Seissler J, Grallert H, Adamski J, Lechner A. PLoS One. 2014; 9

(15). Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C, Mendes J, Wahl S, Roemisch-Margl W, Ceglarek U, Polonikov A, et al. PLoS One. 2011; 6

(16). Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T. PLoS One. 2011; 6

(17). Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Römisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, et al. Aging Cell. 2012; 11:960–967. [PubMed: 22834969]

(18). Floegel A, Wientzek A, Bachlechner U, Jacobs S, Drogan D, Prehn C, Adamski J, Krumsiek J, Schulze MB, Pischon T, Boeing H. Int J Obes (Lond). 2014; 38:1388–1396. [PubMed: 24608922]

(19). Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, Perstorfer K, Jaremek M, Roemisch-Margl W, Rathmann W, Li Y, et al. BMC Med. 2013; 11:60. [PubMed: 23497222]

(20). Davies SK, Ang JE, Revell VL, Holmes B, Mann A, Robertson FP, Cui N, Middleton B, Ackermann K, Kayser M, Thumser AE, et al. Proc Natl Acad Sci USA. 2014; 111:10761–10766. [PubMed: 25002497]

(21). Nicholson G, Rantalainen M, Li JV, Maher AD, Malmodin D, Ahmadi KR, Faber JH, Barrett A, Min JL, Rayner NW, Toft H, et al. PLoS Genet. 2011; 7

(22). Biocrates Life Sciences AG. User manual: UM_p180_AB SCIEX_9; 2014, User manual: UM_p180_Waters_5; 2014, Analytical specifications: AS_p180_4; 2014. Innsbruck, Austria: 2014.

(23). Phinney KW, Ballihaut G, Bedner M, Benford BS, Camara JE, Christopher SJ, Davis WC, Dodder NG, Eppe G, Lang BE, Long SE, et al. Anal Chem. 2013; 85:11732–11738. [PubMed: 24187941]

(24). Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG, Bearden DW, Schock TB, Tchekhovskoi DV, Blonder N, et al. Anal Chem. 2013; 85:11725–11731. [PubMed: 24147600]

(25). National Institute of Standards & Technology. Certificate of Analysis, Standard Reference Material 1950: Metabolites in Human Plasma. Certificate Issue Date: 07 November 2012

(26). Keun HC, Ebbels TM, Antti H, Bollard ME, Beckonert O, Schlotterbeck G, Senn H, Niederhauser U, Holmes E, Lindon JC, Nicholson JK. Chem Res Toxicol. 2002; 15:1380–1386. [PubMed: 12437328]

(27). Ward JL, Baker JM, Miller SJ, Deborde C, Maucourt M, Biais B, Rolin D, Moing A, Moco S, Vervoort J, Lommen A, et al. Metabolomics. 2010; 6:263–273. [PubMed: 20526352]

(28). Viant MR, Bearden DW, Bundy JG, Burton IW, Collette TW, Ekman DR, Ezernieks V, Karakach TK, Lin CY, Rochfort S, de Ropp JS, et al. Environ Sci Technol. 2009; 43:219–225. [PubMed: 19209610]

(29). Alwood JW, Erban A, de Koning S, Dunn WB, Luedemann A, Lommen A, Kay L, Löscher R, Kopka J, Goodacre R. Metabolomics. 2009; 5:479–496. [PubMed: 20376177]

(30). Piccinonna S, Ragone R, Stocchero M, Del Coco L, De Pascali SA, Schena FP, Fanizzi FP. Food Chem. 2016; 199:675–683. [PubMed: 26776024]

(31). Klavins K, Neubauer S, Al Chalabi A, Sonntag D, Haberhauer-Troyer C, Russmayer H, Sauer M, Mattanovich D, Hann S, Koellensperger G. Anal Bioanal Chem. 2013; 405:5159–5169. [PubMed: 23604417]

(32). Cheema AK, Asara JM, Wang Y, Neubert TA, Tolstikov V, Turck CW. J Biomol Tech. 2015; 26:83–89. [PubMed: 26290656]

(33). Martin JC, Maillot M, Mazerolles G, Verdu A, Lyan B, Migné C, Defoort C, Canlet C, Junot C, Guillou C, Manach C, et al. Metabolomics. 2015; 11:807–821. [PubMed: 26109925]

(34). Dane AD, Hendriks MM, Reijmers TH, Harms AC, Troost J, Vreeken RJ, Boomsma DI, van Duijn CM, Slagboom EP, Hankemeier T. Anal Chem. 2014; 86:4110–4114. [PubMed: 24650176]

(35). Benton HP, Want E, Keun HC, Amberg A, Plumb RS, Goldfain-Blanc F, Walther B, Reily MD, Lindon JC, Holmes E, Nicholson JK, et al. Anal Chem. 2012; 84:2424–2432. [PubMed: 22304021]

(36). Pham HT, Arnhard K, Asad YJ, Deng L, Felder TK, John-Williams LS, Kaever V, Leadley M, Mitro N, Muccio S, Prehn C, et al. JALM. 2016

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

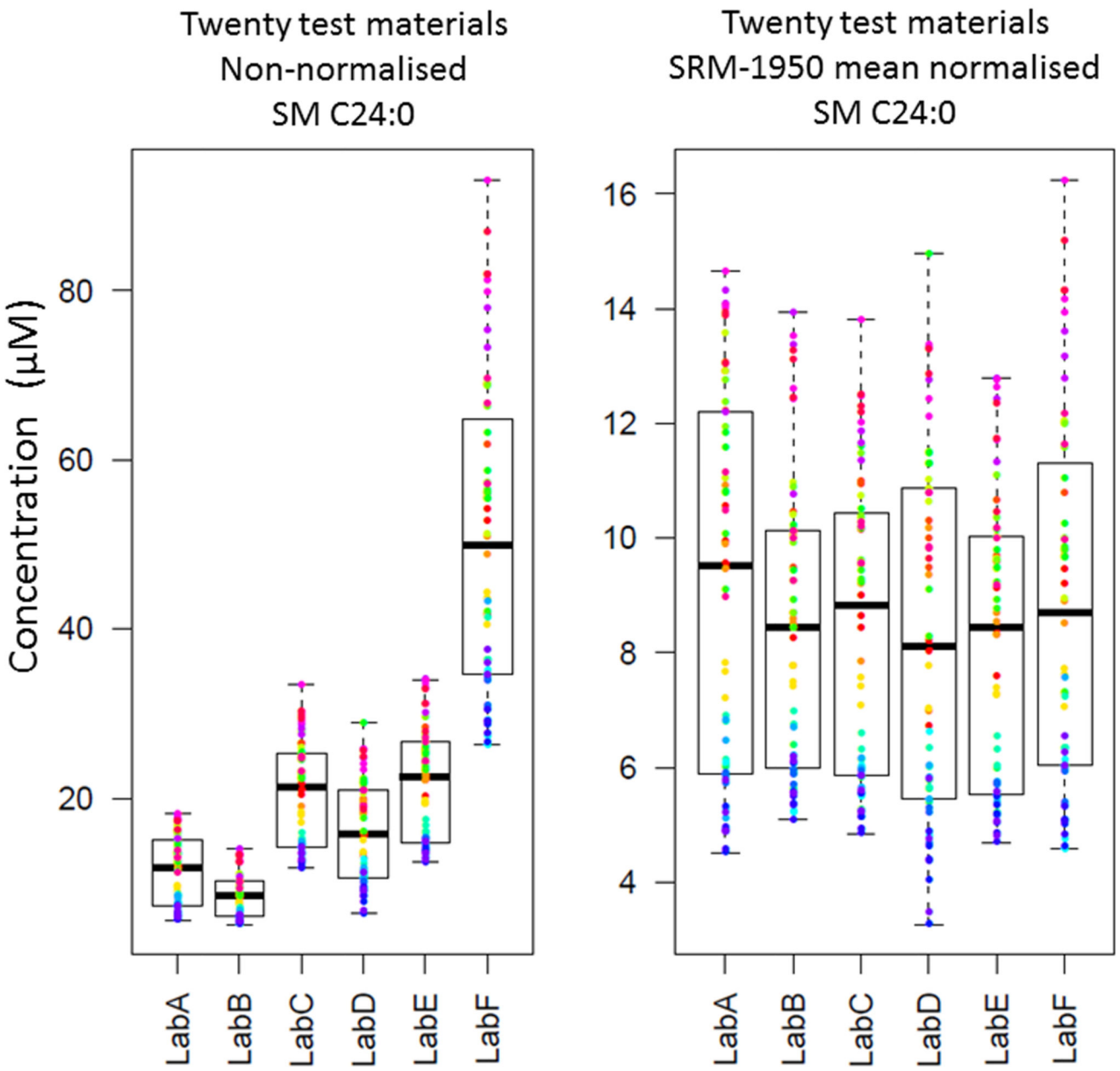

Figure 1.

An example of the effect of normalisation using the NIST SRM 1950 (TS-06), to the reported values of SM C24:0 by each laboratory, for our primary test set of 20 test materials.

Europe PMC Funders Author Manuscripts

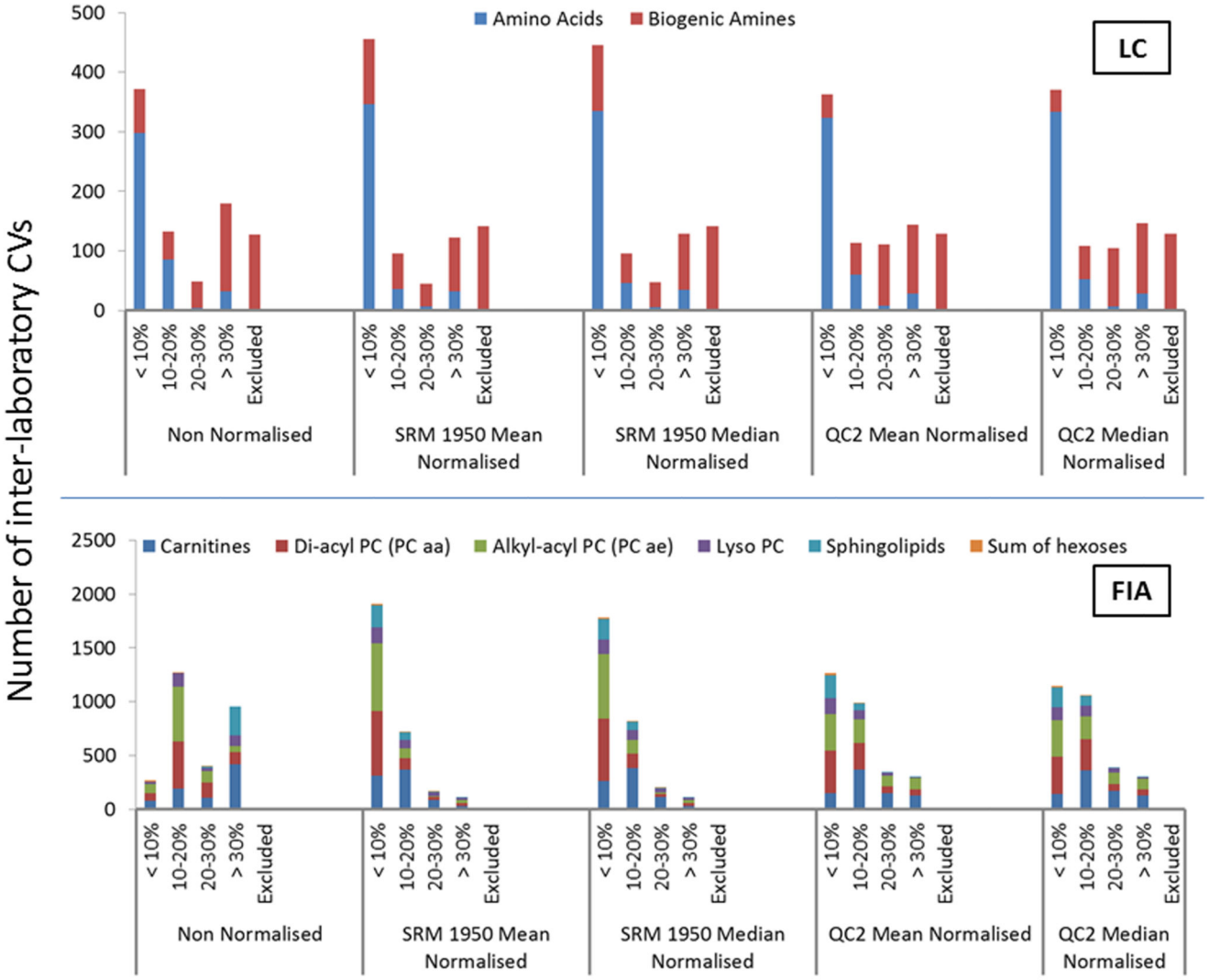

Figure 2.

Distribution of inter-lab %CVs for the 20 test material and 189 metabolites (total 3780 inter-lab %CV values), depicted per metabolite class, for the LC-assay and the FIA-assay. Data is shown for the non-normalised data, the data normalised to the mean or median values from SRM 1950 and data normalised to the mean or median values from QC level 2 provided with the kit.

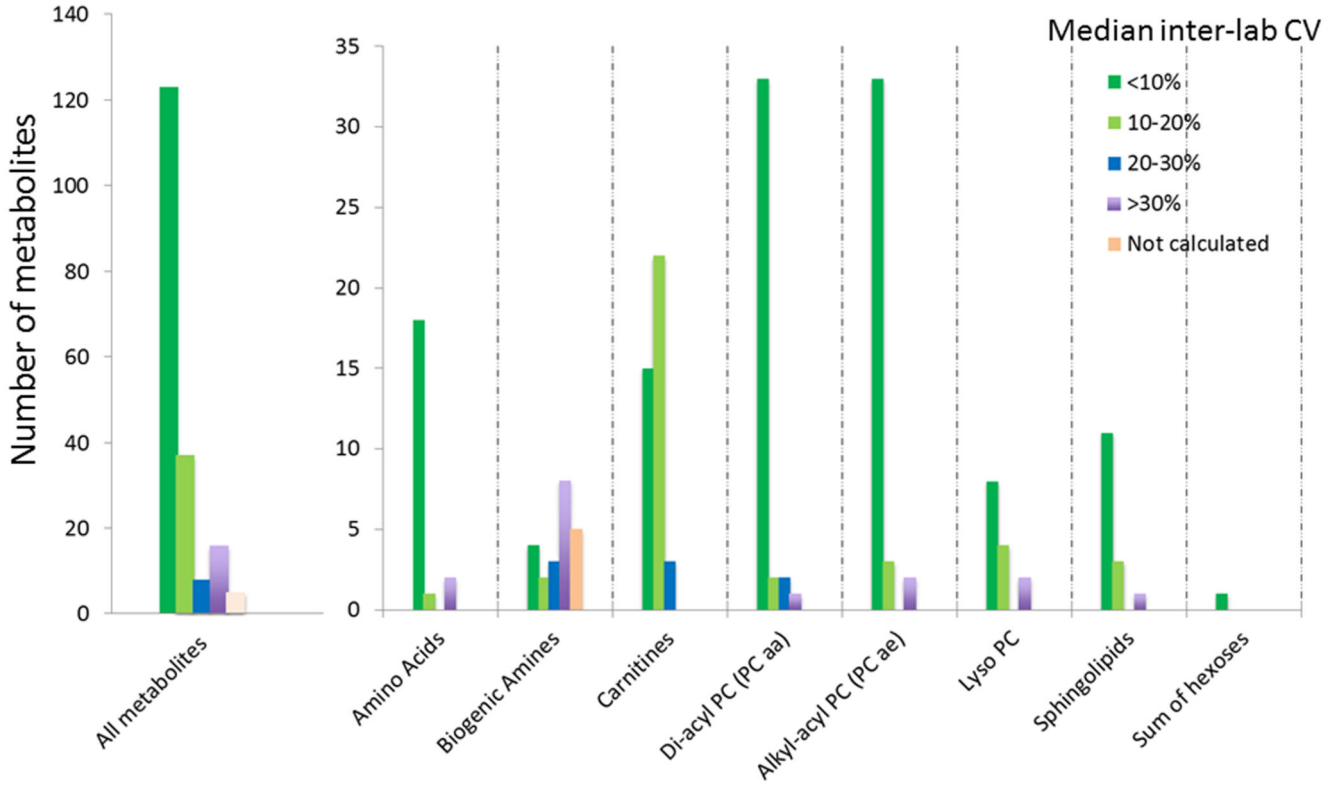

Figure 3.

Distribution of the median inter-laboratory precision (%CV) of the 189 metabolites for six laboratories and 20 biological test materials, and also depicted per metabolite class.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

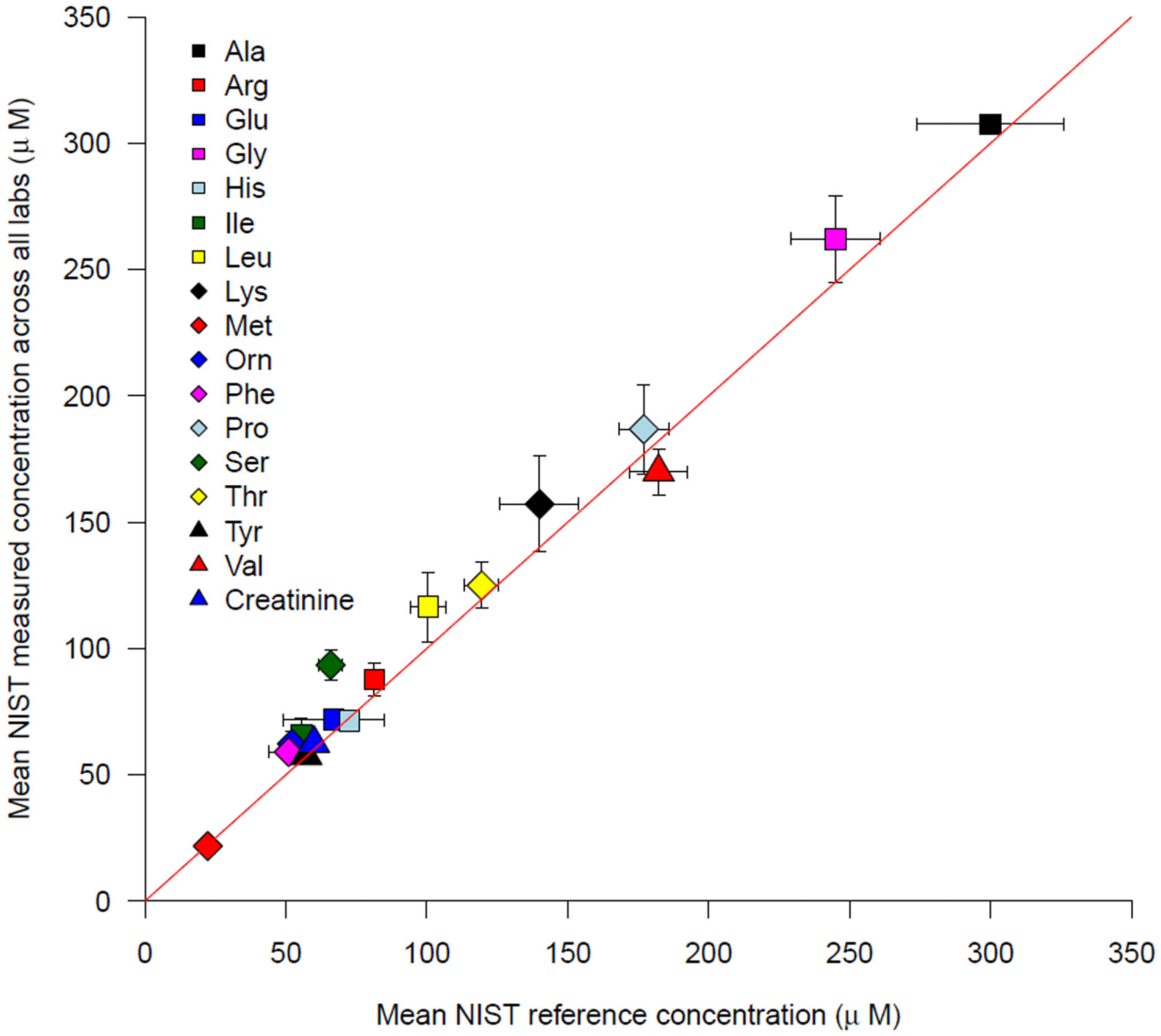

Figure 4.

Accuracy comparison of the estimated concentrations measured by the Absolute*IDQ*™ p180 Kit with the NIST reference values. The mean accuracy across all laboratories was used. Error bars represent the SD around the mean.

Europe PMC Funders Author Manuscripts

Table 1

Median inter-laboratory %CV (prior and post normalisation), and total % BLD and % missing values, for the 20 test material and 189 metabolites. The column indicating missing includes the total of not acquired data ‘NQ’, ‘NA’ values, zero values and outliers.

Metabolite class	Median InterLab %CV [90%CI] Non-Normalised	Median InterLab %CV [90%CI] Normalised	% total (BLD+missing)
Amino Acids	7.1 [4.5-13]	-	1.3
Biogenic Amines	27 [6.6-100]	-	43.4
Carnitines	34 [11-61]	12 [7.2-18]	48.0
Di-acyl PC (PC aa)	17 [11-36]	6.4 [4.8-12]	6.7
Alkyl-acyl PC (PC ae)	15 [10-29]	5.9 [4.8-11]	4.0
Lyso PC	18 [12-66]	8.1 [6.5-28]	10.4
Sphingolipids	68 [37-120]	6.7 [5.1-14]	5.3
Sum of hexoses	9.3 [7.5-16]	6.3 [3.9-13]	0.8
Total	**18 [8.2-61]**	**7.6 [5-27]**	**18.7**