analytical chemistry

Article

pubs.acs.org/ac

Liquid Chromatography–Mass Spectrometry Calibration Transfer and Metabolomics Data Fusion

Andrew A. Vaughan,*[,†] Warwick B. Dunn,[‡,§] J. William Allwood,[†] David C. Wedge,[†,∥] Fiona H. Blackhall,[⊥] Anthony D. Whetton,[#] Caroline Dive,[⊥] and Royston Goodacre[†,§]

[†]School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[‡]Centre for Advanced Discovery and Experimental Therapeutics (CADET), Central Manchester NHS Foundation Trust and School of Biomedicine, University of Manchester, Manchester Academic Health Science Centre, York Place, Oxford Road, Manchester, M13 9WL, United Kingdom

[§]Manchester Centre for Integrative Systems Biology, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

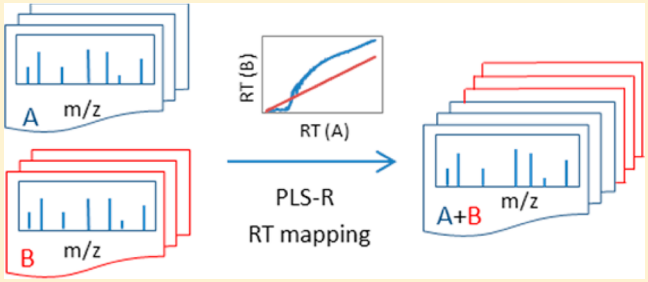
[∥]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom

[⊥]Clinical and Experimental Pharmacology Group, Paterson Institute for Cancer Research and Manchester Cancer Research Centre (MCRC), Manchester Academic Health Science Centre, University of Manchester, Wilmslow Road, Withington, Manchester, M20 4BX, United Kingdom

[#]School of Cancer and Enabling Sciences, Manchester Academic Health Science Centre, University of Manchester, Manchester, M20 3LJ, United Kingdom

Supporting Information

ABSTRACT: Metabolic profiling is routinely performed on multiple analytical platforms to increase the coverage of detected metabolites, and it is often necessary to distribute biological and clinical samples from a study between instruments of the same type to share the workload between different laboratories. The ability to combine metabolomics data arising from different sources is therefore of great interest, particularly for large-scale or long-term studies, where samples must be analyzed in separate blocks. This is not a trivial task, however, due to differing data structures, temporal variability, and instrumental drift. In this study, we employed blood serum and plasma samples collected from 29 subjects diagnosed with small cell lung cancer and analyzed each sample on two liquid chromatography–mass spectrometry (LC-MS) platforms. We describe a method for mapping retention times and matching metabolite features between platforms and approaches for fusing data acquired from both instruments. Calibration transfer models were developed and shown to be successful at mapping the response of one LC-MS instrument to another (Procrustes dissimilarity = 0.04; Mantel correlation = 0.95), allowing us to merge the data from different samples analyzed on different instruments. Data fusion was assessed in a clinical context by comparing the correlation of each metabolite with subject survival time in both the original and fused data sets: a simple autoscaling procedure (Pearson's R = 0.99) was found to improve upon a calibration transfer method based on partial least-squares regression (R = 0.94).



In metabolomics, liquid chromatography coupled to mass spectrometry (LC-MS) is becoming a well-established technique for targeted[1,2] and untargeted studies[3,4] and includes the recent advanced techniques of ultraperformance LC (UPLC) and ultrahigh performance LC (UHPLC).[5] These platforms provide appropriate levels of sensitivity (typically low μmol/L limits of detection) and, via the use of different LC stationary phases, the ability to detect metabolites with a wide range of different physicochemical properties. Reversed phase LC is applied to the study of hydrophobic metabolites, including lipids,[6] whereas hydrophilic interaction LC (HILIC) is applied to the study of hydrophilic metabolites.[7] However, in metabolomics it is widely acknowledged that a single analytical platform is unable to detect all of the metabolites in a biological sample and it is therefore necessary to use multiple analytical platforms to increase the coverage of detected metabolites.[3] Other platforms commonly used for metabolomic studies include gas chromatography coupled to MS (GC-MS)[8] and nuclear magnetic resonance (NMR) spectroscopy.[9] The ability to combine metabolomics data arising from different sources is

Received: August 3, 2012
Accepted: October 16, 2012
Published: October 16, 2012

therefore necessary, particularly for large-scale[3,10] and long-term studies,[11] where 100−1000s of samples are collected and analyzed in multiple analytical experiments (blocks), typically spanning months or years. The integration of multiblock data acquired on the *same* instrument is a challenge in itself due to temporal variability, analytical drift, and, in the case of LC-MS, changes in column chemistry or performance,[12] although recent methodological advances, including the use of quality control (QC) samples, have enabled large-scale studies to be performed with the integration of data from multiple analytical experiments.[3]

Data fusion is the process of combining data from multiple sources to achieve improved inferences than could be gained by the use of a single data source alone.[13] Three levels of fusion have been defined, depending on the processing stage at which combination occurs:[14,15] low-level fusion may be regarded as the simple concatenation of raw data matrices, which implies that the data output from the various sources must share a common structure; mid-level fusion is performed using features extracted from each source following a suitable variable selection procedure; high-level fusion occurs after each data set has been modeled separately. Fusion of metabolite data from different analytical approaches has been applied previously to the metabolomic study of plants,[16] microorganisms,[17] and human biofluids such as urine[18] and blood plasma.[19] A mid-level approach has been reported for the fusion of metabolomics and proteomics data for biomarker discovery in the cerebrospinal fluid of rats,[20] and applications within a general systems biology framework have been evaluated.[21,22] The fusion of quantitative data sets, such as those derived from systems biology, is a simpler and more robust process than the fusion of data from untargeted metabolomics studies where the data are not quantitative and the number of variables can be significantly higher.

Examples of metabolomics data fusion reported in the literature have largely been concerned with combining data for (1) the same samples, but from different analytical platforms such as LC-MS and GC-MS,[17] or (2) different samples measured on the same instrument, but separated by batch and time.[3,18,19] One study compared the metabolite profiles of a set of samples by coupling a *single* UPLC separation to two different types of mass spectrometer;[23] however, we are not aware of any previous studies that have investigated the metabolomic analysis of a common set of samples on two entirely separate UPLC-MS platforms. In metabolomics studies the fusion of data from different instruments is advantageous, allowing the transfer of a study from one instrument to another in the event of instrument failure or down-time. In large-scale studies requiring data acquisition over many months or years, study times can be reduced through data acquisition on multiple instruments and subsequent data fusion.

Calibration transfer is the process of creating a model to compensate for differences in data collected for the same samples on two or more different instruments, or on the same instrument at different time-points. The model, commonly based on a multivariate method such as partial least-squares regression or principal components regression,[24] can then be used to eliminate or reduce the variation in response observed between instruments. Multivariate standardization techniques such as direct standardization and reverse standardization (and their piecewise variants) use a set of (transfer) samples analyzed on both instruments to transform the response from one instrument to the other,[25] with the general idea that a 'master' instrument is used to collect most of the data in a database and then 'slave' instruments collect new data that after calibration transfer are matched against the master database. This has some similarities with using a set of pooled quality control (QC) samples to correct for intra- and interbatch drift, which has currently only been applied to data acquired on a single instrument.[3,26] Non-standardization methods are useful when a set of transfer samples is not available[24] and tackle the issue of calibration transfer through the use of data preprocessing methods such as local autoscaling.[18] Calibration transfer methods are commonly used in near-infrared spectrometry[27,28] and have also been applied to pyrolysis mass spectrometry,[29,30] but to date they have not found common usage in transferring the response from one LC-MS instrument to another.

In this study, we performed metabolite profiling on serum and plasma samples collected from 29 subjects diagnosed with small cell lung cancer. Each sample was analyzed on two separate UPLC-MS platforms. The aim of the study was to compare the response matrices from the two platforms, and then transform them so that the data from both instruments could be integrated or fused. This was achieved by developing (1) a method for mapping retention times for different reversed phase LC methods and matching metabolite features between different mass spectrometers of the same type and manufacturer, (2) a low-level fusion approach for combining the raw data for metabolite feature grouping and identification, and (3) a mid-level approach for fusing subsets of matched features for subsequent use in both standardization and non-standardization calibration transfer models. The success of the models at transforming and fusing these data is assessed by multivariate cluster analysis methods, namely, Procrustes analysis[31] and the Mantel test.[32] Procrustes analysis is a method for analyzing the distribution of two shapes or sets of points (or, in the present study, sets of samples located in multivariate space). One "shape" undergoes a linear transformation (incorporating translation, scaling, and rotation components) so that it resembles the target shape as closely as possible. The resulting goodness-of-fit value, the Procrustes dissimilarity, ranges between 0 and 1 and gives an indication of how similar the original shapes are (a value of 0 means that the shapes are identical; values close to 1 indicate that the shapes are very dissimilar). The Mantel test yields the correlation between two pairwise distance matrices. In the present study, this corresponds to the multivariate distances between matched sample pairs analyzed on two different instruments, before and after calibration transfer. These methods are particularly useful for assessing the similarity between identical samples analyzed on different analytical platforms or between different sample types (e.g., serum and plasma) taken from the same subject. We also assess the calibration transfer models in relation to their predicted correlations of metabolite levels with subject survival time. Moreover, we demonstrate the advantages to be gained from combining metabolomics data from multiple sources as applied to clinical samples.

EXPERIMENTAL SECTION

Sample Collection, Storage, and Selection. Blood samples from patients (prior to undergoing treatment with standard chemotherapy) at The Christie Hospital, Manchester, were collected, processed, logged, tracked, and stored according to standard operating procedures and Good Clinical Laboratory Practice standards. Patients gave written, informed consent to

ethically approved protocols. Blood was collected in either Monovette serum gel tubes (for processing to serum) or in Monovette Li-heparin tubes (for processing to plasma). Serum samples were left to clot for up to 120 min at room temperature (ca. 22 °C), centrifuged at 2000*g* for 10 min, and then stored at −80 °C. Plasma samples were stored at room temperature and were processed within 120 min of collection by centrifuging at 1000*g* for 10 min followed by storage at −80 °C. All samples were transported and stored using appropriate health and safety protocols[3] to eliminate or reduce risks associated with handling human blood products.

Materials. All chemicals used were of analytical reagent (AR) or a higher purity grade. HPLC grade methanol and water were purchased from Sigma-Aldrich (Gillingham, U.K.). Formic acid (BDH Aristar 1) was purchased from VWR International (East Grinstead, U.K.).

Extraction of Metabolites from Plasma and Serum. Metabolite extraction from serum and plasma was based upon methanolic deproteinization.[33] Extraction involved thawing samples on ice, transferring 400 μL aliquots to 2 mL microcentrifuge tubes (Eppendorf, Cambridge, U.K.) and adding 1200 μL of methanol (room temperature), followed by vortex mixing for 15 s and centrifugation at 13363*g* for 15 min. Four aliquots of the supernatant (each of volume 370 μL) were transferred to separate 2 mL microcentrifuge tubes and were lyophilized by speed vacuum concentration at 45 °C for 16 h (HETO VR MAXI vacuum centrifuge attached to a Thermo Svart RVT 4104 refrigerated vapor trap; Thermo Life Sciences, Basingstoke, U.K.) and stored at −80 °C prior to analysis. Pooled quality control (QC) samples were prepared by combining 100 μL aliquots from each subject sample and extracting as described above. QC samples were prepared separately for serum and plasma.

UPLC-MS Analysis. The extracted samples were reconstituted in 90 μL of water, vortex mixed, and centrifuged for 15 min at 13363*g*. Each supernatant was transferred to a single analytical vial with 200 μL fixed insert, stored in the autosampler at 5 °C, and analyzed within 48 h of reconstitution. All of the reconstituted samples were analyzed on two separate UPLC-MS instruments:

Instrument A was an Acquity UPLC system (Waters, Elstree, U.K.) coupled to a hybrid LTQ-Orbitrap XL mass spectrometer (ThermoFisher, Bremen, Germany) operating in electrospray ionization (ESI) mode. UPLC separations were performed using an Acquity UPLC BEH C_{18}, 2.1 × 100 mm, 1.7 μm column (Waters, Elstree, U.K.), and a nonlinear water−methanol gradient as follows: ESI+ (Waters curve 5) 100% solvent A (water and 0.1% formic acid) held for 1 min, 0−100% solvent B (methanol and 0.1% formic acid) over 15 min, 100% B held for 4 min, returning to 100% A over 2 min (total run time of 22 min); ESI− (Waters curve 4) 100% A held for 2 min, 0−100% B over 15 min, 100% B held for 5 min, returning to 100% A over 2 min (total run time of 24 min).

Instrument B was an Accela UPLC system (ThermoScientific, Hemel Hempstead, U.K.) coupled to a second hybrid LTQ-Orbitrap XL mass spectrometer (ThermoFisher, Bremen, Germany) operating in ESI mode. UPLC separations were performed using a Hypersil GOLD C_{18}, 2.1 × 100 mm, 1.9 μm column (FisherScientific, Loughborough, U.K.), and a linear water−methanol gradient as follows: ESI+ 100% A held for 1 min, 0−100% B over 11 min, 100% B held for 8 min, returning to 100% A over 2 min (total run time of 22 min); ESI− 100% A held for 1 min, 0−100% B over 16 min, 100% B held for 5 min, returning to 100% A over 2 min (total run time of 24 min).

Both UPLC systems operated at a flow rate of 360 μL min^{-1} in ESI+ mode and 400 μL min^{-1} in ESI− mode, and the columns were maintained at a temperature of 50 °C. Prior to the first analytical batch, new UPLC columns were conditioned for 40 min under the same initial gradient conditions applied to sample analysis. A sample injection volume of 10 μL was employed in wasteless mode. Both LTQ-Orbitrap XL MS systems were controlled under Xcalibur software (ThermoFisher Ltd. Hemel Hempstead, U.K) and were tuned to optimize conditions for the detection of ions in the *m/z* range 50−1000 (tuning parameters are shown in the SI, Table S-1). They were calibrated in both ESI polarities using the manufacturer's predefined methods and recommended calibration mixture, consisting of caffeine, sodium dodecyl sulfate, sodium taurocholate, the tetrapeptide MRFA, and Ultramark 1621. Data were acquired, without lock mass recalibration, in the Orbitrap mass analyzers operating at a mass resolution of 30000 (FWHM defined at *m/z* 400) and a scan speed of 0.4 s. Prior to each analytical batch, 10 (instrument A) or 20 (instrument B) injections of QC sample were performed for column conditioning. The QCs were then analyzed every sixth injection throughout each batch for use in subsequent intrabatch signal correction and the removal of poor quality peaks. Between analytical batches, the ESI ion tube and spray deflector were cleaned using methanol acidified with 0.1% formic acid and ultrasonication for 15 min. Serum and plasma samples were analyzed in separate analytical batches, in both ESI+ and ESI− modes, each randomized according to the available metadata on gender, body mass index, and disease severity (i.e., the number of days of survival after sample collection). The same analysis order was used for both instrument A and instrument B. The total number of analyses on each instrument, including QC samples and technical replicates, was 52 for each biofluid/ionization combination (SI, Table S-2).

Data Processing. The raw UPLC-MS data profiles were first converted into NetCDF format within the Xcalibur software's file converter program. The data for each analytical batch were then deconvolved separately using XCMS,[34] as described previously,[35] resulting in a matrix of metabolite features (with accurate *m/z* and retention time) and related peak area for each sample. Signal correction and quality control were performed, separately for each data set, using a previously reported method.[3,11] Peaks were aligned by fitting to a low-order, nonlinear locally weighted spline (LOESS). All peaks were then removed for which more than 40% of the QCs (excluding conditioning QCs) were either missing or had an area that differed from the mean area by >20%.

Matching of Metabolite Features between UPLC-MS Platforms. The retention times for each biofluid/ionization combination were mapped between instruments by (1) identifying features that had a *m/z* difference of <0.01 and correlation coefficient >0.5, (2) constructing a scatter plot of the retention times for these features (instrument A vs instrument B) and manually discarding any outliers from the resulting curve, (3) applying a linear interpolation algorithm to generate a piecewise polynomial form of the curve, and (4) evaluating the piecewise polynomial for the full set of retention times for instrument B (SI, Script S-1). Metabolite features were then matched between instruments by identifying those that had a *m/z* difference of <0.01 and a mapped retention time

dx.doi.org/10.1021/ac302227c | *Anal. Chem.* 2012, 84, 9848−9857

difference of <10 s, as used previously for serum/plasma comparisons.[33]

Metabolite Identification. Putative annotation of metabolite features was performed by applying the PUTMEDID-LCMS set of workflows[36] to the fused set of raw, mapped features from both instruments, using a mass accuracy window of 5 ppm. Metabolite feature groups were constructed using a retention time window of 5 s, a correlation coefficient of ≥0.9, and mass differences corresponding to isotopes, adducts, and other common artifacts as described previously.[37]

Chemometric Methods. Two-way analysis of variance (ANOVA) was performed on samples analyzed in triplicate to assess the relative contributions of the instrument and subject identity to the variation in observed signal for each metabolite feature. Principal components analysis (PCA) was performed on the features matched between instruments to assess the multivariate influence of the instrument and subject identity.

Calibration transfer was performed by building partial least-squares regression (PLS-R) models using between 3 and 26 contiguous samples analyzed on both instruments (transfer samples, A_{cal} and B_{cal}) and between 2 and 10 PLS components (SI, Script S-2). Each model was validated using the final 26 samples analyzed on both instruments (A_{val} and B_{val}), and the root mean squared error of validation (RMSEV) was used to select the optimal model. The selected model was then tested on the remaining samples not used for calibration or validation (A_{test} and B_{test}), and its predictive power was assessed by two multivariate pattern comparison measurements: the Procrustes dissimilarity[31,38] and the Mantel correlation.[32] The PLS-R model was also assessed within a clinical context by calculating correlation coefficients between the area of each metabolite feature (across subjects) and the survival times of the subjects for (1) the full, original sample set B, and (2) the PLS-fitted B_{test} samples fused with the original B_{cal} and B_{val} samples. These two sets of correlations were then compared to assess the effectiveness of merging data from two instruments.

All data analysis was performed in MATLAB version R2010a (The MathWorks, Inc.). Figure 1 shows a flowchart of the steps involved in data processing, calibration transfer, and data fusion.

RESULTS AND DISCUSSION

Comparison of Metabolite Features. Following data deconvolution, the number of features detected by instrument A for serum (ESI+), serum (ESI−), plasma (ESI+), and plasma (ESI−) were 5282, 3479, 6321, and 2822, respectively. For instrument B, the corresponding figures were 3152, 2689, 3406, and 2433. As can be seen, more features were detected in plasma than in serum in ESI+ for both instruments, and more features were detected in serum in ESI−; the compositional differences between plasma and serum have been discussed previously.[33] Considerably more features were detected by instrument A than instrument B, ranging from 16% more for plasma ESI− to 86% more for plasma ESI+. Figure 2 shows that the m/z distribution profiles are similar for the two instruments at lower m/z values (~50−500), with the additional metabolite features for instrument A being observed at higher m/z (>600). These data suggest that instrument A detects higher molecular weight lipid-based metabolites with a higher sensitivity than instrument B under the operating conditions employed. Subsequent QC test runs and consultation with the instrument manufacturer suggest that the lack of higher m/z ions detected by instrument B may be due to its operating with a version of the mass spectrometer's C-trap

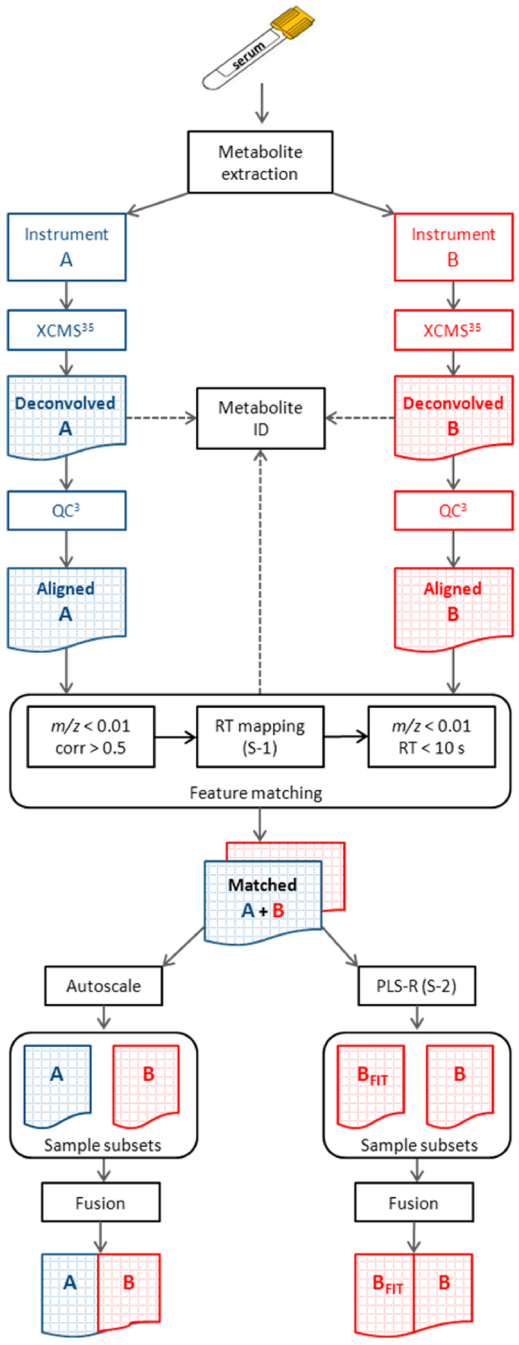


Figure 1. Flowchart of the steps involved in data processing, calibration transfer and fusion (figures in parentheses refer to scripts in the SI).

software different to that used by instrument A. Figures S-1A,B show that by adjusting the lower end of the scan range of instrument B, considerably more ions are detected at higher m/z. In general, differences in feature numbers such as these could also be a result of (1) tuning parameter differences (Table S-1), (2) applying different UPLC systems, (3) operator differences, (4) suboptimal XCMS parameter settings, or (5) the formation of dimers (which is dependent on instrument settings,[19] however, our identification procedures suggest that, in the present study, these additional features are not dimers). These results demonstrate how the metabolite profile for identical samples can vary dramatically due to the instrument

dx.doi.org/10.1021/ac302227c | Anal. Chem. 2012, 84, 9848−9857

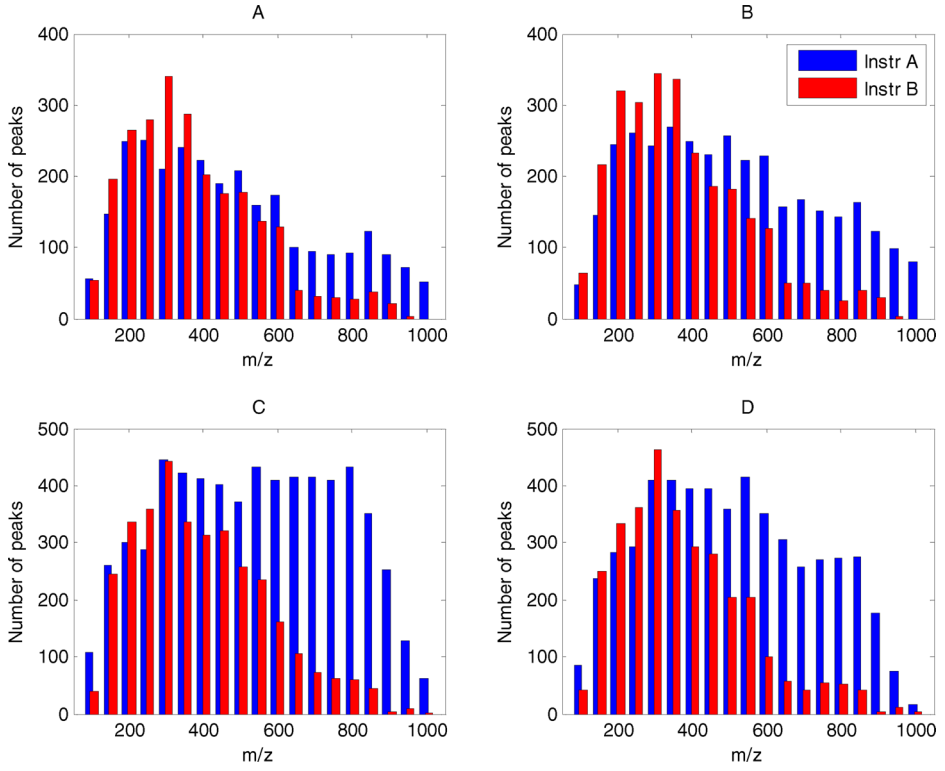


Figure 2. Distribution of features (m/z) detected by instruments A and B in (A) plasma ESI−, (B) serum ESI−, (C) plasma ESI+, and (D) serum ESI+.

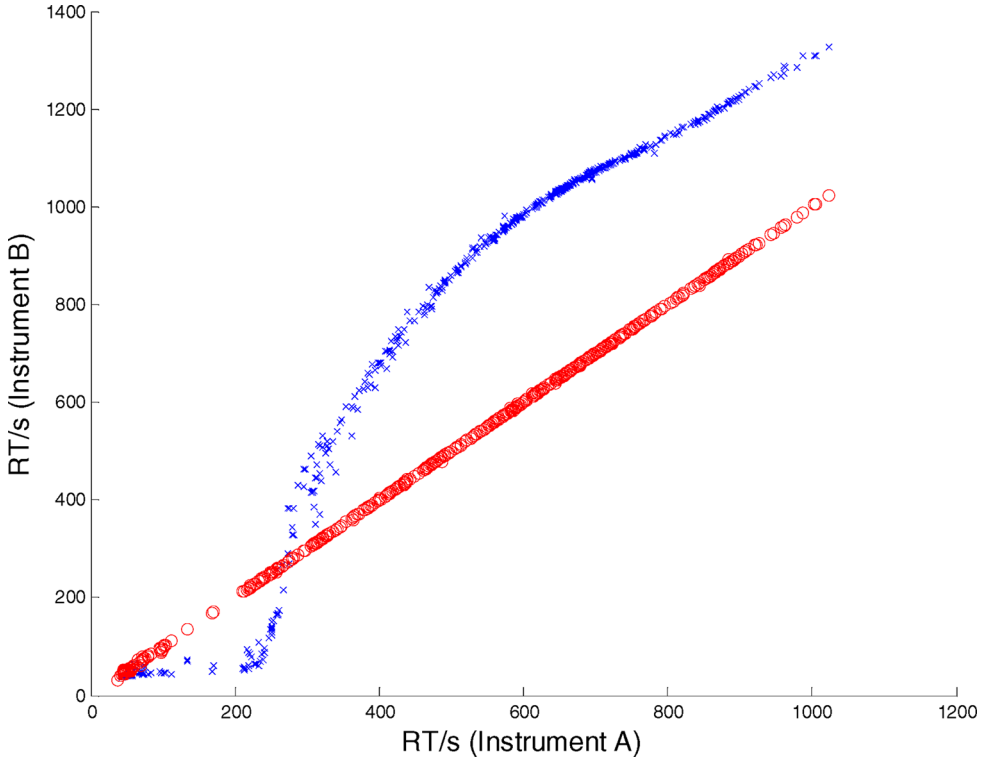


Figure 3. Retention times for features matched between instrument A and instrument B (plasma ESI−) before (blue crosses) and after (red circles, $R^2 > 0.99$) retention time mapping.

configuration and parameters used, which can present difficulties in fusing data sets from different instruments.

Technical Reproducibility. The technical reproducibility of the data acquired from each instrument was assessed by (1) calculating the pairwise Pearson correlation coefficients between the analytical triplicates for each of six samples and (2) calculating the relative standard deviation (RSD) for each metabolite feature for each set of replicates. Reproducibility was

dx.doi.org/10.1021/ac302227c | *Anal. Chem.* 2012, 84, 9848−9857

high and virtually identical between instruments; averaged across all biofluid/ionization combinations, the pairwise correlation between replicates was >0.99 for both instruments, and the median (lower quartile, upper quartile) RSD was 6% (3%, 10%) for instrument A and 6% (4%, 10%) for instrument B.

Matching of Metabolite Features between Data Sets. Following quality control procedures[3] (which removed between 5 and 10% of the deconvolved features), the numbers of metabolite features matched between instruments for serum (ESI+), serum (ESI−), plasma (ESI+) and plasma (ESI−) were 1755, 1261, 1851, and 1100, respectively. This represents between 49% (plasma ESI−) and 61% (serum ESI+) of the features, after QC, detected by instrument B. The inability to match other features is a result of (1) the strict retention time matching criterion employed (mapped RT difference <10 s), (2) the different sensitivities of the two instruments, and (3) the formation of different ion types for the same metabolite on different instruments. Figure 3 shows a plot of the retention times of the matched features measured for plasma (ESI−) on instrument A versus instrument B, before and after retention time mapping. This shows that the retention times for matched features differ by up to 400 s between the two instruments; for example, features with retention times ∼600 s on instrument A have retention times ∼1000 s on instrument B. These differences in retention times are due to the two instruments using different elution gradients (linear vs nonlinear), different UPLC columns and having different system dead volumes. The retention time profiles for the matched metabolite features were identical for serum and plasma, but differed between ionization polarity due to the different gradient elution programs used for ESI+ and ESI−, as shown in the SI (Figure S-2). Mapping the retention times from instrument B to instrument A resulted in a linear profile across the entire LC separation. Mapping the retention times in this way allowed us to exploit the difference in sensitivities observed across the m/z range for the two instruments by merging the two UPLC-MS data sets to create a single data set with greater metabolite coverage than either data set alone. Figure S-3 shows that, despite employing a m/z tolerance of <0.01 for matching features between instruments, 88% of the total matched features differed by a m/z of <5 ppm (corresponding to an absolute error of 0.00025 at m/z 50, and 0.005 at m/z 1000). These low mass errors provide confidence that the features identified by our matching procedures do correspond to the same ions.

During the metabolite identification workflows, some of the unmatched metabolite features from each data set were found to complement each other and assisted in reducing the number of false putative annotations by eliminating inaccurate matching of peaks.[36] For example, the H^+ and Na^+ adducts of indoleacetic acid were detected by both instruments, but ^{13}C isotope peaks of these adducts were unique to each instrument (MH^+ isotope detected on A only; MNa^+ isotope detected on B only). Our feature matching and identification procedures ensured that all of these metabolite features were grouped together, thus increasing our confidence in the identification of the parent molecular ion and preventing, in this example, the isotopes being falsely identified as other metabolites. These data show that different ion types (or the same ion types at different relative ratios) can be created in identical instruments from the same manufacturer.

Multivariate Analysis of Matched Metabolite Features. PCA was performed on the metabolite features matched between instruments following normalization to total peak area and with (1) A and B samples autoscaled[39] together and (2) A and B samples autoscaled separately. The resulting scores plots for plasma ESI− (Figure S-4A) show that the A and B samples are clearly separated by the first principal component when they are combined *before* autoscaling. However, when they are combined *after separate* autoscaling, there is considerable intersubject variability, yet the samples from the same subject are more similar to each other than they are to other subject samples analyzed on the same instrument. This demonstrates the need for transforming one or both of the data sets before fusion in order to remove the instrument-specific variation.

This multivariate similarity was measured numerically by calculating (1) the Mantel correlation coefficient and (2) the Procrustes dissimilarity, d, between the scores from sufficient PCs to explain 75% of the variance of the separately autoscaled, then combined, data sets from both instruments. The resulting correlations were >0.9, and values of d were <0.1 for all biofluid/ionization combinations (PCA scores plots shown in Figures S-4A−D). These high correlations and low values of d confirm that the subject samples cluster in a very similar way between instruments, and show that the analytical (interinstrument) variation is very low in comparison to the biological (intersubject) variation when the data sets are preprocessed separately prior to PCA.

PCA was also performed on the *full* set of metabolite features, separately for each instrument, following normalization and autoscaling. The resulting Mantel correlations for these PCA-reduced data sets were all >0.8 and the Procrustes d values were all <0.2. These values are only slightly poorer than those obtained for the matched metabolite features and indicate that there is a very good correlation between the full data matrices arising from the two instruments, despite the large difference in the number of features detected by each one.

Univariate Analysis of Matched Metabolite Features. ANOVA was performed on six samples that were analyzed in triplicate on both instruments for each biofluid/ionization combination. Averaged across all matched metabolite features and across all biofluid/ionization combinations, the percentage variance explained by the subject identity, the instrument, and the interaction effect (subject × instrument) was 81.1% (±1.1%), 1.2% (±0.2%), and 8.9% (±0.9%), respectively. The remaining 8.8% (±0.6%) variance was attributed to technical variance. These figures indicate that the subject identity has a much greater influence than the instrument and interaction effect on the variation observed for each metabolite feature and support the evidence from the multivariate analysis that the choice of instrument does not affect the ability to discriminate between subjects.

Calibration Transfer. One objective of this study was to compare the data acquired on two UPLC-MS instruments for the same set of samples. A second objective was to integrate or fuse different subsets of the data from each instrument. This scenario is representative of an intra- or interlaboratory experiment, in which the analysis of samples from the same study must be distributed across multiple instruments. To combine the resulting data, one must first use a method to transform one or more of the data sets to remove any analytical or instrument-specific variation, while retaining the important biological variation. In this study we used a multivariate calibration transfer method known as reverse standardization,[28] in which the response from instrument A (the primary instrument) was transformed to resemble that from instrument

B (the secondary instrument), using PLS-R as a calibration method. This requires a common set of transfer samples to be analyzed on both instruments. The selection of transfer samples is usually determined after data acquisition, using tools such as the Kennard-Stone algorithm,[40] which identifies samples that are representative of the entire data space (QCs would be of limited use as transfer samples because they are not fully representative of the biological variation observed). However, in this study we wished to constrain our transfer set to be a contiguous block of samples located at the midpoint of the analytical batch so that we could preserve our randomized run order and mimic an experimental situation in which the samples were divided almost equally between two instruments. Using correlation as a variable selection criterion for the PLS-R model,[41] the 500 most correlated features matched between the two instruments were used to determine the required number of transfer samples. The RMSEV curve was observed to reach a local minimum at 15 transfer samples (using 10 PLS components) for the majority of the biofluid/ionization combinations (SI, Figures S-5A–D); these model parameters were used for the prediction of the 11 remaining test samples, B_{test} from A_{test}. A schematic of the samples used for calibration transfer is shown in Figure 4.

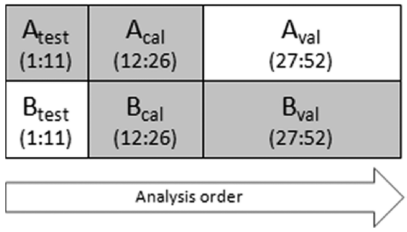


Figure 4. Schematic of samples used for calibration transfer. The shaded area shows the samples available in our experimental scenario, and used for PCA training ($_{test}$, $_{cal}$, and $_{val}$ refer to the samples used for testing, calibrating and validating the PLS-R model, respectively).

Four separate PCA models were constructed for each biofluid/ionization combination to assess the predictive power of the PLS model: (1) PCA of the 52 original A samples and the 52 original B samples, (2) PCA of the 52 original A samples and the 52 PLS-fitted B samples, (3) PCA of A_{test}, A_{cal}, B_{cal}, and B_{val} (shaded samples in Figure 4), with the original B_{test} samples projected onto the model, (4) PCA of A_{test}, A_{cal}, B_{cal}, and B_{val}, with the PLS-fitted B_{test} samples projected onto the model. The A and B samples of each model were normalized and autoscaled *separately* prior to PCA. The resulting PCA scores plots for plasma (ESI+) show that the original A/B sample pairs lie very close together (Figure 5A) and, therefore, have a low Procrustes dissimilarity (d = 0.005). The PLS-fitted B samples lie slightly further away from their original A sample pairs (Figure 5B) but are still very close (d = 0.018). The Mantel correlation between the A/B distance matrices (using 2 PCs) of these two models is 0.96, indicating that the PLS-R model is successful at predicting instrument B's response to the test samples. However, these two PCA models take into account *all* of the 52 samples analyzed on both instruments. In our mock scenario, we would only have samples A_{test}, A_{cal}, B_{cal}, and B_{val} (PCA training set; shaded in Figure 4) to input to our PCA model, so a more realistic assessment of the PLS-R model's performance is achieved by using the training set's loadings to project the test samples onto the PCA model. Figure 5C shows that the projected original B_{test} samples have a very similar multivariate pattern to the A_{test} samples and, as such, have a low Procrustes dissimilarity (d = 0.008). The projected PLS-fitted B_{test} samples (Figure 5D) have a slightly different pattern compared to the A_{test} samples but are still very similar and have a low Procrustes value (d = 0.041). The Mantel correlation between the A/B distance matrices (using two PCs) of these two models is 0.95, indicating that instrument B's predicted response for the test samples is still good, from a multivariate perspective, when a reduced sample set is used for the PCA.

Similar results were obtained for the calibration transfer of the plasma (ESI−), serum (ESI+), and serum (ESI−) data sets using the 500 most correlated features. The PCA plots are shown in the SI (Figures S-6–8) and the corresponding Procrustes dissimilarities and Mantel correlations are summarized in Table S-3. Calibration transfer was also assessed for the *full* set of features for each biofluid/ionization combination. It is clear that when more variables are included in the PLS-R model, the A/B samples cluster slightly differently (SI, Figures S-9–12). However, the multivariate patterns are still similar (Table S-3), with the lowest Mantel correlation (0.75) and highest Procrustes d (0.141) being observed for serum (ESI−).

Data Fusion Applied to Clinical Study. The success of the calibration transfer model's predictive ability was also assessed in a clinical context. As it is not possible to accurately quantify every metabolite detected in an untargeted metabolomics analysis (as it is impractical to prepare calibration curves for 100–1000s of metabolites), we used a surrogate, univariate measurement, the correlation of each feature with subject survival time, to assess the impact of fusing subsets of the data sets from each instrument. The correlation of each feature from the full original B data set with survival time was calculated, and this was then repeated for the same data set but with the B_{test} samples replaced with those predicted by PLS-R. For plasma (ESI+), the Pearson correlation coefficient between these two sets of correlations was 0.94. The Spearman correlation, which allows us to assess the rank of each peak's correlation with survival time, was 0.93 for plasma (ESI+). These results suggest that the correlations with survival time were not greatly affected by replacing 21% of the samples with PLS-fitted samples and that the same clinical conclusions could be drawn based on these data. However, it was noted in the calibration transfer that the original B_{test} samples were always more similar to the corresponding A_{test} samples than were the PLS-fitted B_{test} samples (Figure 5 and Table S-3). As a result, we performed the same survival time correlation analysis, but with the B_{test} samples replaced by the autoscaled A_{test} samples. The resulting Pearson and Spearman correlations between this fused data set and the original B data set were both 0.99 for plasma (ESI+), indicating that a simple autoscaling procedure was sufficient to transform subsets from the two original data sets in order to fuse them successfully. As an additional measure of success, the 10 features most correlated with survival time were compared between the fused and original data sets, of which 9 were matched. The autoscaling method of transformation was extended to fuse the first half of the A data set with the second half of the B data set (26 samples in each, with no overlap); the resulting Pearson and Spearman correlations between the original and fused data sets were 0.96 and 0.95, respectively, for plasma (ESI+), and 9 of the 10 features most correlated with survival time were matched. These correlations are better than those obtained for the PLS-R transfer model,

dx.doi.org/10.1021/ac302227c | *Anal. Chem.* 2012, 84, 9848–9857

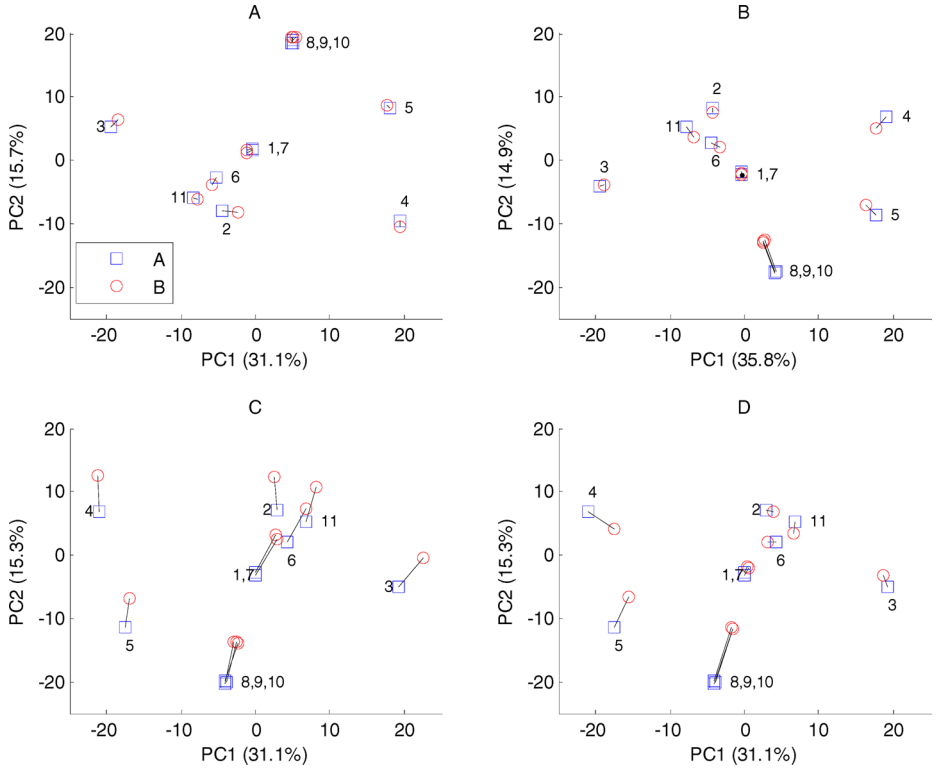


Figure 5. Scores plots for the test samples (plasma ESI+) from PCA of (A) all original A samples and all original B samples, (B) all original A samples and all PLS-fitted B samples, (C) A_{test}, A_{cal}, B_{cal}, and B_{val}, with the original B_{test} samples projected, and (D) A_{test}, A_{cal}, B_{cal}, and B_{val}, with the PLS-fitted B_{test} samples projected.

and the transformation is much simpler to implement, requiring no common transfer samples to be analyzed on both instruments. Similar results were obtained for plasma (ESI−), serum (ESI+), and serum (ESI−) and are summarized in the SI (Tables S-4 and S-5).

While the reporting of significant biomarkers of cancer is not the primary concern of this paper—the novelty lies in the data fusion methods we have developed for untargeted metabolomics—several putatively identified features were found to have modest correlations with survival time. Urea (−0.48), bilirubin (−0.52), and 3-hydroxybutanoic acid (−0.53) were found to be negatively correlated with survival time, suggesting that elevated levels of these metabolites are found in the blood of subjects with a poorer prognosis. These metabolites have long been associated with disease: urea with renal failure,[42] bilirubin with liver disorders,[43] and 3-hydroxybutanoic acid with diabetes.[44] Although these data are preliminary and need to be validated in a larger cohort study, of potential interest is the elevation of the ketone body 3-hydroxybutanoic acid. Recent data suggest a reverse Warburg effect, whereby stromal (support) cells in a tumor provide sustenance to the cancer cells for aerobic respiration.[45] This compound may afford such an opportunity. As such, the poorer prognosis seen with higher levels may be indicative of a richer nutrient environment for the cancer cells or, alternatively (because the cancer cells provoke the stromal cells to produce this type of metabolite), a greater tumor burden to influence the stroma.

CONCLUSIONS

Methods of integrating or fusing metabolomic data sets acquired on different UPLC-MS platforms have been evaluated. Retention time mapping and metabolite feature matching allowed the low-level fusion of data for subsequent metabolite identification, and the metabolite features unique to each instrument were found to complement one another by providing a greater metabolome coverage and reducing the number of false putative annotations. The reverse standardization procedure achieved favorable results in transferring the response of one instrument to the other; however, this required >25% of the samples to be analyzed on both instruments. A method of local data preprocessing, namely, autoscaling each instrument's data separately prior to mid-level fusion, was found to improve upon the predictive power of the calibration transfer procedure. This scaling procedure removes intra- and interinstrumental variations in response and is useful in situations where it is not possible or practical to analyze a set of transfer samples (although, in practice, it may be prudent to analyze a common set of QCs or reference standards in order to validate the procedure; these should be representative of, or compositionally similar to, the samples being analyzed). The ability to fuse data in this way is of great importance to large-scale metabolomic studies where there are too many samples to analyze in a single batch or where the analysis must be distributed between multiple instruments and laboratories. It may also find an application in combining data from clinical studies where, typically, smaller sample sets are collected and analyzed at different time-points, some of which can be many years apart. This would allow issues associated with sample instability during storage to be minimized and could assist with the generation of long-term metabolomics databases. This study has only dealt with metabolomics data; however, it is anticipated that a mid- or high-level fusion procedure could be used to integrate these data with proteomics data measured on the same samples and allow their investigation through a

dx.doi.org/10.1021/ac302227c | *Anal. Chem.* 2012, 84, 9848−9857

systems biology approach. In this study, our objective was to employ well-established or "classical" methods of calibration transfer and apply them, for the first time, to metabolomics data sets acquired on multiple LC-MS instruments. However, there is also scope for the investigation of more recently developed methods of interplatform data integration. Methods such as cross-platform normalization (XPN)[46] and distance weighted discrimination (DWD)[47] have been compared and evaluated in the context of combining gene expression data sets derived from multiple microarray platforms.[48] While these methods are generally effective and may be applicable to metabolomics data sets, they may not be appropriate in all cases of data fusion; a separate study using cancer microarray data sets found that these more advanced and complex methods were often outperformed by simple local data preprocessing methods such as autoscaling and batch mean-centering.[49] In the interest of developing parsimonious models for multivariate data, such simplicity is an important consideration.[50]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional figures, tables, and scripts. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: andrew.vaughan-2@manchester.ac.uk. Phone: +44 (0) 161 3065145.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by Cancer Research UK (including Experimental Cancer Medicine Centre Award) and the Wolfson Foundation. C.D. also wishes to acknowledge the generous core support from Cancer Research UK to her laboratory at the Paterson Institute for Cancer Research. A.D.W. is supported by Leukaemia Lymphoma Research, and R.G. and W.B.D. wish to thank BBSRC for financial support of The Manchester Centre for Integrative Systems Biology (BBC0082191). This work was supported by the NIHR Manchester Biomedical Research Centre. Sample collection was supported by European Union CHEMORES FP6 Contract LSHC-CT-2007-037665.

■ REFERENCES

(1) Bennette, N. B.; Eng, J. F.; Dismukes, G. C. *Anal. Chem.* **2011**, *83*, 3808−3816.

(2) Lu, W.; Bennett, B. D.; Rabinowitz, J. D. *J. Chromatogr., B* **2008**, *871*, 236−242.

(3) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; et al. *Nat. Protoc.* **2011**, *6*, 1060−1083.

(4) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *Trends Anal. Chem.* **2008**, *27*, 251−260.

(5) Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S. *J. Proteome Res.* **2005**, *4*, 591−598.

(6) Gao, X.; Zhang, Q.; Meng, D.; Isaac, G.; Zhao, R.; Fillmore, T. L.; Chu, R. K.; Zhou, J.; Tang, K.; Hu, Z.; Moore, R. J.; Smith, R. D.; Katze, M. G.; Metz, T. O. *Anal. Bioanal. Chem.* **2012**, *402*, 2923−2933.

(7) Cubbon, S.; Antonio, C.; Wilson, J.; Thomas-Oates, J. *Mass Spectrom. Rev.* **2010**, *29*, 671−684.

(8) Fiehn, O. *Trends Anal. Chem.* **2008**, *27*, 261−269.

(9) Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E. *Int. J. Epidemiol.* **2008**, *37*, 31−40.

(10) Bijlsma, S.; Bobeldijk, L.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2006**, *78*, 567−574.

(11) Zelena, E.; Dunn, W. B.; Broadhurst, D.; Francis-McIntyre, S.; Carroll, K. M.; Begley, P.; O'Hagan, S.; Knowles, J. D.; Halsall, A.; Wilson, I. D.; Kell, D. B. *Anal. Chem.* **2009**, *81*, 1357−1364.

(12) Creek, D. J.; Jankevics, A.; Breitling, R.; Watson, D. G.; Barrett, M. P.; Burgess, K. E. V. *Anal. Chem.* **2011**, *83*, 8703−8710.

(13) Hall, D. L.; Llinas, J. *Proc. IEEE* **1997**, *85*, 6−23.

(14) Steinmetz, V.; Sévilla, F.; Bellon-Maurel, V. *J. Agric. Eng. Res.* **1999**, *74*, 21−31.

(15) Roussel, S.; Bellon-Maurel, V.; Roger, J. M.; Grenier, P. *Chemom. Intell. Lab. Syst.* **2003**, *65*, 209−19.

(16) Doeswijk, T. G.; Smilde, A. K.; Hageman, J. A.; Westerhuis, J. A.; van Eeuwijk, F. A. *Anal. Chim. Acta* **2011**, *705*, 41−47.

(17) Smilde, A. K.; van der Werf, M. J.; Bijlsma, S.; van der Werff-van der Vat, B. J. C.; Jellema, R. H. *Anal. Chem.* **2005**, *77*, 6729−6736.

(18) Wagner, S.; Scholz, K.; Sieber, M.; Kellert, M.; Voelkel, W. *Anal. Chem.* **2007**, *79*, 2918−2926.

(19) Draisma, H. H. M.; Reijmers, T. H.; van der Kloet, F.; Bobeldijk-Pastorova, I.; Spies-Faber, E.; Vogels, J. T. W. E.; Meulman, J. J.; Boomsma, D. I.; van der Greef, J.; Hankemeier, T. *Anal. Chem.* **2010**, *82*, 1039−1046.

(20) Blanchet, L.; Smolinska, A.; Attali, A.; Stoop, M. P.; Ampt, K. A. M.; van Aken, H.; Suidgeest, E.; Tuinstra, T.; Wijmenga, S. S.; Luider, T.; Buydens, L. M. C. *BMC Bioinf.* **2011**, *12*, 254.

(21) Van Mechelen, I.; Smilde, A. K. *Chemom. Intell. Lab. Syst.* **2010**, *104*, 83−94.

(22) Richards, S. E.; Dumas, M.-E.; Fonville, J. M.; Ebbels, T. M. D.; Holmes, E.; Nicholson, J. K. *Chemom. Intell. Lab. Syst.* **2010**, *104*, 121−131.

(23) Gika, H. G.; Theodoridis, G. A.; Earll, M.; Snyder, R. W.; Sumner, S. J.; Wilson, I. D. *Anal. Chem.* **2010**, *82*, 8226−8234.

(24) Feudale, R. N.; Woody, N. A.; Tan, H.; Myles, A. J.; Brown, S. D.; Ferre, J. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 181−192.

(25) Wang, Y.; Veltkamp, D. J.; Kowalski, B. R. *Anal. Chem.* **1991**, *63*, 2750−2756.

(26) van der Kloet, F. M.; Jellema, R. H.; Verheij, E. R.; Bobeldijk, I. *J. Proteome Res.* **2009**, *8*, 5132−5141.

(27) Abdelkader, M. F.; Cooper, J. B.; Larkin, C. M. *Chemom. Intell. Lab. Syst.* **2012**, *110*, 64−73.

(28) Pereira, C. F.; Pimentel, M. F.; Galvao, R. K. H.; Honorato, F. A.; Stragevitch, L.; Martins, M. N. *Anal. Chim. Acta* **2008**, *611*, 41−47.

(29) Goodacre, R.; Kell, D. B. *Anal. Chem.* **1996**, *68*, 271−280.

(30) Goodacre, R.; Timmins, E. M.; Jones, A.; Kell, D. B.; Maddock, J.; Heginbothom, M. L.; Magee, J. T. *Anal. Chim. Acta* **1997**, *348*, 511−532.

(31) Gower, J. C. *Psychometrika* **1975**, *40*, 33−51.

(32) Mantel, N. *Cancer Res.* **1967**, *27*, 209−220.

(33) Wedge, D. C.; Allwood, J. W.; Dunn, W.; Vaughan, A. A.; Simpson, K.; Brown, M.; Priest, L.; Blackhall, F. H.; Whetton, A. D.; Dive, C.; Goodacre, R. *Anal. Chem.* **2011**, *83*, 6689−6697.

(34) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.

(35) Dunn, W. B.; Broadhurst, D.; Brown, M.; Baker, P. N.; Redman, C. W. G.; Kenny, L. C.; Kell, D. B. *J. Chromatogr., B* **2008**, *871*, 288−298.

(36) Brown, M.; Wedge, D. C.; Goodacre, R.; Kell, D. B.; Baker, P. N.; Kenny, L. C.; Mamas, M. A.; Neyses, L.; Dunn, W. B. *Bioinformatics* **2011**, *27*, 1108−1112.

(37) Brown, M.; Dunn, W. B.; Dobson, P.; Patel, Y.; Winder, C. L.; Francis-McIntyre, S.; Begley, P.; Carroll, K.; Broadhurst, D.; Tseng, A.; Swainston, N.; Spasic, I.; Goodacre, R.; Kell, D. B. *Analyst* **2009**, *134*, 1322−1332.

(38) Andrade, J. M.; Gomez-Carracedo, M. P.; Krzanowski, W.; Kubista, M. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 123−132.

(39) Goodacre, R.; Broadhurst, D.; Smilde, A.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton, J.; Paternostro, G.; Somorjai, R.; Sjöström, M.; Trygg, J.; Wulfert, F. *Metabolomics* **2007**, *3*, 231−241.
(40) Kennard, R. W.; Stone, L. A. *Technometrics* **1969**, *11*, 137−148.
(41) Hoskuldsson, A. *Chemom. Intell. Lab. Syst.* **2001**, *55*, 23−38.
(42) Berlyne, G. M. *Lancet* **1966**, *1*, 1212−1213.
(43) Watson, D.; Rogers, J. A. *J. Clin. Pathol.* **1961**, *14*, 271−8.
(44) Miles, J. M.; Gerich, J. E. *Clin. Endocrinol. Metab.* **1983**, *12*, 303−319.
(45) Pavlides, S.; Whitaker-Menezes, D.; Castello-Cros, R.; Flomenberg, N.; Witkiewicz, A. K.; Frank, P. G.; Casimiro, M. C.; Wang, C.; Fortina, P.; Addya, S.; Pestell, R. G.; Martinez-Outschoorn, U. E.; Sotgia, F.; Lisanti, M. P. *Cell Cycle* **2009**, *8*, 3984−4001.
(46) Shabalin, A. A.; Tjelmeland, H.; Fan, C.; Perou, C. M.; Nobel, A. B. *Bioinformatics* **2008**, *24*, 1154−1160.
(47) Benito, M.; Parker, J.; Du, Q.; Wu, J.; Xiang, D.; Perou, C. M.; Marron, J. S. *Bioinformatics* **2004**, *20*, 105−14.
(48) Rudy, J.; Valafar, F. *BMC Bioinf.* **2011**, *12*, 467.
(49) Taminau, J.; Meganck, S.; Weiss-Solis, D. Y.; van Staveren, W. C. G.; Dom, G.; Venet, D.; Bersini, H.; Detours, V.; Nowé, A. *Aust. J. Intell. Inform. Process. Syst.* **2009**, *10*, 4−11.
(50) Seasholtz, M. B.; Kowalski, B. *Anal. Chim. Acta* **1993**, *277*, 165−177.