Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2011

Impact Factor: 5.64 · DOI: 10.1021/ac201065j · Source: PubMed

14 AUTHORS, INCLUDING:

Kirill A. Veselkov
Imperial College London
37 PUBLICATIONS **1,112** CITATIONS
SEE PROFILE

Elizabeth J Want
Imperial College London
76 PUBLICATIONS **4,272** CITATIONS
SEE PROFILE

Jia V Li
Imperial College London
52 PUBLICATIONS **1,529** CITATIONS
SEE PROFILE

Richard H Barton
Imperial College London
36 PUBLICATIONS **2,073** CITATIONS
SEE PROFILE

analytical chemistry

ARTICLE

pubs.acs.org/ac

Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery

Kirill A. Veselkov,[†] Lisa K. Vingara,[‡] Perrine Masson,[§] Steven L. Robinette,[†] Elizabeth Want,[†] Jia V. Li,[†] Richard H. Barton,[†] Claire Boursier-Neyret,[§] Bernard Walther,[§] Timothy M. Ebbels,[†] István Pelczer,[∥] Elaine Holmes,[†] John C. Lindon,[†] and Jeremy K. Nicholson*[,†]

[†]Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ, United Kingdom

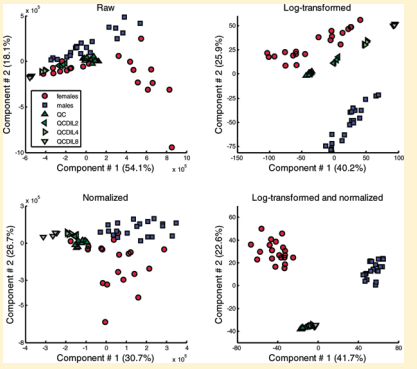
[‡]Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland 97239, United States

[§]Technologie Servier, 27 Rue Eugène Vignat, Orleans 45000, France

[∥]Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1014, United States

Supporting Information

ABSTRACT: Ultra-performance liquid chromatography coupled to mass spectrometry (UPLC/MS) has been used increasingly for measuring changes of low molecular weight metabolites in biofluids/tissues in response to biological challenges such as drug toxicity and disease processes. Typically samples show high variability in concentration, and the derived metabolic profiles have a heteroscedastic noise structure characterized by increasing variance as a function of increased signal intensity. These sources of experimental and instrumental noise substantially complicate information recovery when statistical tools are used. We apply and compare several preprocessing procedures and introduce a statistical error model to account for these bioanalytical complexities. In particular, the use of total intensity, median fold change, locally weighted scatter plot smoothing, and quantile normalizations to reduce extraneous variance induced by sample dilution were compared. We demonstrate that the UPLC/MS peak intensities of urine samples should respond linearly to variable sample dilution across the intensity range. While all four studied normalization methods performed reasonably well in reducing dilution-induced variation of urine samples in the absence of biological variation, the median fold change normalization is least compromised by the biologically relevant changes in mixture components and is thus preferable. Additionally, the application of a subsequent log-based transformation was successful in stabilizing the variance with respect to peak intensity, confirming the predominant influence of multiplicative noise in peak intensities from UPLC/MS-derived metabolic profile data sets. We demonstrate that variance-stabilizing transformation and normalization are critical preprocessing steps that can benefit greatly metabolic information recovery from such data sets when widely applied chemometric methods are used.



Detection of a large number of low molecular weight metabolites in complex mixtures has become an important tool to study human disease and the effects of drug therapy, to identify biochemical functions of unknown genes and assess the impact of mutations, and to investigate the interactions of genes and environment to produce observable phenotypes at the cell and organism level.[1−3] This approach, often referred to as metabonomics,[4] metabolomics,[1] or metabolic profiling, aims to improve understanding of physiology and metabolism by using analytical chemistry techniques to assess metabolic changes in biofluids, tissues, and cell extracts following an experimental perturbation. Unlike microarray based "omics" platforms, metabonomics relies on established analytical chemistry techniques such as liquid chromatography/mass spectrometry (LC/MS), gas chromatography/mass spectrometry (GC/MS) and nuclear magnetic resonance spectroscopy (NMR) to obtain metabolic signatures.[1−3] The recent introduction of ultraperformance LC (UPLC/MS) has greatly enhanced chromatographic performance, increasing the sensitivity and throughput of LC/MS measurements. Although the determination of the exact number of metabolites which can be measured by untargeted UPLC/MS is complicated by sample preparation, chemical diversity of the matrix, and the presence of isotopes and fragment and adduct ions, UPLC/MS routinely detects thousands of features representative of hundreds to thousands of metabolites in biological mixtures.[5]

While these analytical chemistry techniques boast high sensitivity and reproducibility and are capable of untargeted detection

Received: February 23, 2011
Accepted: April 28, 2011
Published: April 28, 2011

of a large number of diverse chemical species, the collection and comparison of a large number of NMR or mass spectra pose formidable data analysis challenges regardless of whether NMR, GC/MS, or LC/MS is applied. While the capabilities of these techniques have been compared elsewhere, spectra generated by each require similar preprocessing steps before comparative analysis.[6] The signatures of thousands of metabolites generated by UPLC/MS require extensive preprocessing prior to comparative statistical analysis.[7−10] A key step in the analysis of UPLC/MS data sets is the transformation of ion intensities resolved by elution time into a matrix of features (m/z−retention time pairs; columns) present in each sample (rows) by peak detection, alignment, and area extraction algorithms. A number of open source tools such as XCMS,[10] MAVEN,[11] MZmine,[12,13] OpenMS,[14,15] and XAlign,[16] as well as proprietary software packages such as MarkerLynx (Waters Corp), Mass Profiler (Agilent), and others from Thermo Scientific and Applied Biosystems, are currently available to achieve these objectives. Subsequent statistical analysis tools operate on this matrix of ion intensities.

Despite the significant attention paid to peak identification methods, instrumental and experimental challenges complicating the direct interpretation of ion intensities as metabolite concentrations have been largely neglected in the experimental literature.[17,18] This is especially true with respect to the urinary matrix commonly used in metabonomics studies. While urine is an ideal biofluid for metabonomics studies since it is obtained noninvasively, and its composition is affected by genetic and environmental factors reflecting the physiology of multiple organs, urine dilution effects and instrumental variation from the analytical method play a significant confounding role when one attempts to characterize these biological and physiological factors through NMR and MS measurements of small molecule concentrations.[19] Ensuring that NMR- and MS-generated urinary profiles are directly comparable by reducing the impact of variance associated with dilution and noise by data normalization and transformation are critical steps in MS data preprocessing prior to multivariate statistical analysis.[18,20]

Normalization is used to identify and remove sources of systematic variation between sample profiles due to factors that are irrelevant with regard to biological processes, such as sample dilution or variation in instrument detector sensitivity, in order to ensure that spectra are comparable across runs and across related sample sets.[20] Up to 15-fold changes in urine volume are commonly observed under normal conditions, resulting in significant variation in dilution of metabolite concentrations across many samples.[21] Additionally, disease or toxic effects may cause further changes in overall urine volume, further complicating interpretation of changes in observed metabolite concentrations. Technical variation in UPLC/MS arises from sources such as differences in LC separation efficiency (e.g., due to column aging), and drifts in ionization and detector efficiencies (e.g., due to source contamination). In order to obtain repeatable and interpretable LC/MS metabolic profiling data, the chromatographic retention time, ion signal intensity, and mass accuracy must be stable both within a sample batch and between multiple batches. It is well-known that the first few injections of sample on an LC/MS system give unrepresentative results for global metabolic analysis, due to small changes in chromatographic retention time and/or signal intensity. Usually after 5−10 injections of the matrix, retention times stabilize as the column becomes "conditioned" and the system then shows little variability through the remainder of the run. Additionally, the source of the mass spectrometer can become contaminated, leading to gradual changes in instrument sensitivity over time. Providing that these changes are not major, the subsequent data treatment will not be too adversely affected, provided that careful randomization of the samples has been performed to ensure that all of the experimental groups are affected to the same extent. Quality control samples, ideally a pooled sample consisting of aliquots from all samples in a study batch, can be used to condition the column at the start of the run and are then injected periodically, for example, every 10 samples throughout the run to assess instrument stability. These QC samples can also be used to assess stability across multiple instruments.

In order to ensure that samples are comparable within a batch and between multiple sets of samples, a variety of methods can be employed to estimate normalization factors.[17,20,22−24] Each technique relies on a set of assumptions about the analytical measurements and the biological variability of a data set in question. The normalization approaches applied to LC/MS metabolic data sets have been thus far largely limited to basic techniques that scale all spectra by a common factor, usually the total metabolite peak intensity or physiological variables such as urine volume, osmolality, or creatinine level.[24] However, both the total metabolite concentration and physiological variables of urine can vary considerably with reference to a variety of factors such as age, gender, diet, and time of urine collection. The instrumental effects of ion suppression and detector saturation can also cause highly abundant metabolite peaks to respond differentially to dilution, and thus the same scaling factor may not necessarily be required across the peak intensity range.[20] More recently, a normalization method was developed to correct for UPLC/MS peak intensity attenuation of serum profiles with respect to the order of injection based on QC data.[25] It is a robust method to account for technical variance due to factors such as loss of instrumental sensitivity, but it is not directly applicable to urine studies where additional and often more substantial variation of total urinary volume (and hence metabolite concentrations) is expected.

In addition, the UPLC/MS peak intensities of metabolic profiles are subject to noise from various sources and hence of different types resulting in a heteroscedastic noise structure, characterized by increasing technical variance as a function of increased signal intensity.[17,18] This can adversely affect standard statistical and pattern recognition tools,[26,27] which assume the data noise to be consistent across the whole intensity range (i.e., homoscedastic). This issue can be addressed either by modifying a statistical model to accommodate heteroscedastic noise structure[26,28] or by validating an appropriate variance-stabilization transformation to make peak intensity consistent across the whole intensity range.[27] Where possible, the last is often a preferred option. Here, we introduce a statistical model for UPLC/MS peak intensities of metabolite levels that comprises data normalization and quantification of measurement noise. The objective is to remove sources of systematic variation between sample profiles due to factors that are irrelevant with regard to biological processes such as sample dilution (normalization) and to stabilize the technical variance as a function of signal intensity (variance-stabilizing transformation). Four different normalization techniques [locally weighted scatter plot smoothing (LOESS), quantile, median fold change, and total intensity] are compared for adjusting peak intensities between samples to a common scale. The choice is made to cover a range of assumptions regarding UPLC/MS metabolic measurements and biological variability of a data set. The utility of logarithmic-based variance-stabilizing transformations is evaluated.

The proposed approach could be adapted for use in other types of biochemical mixtures and is thus generally applicable to a wide range of analytical chemistry applications.

MATERIALS AND METHODS

Study Details. Urine samples analyzed as part of a study on gender differences in rat metabolic profiles were used to assess the different preprocessing procedures. After 1 week of acclimatization, 40 Wistar rats (20 males and 20 females), aged 9 weeks, were housed individually in metabolism cages for 24 h under a conventional environment with controlled conditions [temperature 20−24 °C, relative humidity 40−70%, 12-h light cycle (7 a.m. to 7 p.m.)]. Filtered tap water was available ad libitum. Food [sterilized A04C-10 feed pellets from SAFE (Villemoisson-sur-Orge, France)] was available ad libitum for the first 8 h, and then the animals were fasted for the subsequent 16 h. Urine was collected over the intervals 0−8 h and 8−24 h into refrigerated tubes protected from light containing sodium azide (500 μL of 1% solution). After collection, all samples were centrifuged and 225 μL aliquots were stored at −40 °C until UPLC/MS analysis.

Sample Preparation. Water (150 μL) was added to 50 μL of urine. Samples were vortexed and centrifuged (16000*g*, 10 min), and 140 μL of supernatant was transferred into MS 96-well plates. A quality control (QC) sample, representative of the sample batch, was prepared by combining 40 μL of each sample. Dilutions 1/2, 1/4, and 1/8 of this QC sample in water were also prepared.

UPLC/MS Analysis. Analyses were performed on an Acquity UPLC system (Waters, Elstree, U.K.) coupled to a LCT Premier time-of-flight mass spectrometer (Waters, Manchester, U.K.), operated in the positive electrospray ionization (ESI+) mode. Chromatography was carried out at 40 °C with a 0.5 mL/min flow rate on a Waters Acquity UPLC HSS T3 column (1.8 μm, 2.1 × 100 mm), with the following solvent system: A = 0.1% formic acid in water, B = 0.1% formic acid in acetonitrile. A 10 min gradient was used, followed by a 2 min re-equilibration phase (0−1 min, hold at 0% B; 1−3 min, 0−15% B (linear); 3−6 min, 15−50% B (linear); 6−9 min, 50−95% B (linear); 9−10 min, hold at 95% B; 10.1−12 min, hold at 0% B). Gradient type = 6 (Waters). Samples were kept at 4 °C in the autosampler during the analysis. The injection volume was 5 μL. ESI conditions were as follows: source temperature 120 °C, desolvation temperature 350 °C, cone gas flow 25 L/h, desolvation gas flow 900 L/h, capillary voltage 3200 V, cone voltage 35 V. The instrument was operated in V optics mode and set to acquire data over the m/z range 50−1000 with scan time of 0.2 s and an interscan delay of 0.01 s. Data were collected in centroid mode. Leucine enkephalin (MW = 555.62) (200 pg/μL in acetonitrile/water 50:50) was used as a lock mass for accurate mass measurements. The instrument was calibrated before analysis by use of 0.5 mM sodium formate solution. The run order was randomized to minimize its correlation with the factors of the experimental design (gender and collection time point). The QC sample was injected regularly throughout the run (after every 10 samples) to monitor the stability of the analytical platform.[5,9,29] This QC sample was also used to condition the column at the beginning of the UPLC/MS run,[5,9] because the first injections of a batch are often not reproducible.

Raw Data Preprocessing. UPLC/MS raw data files were converted to netCDF format by use of the DataBridge tool implemented in MassLynx4.1 software (Waters). They were then preprocessed by use of the freely available XCMS software.[10] The CentWave algorithm[30] was used for peak picking with a peak width window of 3−15 s; the m/z width for the grouping step was changed to 0.1 Da; and the bandwidth parameter was kept to default (30 s) for the first grouping and then determined from the retention time deviation profile after retention time correction. This preprocessing step produced an output table of time-aligned detected features with their retention time, m/z ratio, and intensity in each sample. Isotope peaks, fragments, and adducts were treated as separate metabolite features. For the work carried out in this paper, the urine samples of the collection interval 8−24 h only were used, resulting in 40 urine samples (20 males and 20 females) plus the QC samples.

Statistical Model. An UPLC/MS data set following the above routine preprocessing stages can be represented by a rectangular table ($\mathbf{X}$) of real number elements (x_{ik}). The rows and columns of $\mathbf{X}$ represent, respectively, samples and metabolite peak intensities. Due to a variety of experimental, instrumental, environmental, and signal processing influences, the resulting peak intensities are a combination of real signal and noises of varying characteristics, which are generally of additive or multiplicative nature[18,31] and thus can be modeled by

$$x_{ik} = \beta_{ik} + n_{ik} s_{ik} e^{\eta_{ik}} \quad (1)$$

where x_{ik} is the measured peak intensity (level) of the ith metabolite in the kth sample, s_{ik} is the expected peak intensity (level), β_{ik} is the random background (electronic) noise, η_{ik} is the multiplicative random noise (for example, due to sample preparation variation or fluctuations in the ion source or the sample-introduction equipment), and n_{ik} is the normalization scaling factor of signal intensities of the kth sample as a result of variable sample dilution assumed to be either fixed (n_k) or peak-intensity-dependent (n_{ik}). Additive noise is characterized by random fluctuations in the baseline, irrespective of the presence of metabolite signals, and is derived from electronic noise in the equipment. Conversely, multiplicative noise grows with the signal intensity of a metabolite and is often proportional to it. As a result of the influence of multiplicative noise, metabolites with higher peak intensities would exhibit larger variability when repeatedly measured, and thus weak signals can be buried in the noise of strong signals. The objective of the preprocessing strategies considered here is to remove the systematic variation due to differences in overall sample concentration (normalization) and to convert multiplicative noise into additive noise (variance-stabilizing transformation):

$$\mathrm{vst}\left(\frac{x_{ik}}{n_{ik}}\right) \approx \mu_{ik} + \varepsilon_{ik} \quad (2)$$

where vst denotes a variance-stabilizing transformation, μ_{ik} is the transformed peak intensity, and ε_{ik} is the random noise. The overall procedure is quite straightforward. The fixed or peak-intensity-dependent normalization factor is initially estimated for each sample profile. The peak intensity is then divided by the sample-dependent normalization factor. The generalized logarithm or logarithm variance-stabilizing transformation is finally applied to convert multiplicative noise into additive noise.

Normalization: Accounting for Systematic Bias between Samples. Four different normalization techniques were selected for adjusting peak intensities between samples to a common scale. The choice was made to cover a range of assumptions regarding UPLC/MS metabolic measurements and biological variability of a data set. The first two approaches considered here

are the total intensity and median fold change normalizations. Both methods assume that measured peak intensities are directly proportional to concentrations of metabolites in solution. Under this assumption, the change in intensity of a profile due to variable sample dilution and/or inconsistency of UPLC/MS acquisition parameters is expected to be uniform across all peaks and thus a fixed scaling factor is employed.

Total intensity normalization forces all samples in a set of experiments to have equal total intensity:

$$n_k^{\mathrm{TI}} = \sum_i x_{ik} \tag{3}$$

where n_k^{TI} is the total intensity of the kth sample.[24] This technique assumes that the total concentration for all metabolites in a sample does not vary across samples in a data set. This assumption is not generally fulfilled in urinary metabolic studies, due to changes in the number and concentration of small molecules excreted.[32] Changes in the peak intensities of a few high-concentration metabolites can compromise noticeably the normalization performance because of their substantial contribution to the total peak intensity.

Median fold change normalization adjusts the median of log fold changes of peak intensities between samples in a set of experiments to be approximately zero:

$$n_k^{\mathrm{MFC}} = \mathrm{median}\left(\frac{x_{ik}}{x_{ir}}\right) \tag{4}$$

where n_k^{MFC} is the median fold change of peak intensities between the kth sample and the target sample. This normalization is based on the reasonable premise that metabolite peaks affected purely by dilution will exhibit the same fold changes between two sample profiles. This should be the case if peak intensities are directly proportional to metabolite levels in solution. The median value of fold change between sample and target profiles has been shown to be a robust estimate of this dilution factor.[32] The performance of this method is robust against at least 50% peak intensities exhibiting asymmetrical increase or decrease in response to biological factors such as disease or toxic challenge. The choice of a target profile has been shown not to be crucial for the method performance; it can be any sample of a given data set or, as typically selected, the median profile.

The other two approaches considered here are quantile and LOESS normalizations. These methods assume that systematic sample-to-sample variation is peak-intensity-dependent, that is, not necessarily the same scaling factor is used over the peak intensity range. This variation may arise due to the effects of ion suppression and detector saturation on measured metabolite levels.

Quantile normalization enforces all samples in a set of experiments to have identical peak intensity distribution:[22]

$$n_{ik}^{\mathrm{QN}} = \frac{x_{ik}}{x_{r(i)t}} \tag{5}$$

where n_{ik}^{QN} is the quantile-intensity-dependent normalization factor of the kth sample and the subscript (i) enclosed in parentheses indicates the order statistic of the ith peak intensity of the kth sample and x_t is a target sample. Each peak intensity of a sample by the quantile normalization is simply substituted by the peak intensity with the same order statistic $r(i)$ of a target sample. For example, the highest intensity in a given sample takes the value of the highest peak intensity in the target sample, the second highest peak intensity takes the second highest in the target sample, and so on. The target profile is typically selected to be the mean or median profile of a data set. The method assumes that the distribution of peak intensity across a data set is nearly the same for all samples and can be problematic with high-intensity values since they typically vary noticeably from sample to sample. It is also assumed that there is a similar number of metabolites with increased and decreased signals across the peak intensity range.

Locally weighted scatter plot smoothing (LOESS) normalization adjusts the local median of log fold changes of peak intensities between samples in a data set to be approximately zero across the whole peak intensity range:[23,33]

$$n_{ik}^{\mathrm{LOESS}} = \mathrm{f}(x_{ik}) \tag{6}$$

where n_{ik}^{LOESS} is the LOESS-intensity-dependent normalization factor of the kth sample calculated by nonlinear mapping of sample intensities into those of a reference sample. It is assumed that the proportion of metabolic changes across biological samples is relatively small or that there is a similar number of metabolites with increased and decreased signals across the peak intensity range.

Variance-Stabilizing Transformation: Accounting for Heteroscedastic Noise Structure. To account for heteroscedastic noise structure observed in UPLC/MS metabolite measurements, two logarithm-based transformations are chosen on the basis of the model (eq 1). It is first assumed that the noise is mainly multiplicative in nature, that is, the influence of background random noise is negligible (i.e., $\beta_{ki} \approx 0$). This assumption has been previously used in modeling UPLC/MS data[17] and other types of high-throughput data generated by, for example, microarray technology.[26,34] In this noise model, the standard deviation of peak intensity is proportional to its expected value. The logarithm is an appropriate variance-stabilizing transformation, as it makes peak intensity variance independent of its expected value by transforming multiplicative error into additive error:

$$\log\left(\frac{x_{ik}}{n_{ik}}\right) \approx \log(s_{ik}) + \eta_{ik} \tag{7}$$

It is second assumed that the error structure of UPLC/MS measurements has both additive and multiplicative noises (eq 1). The additive random component may arise from the integration of count-based signal inherent in the majority of MS instrumentation and/or the presence of a small basal unspecific background signal component. The variance of peak intensity under this model is a quadratic function of its expected value.[26] The same model has been previously considered in modeling LC/MS data[18] and is commonly used in gene expression studies.[26,27,31] The inverse generalized logarithm transformation has been derived as an appropriate variance-stabilizing transformation in this case:[27]

$$\mathrm{glog}\left(\frac{x_{ik}}{n_{ik}}\right) \approx \mu_{ik} + \varepsilon_{ik} \tag{8}$$

$$\mathrm{glog}(x) = \log\left(\frac{x + \sqrt{x^2 + c}}{2}\right) \tag{9}$$

where c is a fitted parameter. This transformation (eq 9) coincides with the logarithm for large intensities but is approximately linear for low intensities and interpolates smoothly in between. The scripts for

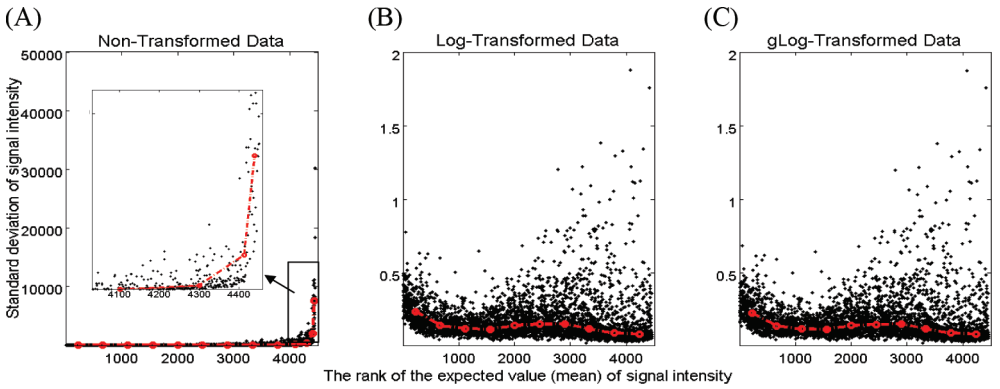


Figure 1. Standard deviation as a function of the rank of mean signal intensity for QC samples: (A) nontransformed data, (B) logarithm- (log-) transformed data, and (C) generalized logarithm- (glog-) transformed data. The running median of the standard deviation is shown as a dotted red line in the plot. In the absence of heteroscedastic noise structure, it should be approximately a horizontal line with minor fluctuations but no overall trend.

performing variance-stabilizing normalizations described above are available, upon request, in the open source statistical language R.

Principal Component Analysis. Formally, principal component analysis (PCA) summarizes the data matrix in terms of scores and loadings:[35]

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \qquad (10)$$

where **X** is a data matrix of sample profiles, **T** is a matrix of scores, **P** is a matrix of weights (loadings) of spectral variables into the scores, **E** is a residual, and the prime (′) denotes a transpose operation. PC scores summarize interrelationships between observations, for example, groupings according to physiological traits or gender and dose- or time-dependent changes. A subset of metabolites correlated with PC scores can be identified by use of the corresponding PC loadings. PCA was performed by using the built-in function of Matlab for eigenvalue decomposition (http://www.mathworks.com/).

RESULTS AND DISCUSSION

Performance of Logarithmic-Based Variance Stabilizing Transformations. To evaluate the changes in variation of peak intensities across the entire measurement intensity range, replicate injections of a quality control (QC) sample, obtained by pooling equal aliquots of each urine sample in the study, were used, as this QC sample represents all metabolites present in the experimental set and can serve as a measure of technical variability. The QC sample was measured eight times in the experimental run. The standard deviation as a function of the rank of the mean of peak intensity for replicate injections of the QC sample is shown in Figure 1A. In the absence of heteroscedastic noise structure, the running median of the standard deviation should approximate a horizontal line with minor fluctuations but no overall trend. This is not the case for the UPLC/MS data, where the variation of peak intensity grows with the rank of mean intensity, that is, as intensity increases. As a consequence for replicate measurements, larger peak intensities are progressively more variable. This means that assumptions underlying many widely applied statistical methods are not valid, and this in turn could lead to incorrect statistical inferences, having consequences for metabolic information recovery. To correct for heteroscedastic noise structure, the log and glog transformations were applied. Both transformations performed comparatively well in stabilizing the variance of transformed peak intensities across the entire measurement intensity range (Figure 1B,C). This indicates the negligible influence of background additive noise and the predominant influence of multiplicative noise in the UPLC/MS peak intensity data. Such a dominating influence of multiplicative noise has been previously observed in LC/MS measurements of protein levels.[18]

Performance of Normalization Methods in the Absence of Biological Variation. To assess the ability of normalization to account for dilution effects, the QC sample was systematically diluted, ranging from 2- to 8-fold. An 8-fold dilution of urine samples is quite common in metabolic studies. The objective here is to evaluate whether sample dilution is likely to induce uniform or peak-intensity-dependent changes in metabolic profiles. The uniform change due to sample dilution occurs when measured peak intensities are directly proportional to metabolite levels in solution (e.g., 2-fold sample dilution would result in 2-fold decrease in all peak intensities). However, various instrumental influences can potentially introduce peak-intensity-dependent biases between samples. As a result of ion suppression or detector saturation, for example, the peak intensities of highly abundant peaks may respond differently to sample dilution or not be affected by it at all. One way to test this is to visualize the log fold changes of peak intensities between diluted and undiluted samples as a function of log peak intensity; in microarray analysis this referred to as the M versus A plot. Typical behavior of peak intensities in response to 2- and 4-fold dilutions is shown in Figure 2 panels A and B, where red curves (calculated by the nonlinear LOESS regression) capture the adjustment required for a particular intensity range between diluted and undiluted samples. Clearly, the dilution-induced adjustment of peak intensities is comparable across the entire peak intensity range (e.g., the scaling factor of around 2 is necessary to bring all peak intensities of the 2-fold diluted sample to a common scale with the undiluted sample). The majority of peak intensities are thus likely to respond linearly to sample dilution. This is further confirmed by the fact that the reduction of the percentage of dilution-induced variation for individual peak intensities (calculated by one minus ratio of variance of transformed peak intensity after and prior to application of normalization) is comparable (on average 70−95%) over the peak intensity range between four studied normalization methods (Figure 2C). It should be noted that high sample dilution might limit the detection of low-abundance metabolites in biological samples by UPLC/MS.

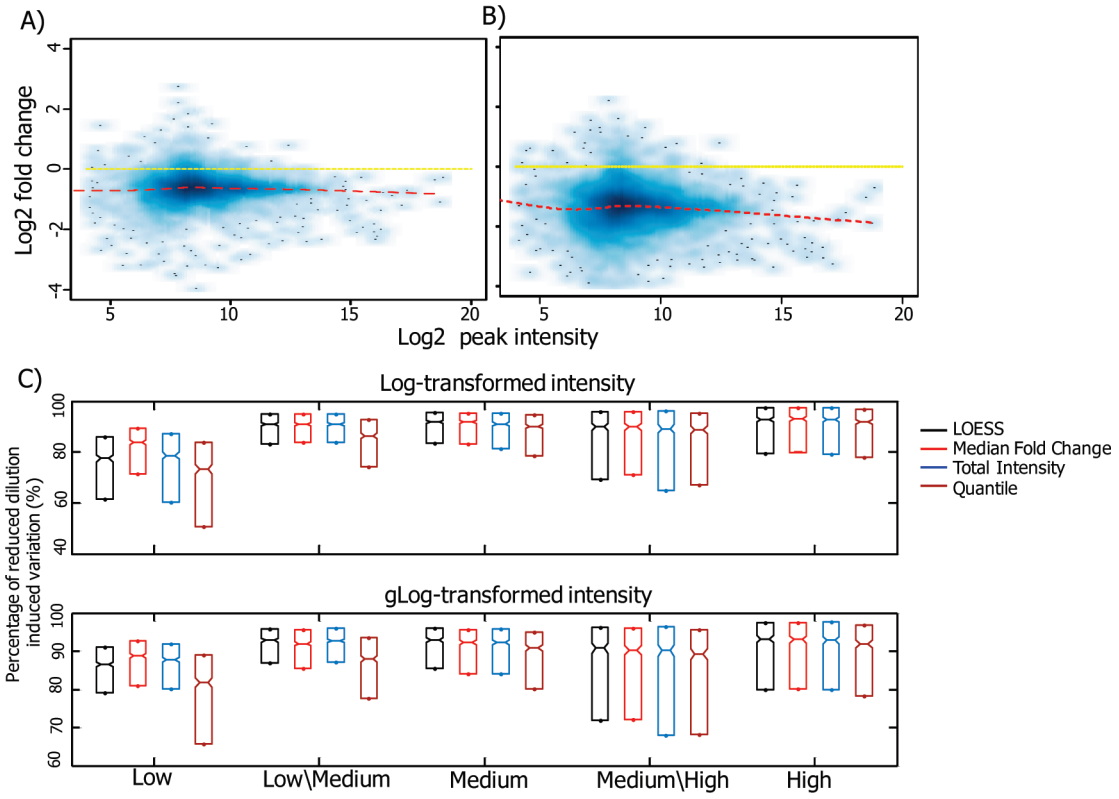


Figure 2. Assessment of peak-intensity-dependent bias due to variable sample dilution in UPLC/MS metabolic measurements. The log fold change in peak intensities between diluted and undiluted samples as a function of log peak intensity is shown for (A) 2-fold and (B) 4-fold sample dilutions. Red curves (calculated by nonlinear LOESS regression) capture the adjustment required for a particular intensity range between diluted and undiluted samples. (C) Reduction of the percentage of dilution-induced variation for individual log- or glog-transformed peak intensities across peak intensity range. Reduction of dilution-induced variation is calculated by $1 - var_{norm}/var_{non\text{-}norm}$, where var_{norm} and $var_{non\text{-}norm}$ are variances of transformed peak intensities after and prior to application of normalization, respectively. The sorted mean intensity values for 4453 peaks were divided into five approximately equal-sized data subsets, denoted by low, low/medium, medium, medium/high, and high peak intensity ranges. The upper edge of the box is the 75th percentile and the lower edge is the 25th percentile of values.

However, this issue relates only to the instrumental data acquisition and not to the subsequent data normalization.

Comparison of Normalization Methods in the Presence of Biological Variation. A robust normalization method not only must account for dilution effects in samples of the same composition but also must be applicable to experimental data where changes in mixture composition are to be expected. Certain assumptions with regard to the nature of biological variance are inherent to all normalization methods (see the Materials and Methods section for a detailed description). As all four studied methods performed comparatively well in the absence of biological variation, a method whose performance is robust to biologically induced changes in peak intensity would be of comparative advantage. The total intensity, quantile, and LOESS normalizations rely heavily on the assumption that there are similar numbers of metabolites with increased and decreased signals.[22,23] This assumption may not generally be fulfilled in metabolic studies since the increases in the level of one metabolite in response to stimuli may not necessarily be accompanied by the decreased level of another metabolite. The application of the LOESS and quantile methods is thus limited to cases in which a low percentage of metabolites have asymmetrically increasing or decreasing abundances. The performance of total intensity normalization can be compromised by a single large metabolite peak that varies substantially from sample to sample.[32] By contrast, the median fold change normalization allows at least half of metabolite peak intensities to exhibit asymmetrical changes between samples. This is a relaxed assumption with regard to the nature and proportion of differentially abundant metabolite peaks, which is generally applicable to data from metabolic studies.

The performance of normalization methods in the presence of biological variance can be diagnosed by use of summary statistics of sample peak intensity distributions via box plots. If it is assumed that up to a quarter of peak intensities are asymmetrically increased or decreased due to biological effects, the log ratios of nondifferentially abundant metabolite peaks should have a comparably small spread around zero across samples (noted by the box in Figure 3) and a median of zero (a black diamond in the box in Figure 3). This is illustrated on a urine data set of control Wistar rats, in which the majority of peak intensities are expected not to be differentially abundant. The peak intensities of nonnormalized samples exhibit large variation in overall concentration within sample types of male (red) and female (blue) rats (Figure 3A). This variation is within the 8-fold dilution range of the QC sample (green). The effect of this dilution-induced variation is that any informational changes in metabolite concentration between sample types would need to be very large to be detected. As a result, a large set of biomarker candidates would simply not be identified by statistical

dx.doi.org/10.1021/ac201065j |*Anal. Chem.* 2011, 83, 5864–5872

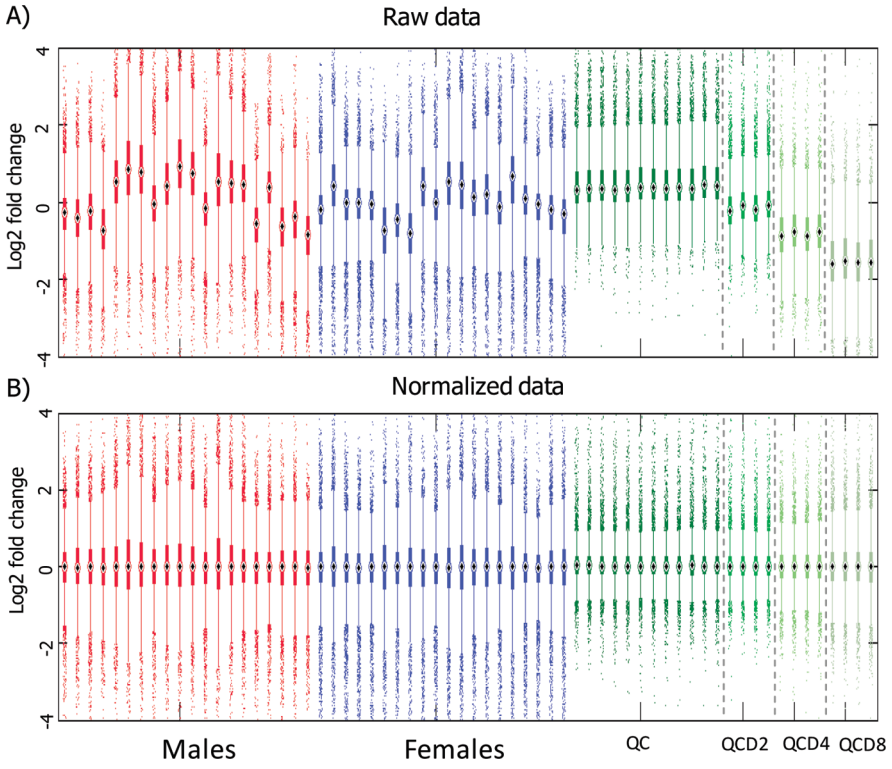


Figure 3. Box plots of the log 2-fold change in peak intensities relative to the median profile calculated across a data set: (A) nonnormalized and (B) median fold change normalized data. The upper edge of the box is the 75th percentile and the lower edge is the 25th percentile of relative intensity values and points noted by '.' are metabolite features which exhibit relative changes. The median is shown as a black diamond in the box. The boxes are colored by sample types. QC, quality control urine samples of Wistar rats; QCD2, 2-fold diluted quality control urine samples; QCD4, 4-fold diluted quality control urine samples; QCD8, 8-fold diluted quality control urine samples.

techniques. The median fold change, quantile, and LOESS normalizations successfully reduce dilution-induced variance, as shown by the near zero log ratios of nondifferentially abundant peaks, indicating the negligible influence of peak intensity biases between samples (Figure 3B and Figure S1, Supporting Information). Total intensity normalization, however, shows inferior performance as indicated by the remaining sample-to-sample variation. This is most likely due to small changes in total metabolite excretion between samples. We suggest that while quantile and LOESS show similar performance to median fold change, the median fold change normalization should be preferred because of its relaxed assumption with the proportion of asymmetric metabolic marker changes.

Impact of Variance-Stabilizing Normalization on Exploratory Data Analysis via PCA. Principal component analysis (PCA) is the most commonly used method to explore relationships between samples in metabolic profiling studies. It aims to extract a small number of latent components that summarize the measured data with minimal information loss by taking advantage of the correlation structure of peak intensities. The information content in PCA is measured as a sum of variances of individual peak intensities, referred to as the total variance. In the presence of multiplicative noise, metabolites with higher peak intensities are progressively more variable when measured repeatedly and therefore constitute a larger share of the total variation. The resulting PC scores are impacted heavily by the random variation of these high-intensity metabolites and therefore can poorly represent the complete variation structure of a data set. This effect is illustrated on the urine data set of control Wistar rats in Figure 4. The sample scores of the first two PCs (raw peak intensities, Figure 4) are scattered, indicating a dominant influence of random variation. As a consequence, the assessment of overall similarities and differences between samples is quite poor; the grouping of technical replicates with respect to the degree of dilution is not even clear. The normalization of raw peak intensities by median fold change results in only slight changes in the PC scores because of the dominant impact of multiplicative noise in high-concentration mixture components (normalized peak intensities, Figure 4). After application of log-based variance-stabilizing transformations, the first two components clearly capture the systematic variation due to the difference in sample composition and variable sample dilution (log/glog-transformed peak intensities, Figure 4). The dispersion in the scores of both male and female Wistar rats along PC1 and PC2 parallel to the QC dilution series indicates a range in concentration similar to the 8-fold dilution changes of the QC sample. Normalization successfully removes this dilution-induced variation, leading to tight clustering of technical replicates and improved clustering of biological replicates. In addition, all clusters of sample types are clearly distinguishable (log/glog-transformed and normalized peak intensities, Figure 4). The sample scores of the first two PCs between log- and glog-based transformations across four normalization methods are somewhat comparable (Figure S2, Supporting Information). This confirms the results of the previous sections: that multiplicative noise is the most significant source of background variation in MS metabolic profiling and that all four normalization methods remove the majority of dilution-induced variation from the data.

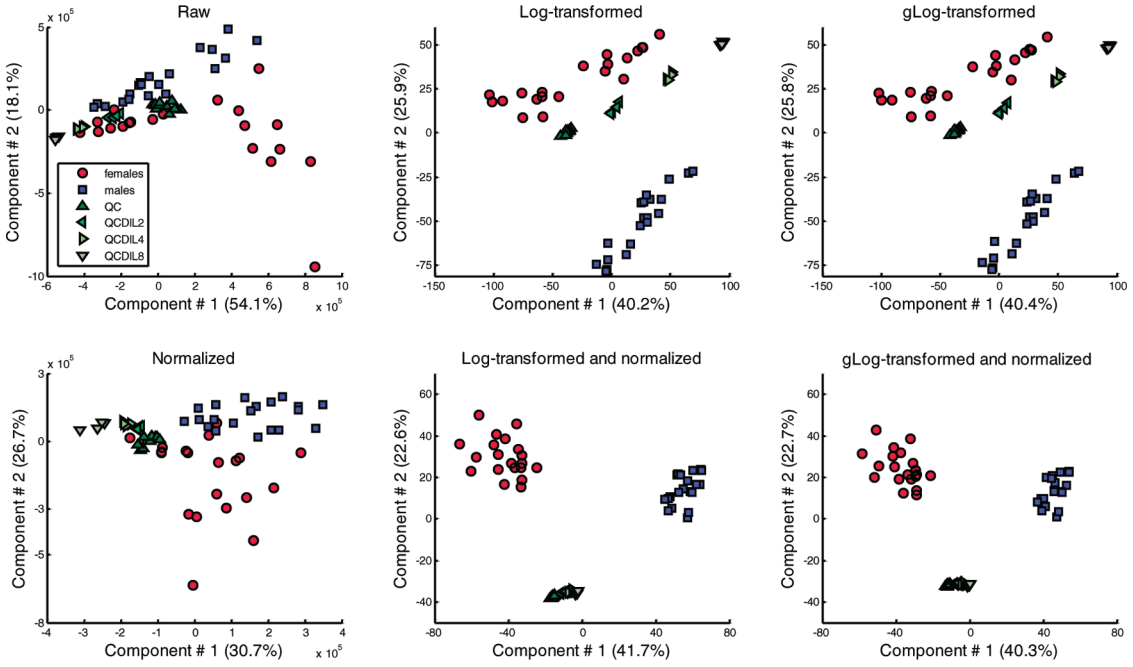


Figure 4. Impact of variance-stabilizing normalization on metabolic information recovery via PCA. QC, quality control urine samples of Wistar rats; QCDIL2, 2-fold diluted quality control urine samples; QCDIL4, 4-fold diluted quality control urine samples; QCDIL8, 8-fold diluted quality control urine samples. The data were normalized by median fold change normalization.

CONCLUSIONS

We have compared several variance-stabilizing normalization strategies based on a statistical error model in order to account for systematic sources of variation, such as that induced by variable sample dilution, and in order to stabilize the technical variance as a function of peak intensity observed in UPLC/MS metabolic profiles of urine.

It can be concluded that the UPLC/MS peak intensities of urine metabolic profiles are likely to respond linearly to variable sample dilution across the intensity range. In other words, peak-intensity-dependent biases between samples as a result of ion suppression or detector saturation that would cause highly abundant peaks to respond differentially to dilution were not observed. This indicates that the majority of urinary metabolites are within the detector dynamic range.

Second, all four studied normalization methods (LOESS, quantile, total intensity, and median fold change) performed comparatively well in reducing dilution-induced variation of urine samples in the absence of biological variation. However, the median fold change normalization is least compromised by biologically relevant changes in mixture components and is thus preferable

Third, the higher peak intensities of urine profiles exhibit larger variability when repeatedly measured, mainly due to the presence of multiplicative noise. This violation of constant variance across the measurement range imposes a serious challenge when standard statistical techniques are applied, as demonstrated by principal component analysis. The UPLC/MS peak intensity of urine samples can be brought in line with this assumption by applying a log-based transformation, which successfully stabilizes the technical variance across the intensity range.

Finally, variance-stabilizing transformation and normalization are critical preprocessing steps that can greatly benefit the metabolic information recovery from urine UPLC/MS data sets via commonly used pattern recognition tools.

Although applied to urine studies, the proposed framework for assessing the performance of normalization methods and variance-stabilizing transformation is general in scope and thus can be applied to other biological sample types.

ASSOCIATED CONTENT

Supporting Information. Two figures, showing box plots of log 2-fold change in peak intensities after various normalization techniques (LOESS, MFC, TI, and QN) and the impact of variance-stabilizing normalizations (LOESS, TI, and QN) on metabolic information recovery via PCA. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: j.nicholson@imperial.ac.uk.

ACKNOWLEDGMENT

K.A.V. and L.K.V. contributed equally to this work. K.A.V. gratefully acknowledges Technologie Servier for financial support.

REFERENCES

(1) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.
(2) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Everett, J. R. *Concepts Magn. Reson.* **2000**, *12*, 289–320.
(3) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–1189.
(4) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nat. Rev. Drug Discovery* **2002**, *1*, 153–161.

(5) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. *Nat. Protocols* **2010**, *5*, 1005–1018.
(6) Pan, Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, *387*, 525–527.
(7) Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. *Anal. Chem.* **2008**, *80*, 6382–6389.
(8) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747–751.
(9) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. *Analyst* **2006**, *131*, 1075–1078.
(10) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
(11) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Anal. Chem.* **2010**, *82*, 9818–9826.
(12) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.
(13) Katajamaa, M.; Miettinen, J.; Oresic, M. *Bioinformatics* **2006**, *22*, 634–636.
(14) Sturm, M.; Bertsch, A.; Gropl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. *BMC Bioinf.* **2008**, *9*, 163.
(15) Bertsch, A.; Gropl, C.; Reinert, K.; Kohlbacher, O. *Methods Mol. Biol.* **2010**, *696*, 353–367.
(16) Zhang, X.; Asara, J. M.; Adamec, J.; Ouzzani, M.; Elmagarmid, A. K. *Bioinformatics* **2005**, *21*, 4054–4059.
(17) Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Oresic, M. *BMC Bioinf.* **2007**, *8*, 93.
(18) Anderle, M.; Roy, S.; Lin, H.; Becker, C.; Joho, K. *Bioinformatics* **2004**, *20*, 3575–3582.
(19) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *NMR Biomed.* **2005**, *18*, 143–162.
(20) Kultima, K.; Nilsson, A.; Scholz, B.; Rossbach, U. L.; Falth, M.; Andren, P. E. *Mol. Cell. Proteomics* **2009**, *8*, 2285–2295.
(21) Tsuchiya, Y.; Takahashi, Y.; Jindo, T.; Furuhama, K.; Suzuki, K. T. *Eur. J. Pharmacol.* **2003**, *475*, 119–128.
(22) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. *Bioinformatics* **2003**, *19*, 185–193.
(23) Ballman, K. V.; Grill, D. E.; Oberg, A. L.; Therneau, T. M. *Bioinformatics* **2004**, *20*, 2778–2786.
(24) Warrack, B. M.; Hnatyshyn, S.; Ott, K. H.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. *J. Chromatogr. B* **2009**, *877*, 547–552.
(25) 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.
(26) Rocke, D. M.; Durbin, B. *Bioinformatics* **2003**, *19*, 966–972.
(27) Huber, W.; von Heydebreck, A.; Sultmann, H.; Poustka, A.; Vingron, M. *Bioinformatics* **2002**, S96–104.
(28) Karakach, T. K.; Wentzell, P. D.; Walter, J. A. *Anal. Chim. Acta* **2009**, *636*, 163–174.
(29) Gika, H. G.; Theodoridis, G. A.; Wingate, J. E.; Wilson, I. D. *J. Proteome Res.* **2007**, *6*, 3291–3303.
(30) Tautenhahn, R.; Bottcher, C.; Neumann, S. *BMC Bioinf.* **2008**, *9*, 504.
(31) Durbin, B. P.; Hardin, J. S.; Hawkins, D. M.; Rocke, D. M. *Bioinformatics* **2002**, S105–110.
(32) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal. Chem.* **2006**, *78*, 4281–4290.
(33) Dudoit, R.; Yang, Y. H.; Callow, M. J.; Speed, T. P. *Stat. Sin.* **2002**, *12*, 111–139.
(34) Ritchie, M. E.; Silver, J.; Oshlack, A.; Holmes, M.; Diyagama, D.; Holloway, A.; Smyth, G. K. *Bioinformatics* **2007**, *23*, 2700–2707.
(35) Wold, S.; Esbensen, K.; Geladi, P. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37–52.