Analytica Chimica Acta 955 (2017) 27–35

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Metabolomic analysis of urine samples by UHPLC-QTOF-MS: Impact of normalization strategies

Yoric Gagnebin [a], David Tonoli [a, b], Pierre Lescuyer [c], Belen Ponte [d], Sophie de Seigneux [d], Pierre-Yves Martin [d], Julie Schappler [a], Julien Boccard [a, b], Serge Rudaz [a, b, *]

[a] *School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Geneva, Switzerland*
[b] *Swiss Centre for Applied Human Toxicology (SCAHT), Universities of Basel and Geneva, Basel, Switzerland*
[c] *Department of Genetic and Laboratory Medicine, Geneva University Hospitals (HUG), Geneva, Switzerland*
[d] *Service of Nephrology, Geneva University Hospitals (HUG), Geneva, Switzerland*

HIGHLIGHTS

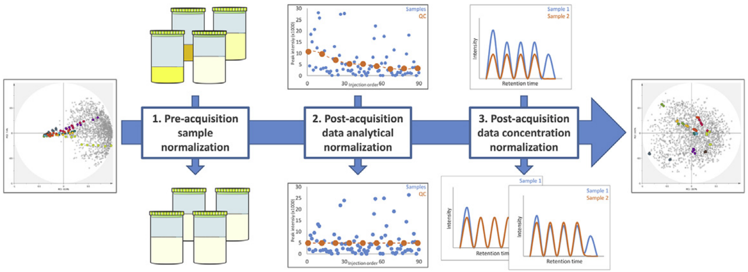
- Sequential normalization strategy for urine metabolomics by UHPLC-QTOF-MS is proposed.
- Pre-acquisition sample normalization enhanced the analytical conditions.
- Post-acquisition data normalization corrected the unwanted variance in the dataset.
- Sequential normalization significantly improved kidney failure patient stratification.

ARTICLE INFO

Article history:
Received 25 August 2016
Received in revised form
8 December 2016
Accepted 20 December 2016
Available online 28 December 2016

Keywords:
Metabolomics
LC-MS
Urine
Normalization
Kidney failure

GRAPHICAL ABSTRACT



ABSTRACT

Among the various biological matrices used in metabolomics, urine is a biofluid of major interest because of its non-invasive collection and its availability in large quantities. However, significant sources of variability in urine metabolomics based on UHPLC-MS are related to the analytical drift and variation of the sample concentration, thus requiring normalization. A sequential normalization strategy was developed to remove these detrimental effects, including: (i) pre-acquisition sample normalization by individual dilution factors to narrow the concentration range and to standardize the analytical conditions, (ii) post-acquisition data normalization by quality control–based robust LOESS signal correction (QC-RLSC) to correct for potential analytical drift, and (iii) post-acquisition data normalization by MS total useful signal (MSTUS) or probabilistic quotient normalization (PQN) to prevent the impact of concentration variability. This generic strategy was performed with urine samples from healthy individuals and was further implemented in the context of a clinical study to detect alterations in urine metabolomic profiles due to kidney failure. In the case of kidney failure, the relation between creatinine/osmolality and the sample concentration is modified, and relying only on these measurements for normalization could be highly detrimental. The sequential normalization strategy was demonstrated to significantly improve patient stratification by decreasing the unwanted variability and thus enhancing data quality.

---

* Corresponding author. School of Pharmaceutical Sciences, University of Geneva, 1 rue Michel Servet, 1211 Geneva 4, Switzerland.
E-mail address: serge.rudaz@unige.ch (S. Rudaz).

http://dx.doi.org/10.1016/j.aca.2016.12.029
0003-2670/© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Because metabolomics aims at the comprehensive monitoring of metabolites (mass < 1000 Da) in biological systems, it constitutes a potent approach for assessing phenotype modifications caused by disease or environmental influences at the molecular level. Among the various biological matrices used in metabolomics, urine is a biofluid of major interest because of its non-invasive collection and its availability in large quantities. However, one important issue remains: the natural variability in the concentration of urine, which depends mostly on the hydration and physiological status of the organism. The total concentration of metabolites in urine was reported to vary by more than 14-fold [1]. The comparison of metabolite levels between different groups of samples constitutes one of the most classic and widespread application of metabolomics to search for biomarkers. Due to the important variability in the concentration of urine, such an approach can lead to misleading results and wrong conclusions without the use of sample or data normalization methods. A literature survey reported that information related to normalization was lacking in almost half of the urinary metabolomics studies published between 2012 and 2014 [2]. Furthermore, when a normalization strategy applied was mentioned, useful details were rarely reported.

Several normalization strategies based on different principles have been developed for metabolomics and can be classified into three main categories: (i) pre-acquisition, by specific dilution or injection volume [3,4]; (ii) post-acquisition, by data treatment [5]; and (iii) pre- and post-acquisition using a combination of methods [3,6]. The choice of the optimal normalization strategy is often dependent on the context of the study, and efforts still need to be made to develop generic normalization strategies for the analysis of urine. For urine metabolomics, normalization has two distinct aims: to remove differences due to urine concentration and to correct for analytical variability. These two aspects are mandatory to properly compare metabolic profiles.

Strategies were developed to estimate urine concentrations, and the five most popular are based on different approaches: (i) relative concentration to one reference compound (creatinine) or a subgroup of molecules [7,8], (ii) measurement of the total solute concentration (osmolality) [6,9], (iii) urine/pure water density ratio (specific gravity) [4], (iv) 24-h urine volume [5,10], or (v) signal integration (total ion current) [11,12]. Each concentration estimator was used in pre-acquisition sample or post-acquisition data normalization. Creatinine excretion is a commonly used clinical parameter that is directly linked to glomerular filtration in normal physiological conditions [1,2]. Nevertheless, human urine metabolomic studies often target disease or abnormal physiological conditions; therefore, this estimator can be altered [13,14]. Furthermore, relying on a single compound to evaluate the overall sample concentration constitutes a risk when its proper measurement cannot be guaranteed [15]. Alternatively, osmolality allows a more comprehensive evaluation of the sample concentration and is considered the gold standard in clinical study [16]. Specific gravity is a potent osmolality estimator but has limited interest when osmolality is available [16]. *Warrack* et al. reported better results when applying normalization based on osmolality rather than creatinine and 24-h urine volume. Collecting the latter for all volunteers may be problematic in clinical studies. Alternatively to osmolality, a data treatment called MS total useful signal (MSTUS) based on a normalization factor calculated from the sum of features common to all samples can be applied [5]. This treatment limits the impact of exogenous compounds, such as xenobiotic and orphan features, that could be detected in a limited number of samples. Most recently, *Filzmoser* et al. described the concept of “closure” induced by total sum normalization (TSN) or MSTUS involving the distribution of variation over all signals when one component decreases or increases, thus creating links between features. These normalizations require the assumption of closed data, where the features add to a constant sum. *Filzmoser* et al. demonstrated the benefit of using probabilistic quotient normalization (PQN) over these relatively classic data pre-treatment methods in a simulation study when data were not initially closed [17].

In addition to the natural variability of urine concentrations, several studies reported analytical drift in UHPLC-QTOF-MS analysis [18]. To monitor and overcome this effect, quality control (QC) samples are injected at regular intervals during the experimental sequence as relevant landmarks of the analytical process stability to ensure the quality of data acquisition [19–21]. In most metabolomic studies, normalization is used to correct for the analytical variability during experimental runs by ensuring the repeatability of the QCs [18]. *Dunn* et al. reported that every feature fluctuated differently during the sequence, and global normalization is therefore sub-optimal. In that context, a quality control–based robust LOESS signal correction (QC-RLSC) selectively applied to each detected ion feature was proposed. For each feature in the QC samples, a specific locally weighted scatterplot smoothing (LOESS) fitted curve is obtained based on the QC samples to evaluate the temporal drift with respect to the processing order. This function is then used to correct for the same feature in each sample [18]. *Kirwan* et al. demonstrated the advantage of using data treatment based on robust spline correction (QC-RSC) in terms of execution speed compared to QC-RLSC [22]. Most recently, support vector regression correction (QC-SVRC) was reported to slightly outperform QC-RSC [23].

A final important aspect is the implementation of peak filtering in the workflow to ensure the relevance of the data. Using non-relevant features, such as noisy or saturated signals, in the data normalization step affects data quality and integrity. For this reason, peak filtering has to be implemented before data normalization. When applied after data normalization, peak filtering is strongly dependent on the type of normalization used. *Naz* et al. proposed to filter data by response to dilution [24]. QC and diluted QC (dQC) samples must be injected successively at regular intervals during the experimental sequence to reduce the impact of analytical drift on the filtering. A relevant feature must have a nearly linear response to the concentration, and the dQC/QC ratio is expected to remain constant throughout the sequence [24,25]. Signals filtered before normalization can lead to better data quality.

This study proposes a thorough investigation of various approaches for sequential normalization strategies in the context of urine metabolomic analysis by UHPLC-QTOF-MS based on individual dilution factors. This included the investigation of the creatinine concentration and osmolality and comparison with ^{1}H NMR measurements as absolute estimates of sample concentrations to calculate the individual dilution factors. In a second step, three data pre-processing methods, namely QC-RLSC [18], MSTUS [5] and PQN [17], were investigated and discussed. This sequential normalization strategy was implemented in the context of a clinical study of kidney failure, involving patients suffering from physiological alterations that may have an impact on urine concentration estimators, such as osmolality and creatinine [13,14].

2. Methods

2.1. Preliminary study

2.1.1. Sample collection and preparation

A series of urine samples was collected from 8 healthy volunteers (4 women, 4 men) at 3 time points separated by intervals of one month. The samples were aliquoted and stored at −80 °C. Urine