Journal of Chromatography A, 1437 (2016) 127–136

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Optimization of large-scale pseudotargeted metabolomics method based on liquid chromatography–mass spectrometry

Ping Luo[a], Peiyuan Yin[a], Weijian Zhang[b], Lina Zhou[a], Xin Lu[a], Xiaohui Lin[b], Guowang Xu[a,*]

[a] Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China
[b] School of Computer Science and Technology, Dalian University of Technology, Dalian, China

ARTICLE INFO

Article history:
Received 21 October 2015
Received in revised form
29 December 2015
Accepted 30 January 2016
Available online 4 February 2016

Keywords:
Pseudotargeted analysis
Metabolomics
Metabolic profiling
LC–MS
Quality control
Large-scale

ABSTRACT

Liquid chromatography–mass spectrometry (LC–MS) is now a main stream technique for large-scale metabolic phenotyping to obtain a better understanding of genomic functions. However, repeatability is still an essential issue for the LC–MS based methods, and convincing strategies for long time analysis are urgently required. Our former reported pseudotargeted method which combines nontargeted and targeted analyses, is proved to be a practical approach with high-quality and information-rich data. In this study, we developed a comprehensive strategy based on the pseudotargeted analysis by integrating blank-wash, pooled quality control (QC) sample, and post-calibration for the large-scale metabolomics study. The performance of strategy was optimized from both pre- and post-acquisition sections including the selection of QC samples, insertion frequency of QC samples, and post-calibration methods. These results imply that the pseudotargeted method is rather stable and suitable for large-scale study of metabolic profiling. As a proof of concept, the proposed strategy was applied to the combination of 3 independent batches within a time span of 5 weeks, and generated about 54% of the features with coefficient of variations (CV) below 15%. Moreover, the stability and maximal capability of a single analytical batch could be extended to at least 282 injections (about 110 h) while still providing excellent stability, the CV of 63% metabolic features was less than 15%. Taken together, the improved repeatability of our strategy provides a reliable protocol for large-scale metabolomics studies.

1. Introduction

Genome-wide association studies with metabolomics (mGWAS), combining investigation of genetic variances and metabolic phenotypes in large population cohorts [1–3], are improving the understanding of (patho-) physiological mechanism, diagnosis, prevention and treatment of complex disorders [4–9]. To cover the substantial diversity of phenotypes, large-scale epidemiological studies are required. For example, a GWAS combined with nuclear magnetic resonance (NMR)-based metabolomics study was applied to a large cohort of 8330 Finish individuals [10], and in the HUSERMET project at least 5000 samples were analyzed by gas chromatography–mass spectrometry (GC–MS) [11]. However, detection of thousands of samples increases the difficulty to obtain high quality metabolite data due to the analytical variances, such as the drift of instrument and environment during the long-time analysis [12,13]. Hence, a series of issues about analytical repeatability including machine drift, analysis order, and choice of technology should be taken into consideration in the large-scale metabolomics studies. All these issues present significant challenges, while have not been globally addressed [13,14].

Now most of studies are employing liquid chromatography–mass spectrometry (LC–MS), a mainstream platform for top-down metabolomics attributing to its high sensitivity, wide polarity range, and simple sample pretreatment [15–17]. For the machine drift in LC–MS, scientists are using the strategy of splitting samples into small analytical batches to increase its robustness and then merging multiple data sets into a single one [14]. Nevertheless, significant systematic analytical variation from drift of instrument was exhibited between different batches [18], signal-drift correction had to be carried out. Isotope labeled internal standards (ISs) are often employed to calibrate the instrumental drift for the analogues of ISs. But for the thousands of signals from nontargeted metabolic profiling, limited number of ISs are obviously not enough to adjust all

* Corresponding author. Fax: +86 411 84379530.
E-mail address: xugw@dicp.ac.cn (G. Xu).

http://dx.doi.org/10.1016/j.chroma.2016.01.078
0021-9673/

those metabolic features [18–22]. Recently, the use of pooled quality control (QC) samples are recommended for better stability and repeatability by equally inserting the QCs in the analytical sequence. QC sample has been used as external standards for the calibration of each ion from all studied samples thus to minimize the intra- and inter-batch variations [23–26]. These methods addressed the effects on the post-acquisition signal drift calibration (post-calibration), but the highly multivariate characterized data are difficult to be simulated, for instance, not all the detected metabolites will change in the same tendency [14]. Therefore, novel analytical method is required for the long-term study.

Up to date, the nontargeted metabolic profiling is a main analytical method, which provides a global overview of the metabolome without a priori knowledge of analytes, and the chemical structure of unknown compound can also be identified by using high-resolution mass spectrometer [27]. However, the stability of LC–MS is still a common issue due to its limited repeatability, linear range, and some false results introduced by the complicated peak alignment as the trade-offs of high sensitivity and throughput of the MS instruments [28,29]. Another common technology based on targeted analysis [30–33], is considered as the gold standard of the compound quantitation by using MS with good repeatability and wide linear range. But the limited coverage of metabolites restricts its applications in the -omics studies [34]. To compensate for the above mentioned two techniques, we developed a so called pseudotargeted method [35], which combines the advantages of both nontargeted and targeted methods. It consists of three steps containing metabolic profiling analysis of the pooled sample on quadrupole time-of-flight (Q-TOF) MS, ion-pair selection and multiple reaction monitoring (MRM) measurement of all samples on triple quadrupole (TQ) MS. Metabolic profiling can provide abundant metabolite information due to the MS priorities of high resolution, sensitivity and throughput. After characteristic ion pairs are selected, these metabolite ions could be quantified on TQMS with better quality. Compared to multiple targeted metabolomics [36,37], pseudotargeted method is not limited to the known metabolites. Similar strategies have also been applied by many other researchers, such as Song et al. screened the components in the QC samples by various mass spectrometric modes. After integrating the transitions lists together, 513 components were quantified by schedule MRM mode, among which 379 were identified [38]. Yan et al. employed data-dependent and data-independent methods to expand the MRM library of human urinary metabolome based on a directly-coupled reversed-phase and hydrophilic interaction chromatography separation system [39]. Pseudotargeted and its similar strategies are increasingly applied to provide good measurement and coverage of the metabolome [40–43].

In this study, we developed a comprehensive strategy based on pseudotargeted analysis by combining blank-wash, large pooled QC sample, and post-calibration to improve the stability of the large-scale metabolomics study. Integrating the advantages of targeted and untargeted methods, pseudotargeted analysis can provide a global metabolite information with the guarantee of better linearity and repeatability data. Considering the current issues about stability in the large-scale study, we first systematically optimized its analytical procedure from both pre- and post-acquisition aspects including the selection of QC samples, frequency of QC insertion, and post-correction method in a batch analysis by assessing the repeatability. Then the proposed strategy was applied to the data combination of 3 independent batches. These results illustrated that the new strategy can guarantee an evidently stable run, and notably improve the repeatability of the integrated data set from multiple batches, and also successfully extends the analytical capacity of a single batch to 282 injections (about 110 h).

2. Materials and methods

2.1. Experimental design

The work flow for the development, optimization and application of the suggested strategy based on pseudotargeted analysis for the improvement of stability and robustness in large-scale metabolomics is displayed in Fig. 1. The goal of the study is to develop a new strategy based on the pseudotargeted technology combined with blank-wash, large pooled QC samples, and post-correction method to address the current issues about stability and robustness in the large-scale metabolomics study. To realize this aim, optimization of the analytical procedure from both pre- and post-acquisition aspects are required. Finally, the proposed comprehensive strategy was used to combine multiple batches and extend the analytical capacity of a single batch to reduce or avoid the significant analytical variances of intra- and inter-batch.

To assess the effect of the new strategy four independent batches were analyzed, and the studied individuals and two QC samples were repeatedly analyzed as Fig. S1, Supplementary material. Two kinds of QC samples were injected at the start and end of each batch, and every four studied samples injections were successively followed by the two QC samples. A blank-wash, which had the same elution gradient as sample analysis and an additional 30 min of rinse time with 100% B phase (organic phase), was inserted every 60 injections to wash the build-up contaminations from the consecutive injections of the LC column [14]. All samples were freshly reconstructed every day, and there were 8, 8, and 3 replicates of studied samples in batch 1, batch 2, and batch 3. The mobile phases were freshly prepared for each batch, and interval time of the former two batches was 4 weeks, and of batch 1 and batch 3 was 5 weeks, and a new column was used in batch 3. Routine instrument maintenance and calibration were done between batches. Batch 4 was the consecutive analysis of QC samples and applied to maximize the analytical capacity of a single batch.

2.2. Sample collection and pretreatment

The fasting serum samples of 5 healthy controls and 5 hepatic fibrosis patients were collected from the Sixth People's Hospital of Dalian (Dalian, China). Informed consent was signed by all participants, and the present study was approved by the ethics committee of the Sixth People's Hospital of Dalian, Dalian, China. The transportation and storage of the samples were according to the standard procedures of metabolomics studies [44]. One mixing sample, defined as batch-QC (b-QC), was prepared by mixing the aliquots from each 10 samples in a batch. The other QC sample, named as pooled-QC (p-QC), was prepared by pooling a large serum samples in the biobank of our laboratory, which was independent of the 10 studied samples. They were subpackaged and stored in the −80 °C.

For sample preparation, a 400 μL acetonitrile including ISs of choline-d4 (0.26 μg/mL), Phe-d5 (3.61 μg/mL), Trp-d5 (4.26 μg/mL), CA-d4 (1.85 μg/mL), carnitine C10:0-d3 (0.10 μg/mL), was added into 100 μL sample for the protein precipitation, then, the sample was vortexed for 60 s and then centrifuged for 10 min at 4 °C and 14,000 rpm (BiofugeStratos, Thermo Scientific, USA). A portion of 200 μL supernatant was transferred to an Eppendorf tube for lyophilizing. 50 μL solution