Journal of Chromatography A, 1459 (2016) 101–111

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Performance evaluation of tile-based Fisher Ratio analysis using a benchmark yeast metabolome dataset

Nathanial E. Watson[a,b], Brendon A. Parsons[a], Robert E. Synovec[a,*]

[a] Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98195, USA
[b] Department of Chemistry and Life Science, United States Military Academy, West Point, NY 10996, USA

ARTICLE INFO

Article history:
Received 8 April 2016
Received in revised form 17 June 2016
Accepted 21 June 2016
Available online 22 June 2016

Keywords:
Comprehensive
Two-dimensional
Gas chromatography
Time-of-flight mass spectrometry
Discovery-based
Metabolomics

ABSTRACT

Performance of tile-based Fisher Ratio (F-ratio) data analysis, recently developed for discovery-based studies using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC × GC–TOFMS), is evaluated with a metabolomics dataset that had been previously analyzed in great detail, but while taking a brute force approach. The previously analyzed data (referred to herein as the benchmark dataset) were intracellular extracts from *Saccharomyces cerevisiae* (yeast), either metabolizing glucose (repressed) or ethanol (derepressed), which define the two classes in the discovery-based analysis to find metabolites that are statistically different in concentration between the two classes. Beneficially, this previously analyzed dataset provides a concrete means to validate the tile-based F-ratio software. Herein, we demonstrate and validate the significant benefits of applying tile-based F-ratio analysis. The yeast metabolomics data are analyzed more rapidly in about one week versus one year for the prior studies with this dataset. Furthermore, a null distribution analysis is implemented to statistically determine an adequate F-ratio threshold, whereby the variables with F-ratio values below the threshold can be ignored as not class distinguishing, which provides the analyst with confidence when analyzing the hit table. Forty-six of the fifty-four benchmarked changing metabolites were discovered by the new methodology while consistently excluding all but one of the benchmarked nineteen false positive metabolites previously identified.

1. Introduction

Knowledge of a biochemical system is incomplete without the inclusion of the metabolome. The study of metabolomics attempts to identify and quantify the low molecular weight compounds which make up the metabolome [1]. Though not specifically mentioned in the central dogma of molecular biology [2], identified metabolites generally complement and reinforce determinations made of the genome and proteome. In other words, an up/down regulated gene should correspond to a correlated change in the proteome, which generally also induces some measureable change in the metabolome [2]. Conversely, unexpected discoveries in the metabolome could suggest changes in the upper realms of the molecular biology hierarchy.

Analytically, the metabolome can be approached in one of two ways, targeted analysis or non-targeted analysis [1]. Targeted analysis, attempts to verify previous genomic and proteomic results through identification of complementary compounds in the metabolome. Targeted approaches focus on expected changes in the metabolome induced by up and down regulated genes and proteins. The foreknowledge of interesting compounds enables the analyst to define detailed methods and techniques to ensure the targeted metabolites are identified and quantified to the necessary level of fidelity. The alternative is to collect a biological sample of interest and submit it for some form of broad analysis. These non-targeted or discovery-based approaches employ non-specific analytical procedures to identify metabolomic changes induced by experimental or environmental perturbations.

Unlike targeted approaches which are amenable to specific assays or similar techniques, non-targeted metabolic investigation requires instrumentation that probes the sample in a wide-ranging and general way. Instrumental platforms that provide the ability to simultaneously separate and identify the resultant sample

* Corresponding author.
E-mail address: synovec@chem.washington.edu (R.E. Synovec).

http://dx.doi.org/10.1016/j.chroma.2016.06.067
0021-9673/

components are desirable. Examples include liquid chromatography (LC) [3], gas chromatography (GC) [4], mass spectrometry (MS) [5], nuclear magnetic resonance spectroscopy (NMR) [6,7], or various combinations of these techniques. Comprehensive two-dimensional gas chromatography coupled with time-of-flight-mass spectrometry (GC × GC–TOFMS) [1] is arguably one of the best analytical instrumentation platforms to study metabolites of interest that reside in the 50–500 Da mass range. The use of GC–MS for metabolomics studies is widespread, but the benefits of GC × GC–TOFMS applied to metabolomics have begun to emerge in more recent years. While GC × GC–TOFMS is an outstanding instrumental platform for biological and metabolomics studies [8–26], there is an ongoing need to develop software methods to glean useful information from the immensely complex data. For this purpose Fisher Ratio (F-ratio) analysis has been found to be a particularly useful algorithm for the analysis of these complex datasets [27–31].

F-ratio analysis compares the variance between classes relative to the variance within the classes [32–34]. Specifically, the F-ratio compares whether these two variances are different relative to a tabulated F-statistic or other means of threshold determination. When the F-ratio is calculated, a quotient is determined, ranging from zero to infinity; as the value increases the results suggests with greater certainty that the compared variances are different implying that the source data is also different. Points with high F-ratios generally correspond to features which distinguish between the classes compared. Previously, F-ratio analysis using GC × GC data was applied to several biological and petrochemical models [26–31]. The prior pixel-based F-ratio studies were fruitful, and provide a benchmark for current software development and evolution.

Specifically, Mohler, et al. [25,26] applied the F-ratio method to intracellular extracts from *Saccharomyces cerevisiae* (yeast) either metabolizing glucose (repressed) or ethanol (derepressed), which define the two classes in the discovery-based analysis to find metabolites that are statistically different in concentration between the two classes. These data were studied in a "pixel-wise" fashion whereby each data point (defined by chromatographic time on column 1 (^{1}D), chromatographic time on column 2 (^{2}D), and the mass-to-charge ratio (m/z)) across all the sample replicates was compared as a subset to calculate an F-ratio, which statistically quantifies the variances for the complete dataset at that one data point pixel. The F-ratio calculation was repeated iteratively for every data point collected. Beneficially, the Mohler et al. dataset was large and well defined biologically. Each culture was thrice replicated providing three samples for each culture class. Also, the samples were extracted in triplicate and each extraction was chromatographically analyzed four times. The copious replication ensured the different sources of variance were sufficiently quantified. The result of this study was a list of metabolites determined to be variant, or up/down regulated and hence changing, between the two sample classes ordered by F-ratio value. The list, referred to as a "hit list," ordered the metabolites from greatest to least F-ratio. This hit list was fully analyzed and quantified through application of parallel factor analysis (PARAFAC) [35]. Usefully, this analysis provided empirical evidence for biochemically suggested metabolic pathways. Regrettably, the identification and quantification of variant metabolites was onerous, since the true positives were found to be intermingled with a significant number of false positive, and data analysis consumed greater than a person-year of labor to execute. The analyst was required to manually interdict the process at many steps and analyze the hit list down to the lowest level achievable without a useful means of deciding when the hit list was complete. In retrospect, the analysis was specifically confounded by two issues we hope to elucidate and correct in this current study: retention time shifting on ^{1}D and ^{2}D, and undesirable false-positive discoveries.

Retention time shifting adds to the difficulty in the analysis of any chromatographic dataset. In the Mohler, et al. study the retention time shift reduced the sensitivity in which true positives were discovered due to the pixel-based comparison of data. Since the yeast experiment was conducted over the course of months and required many extra chromatographic runs in addition to the extracted samples (e.g. solvent blanks, growth medium blanks, etc.), retention time shifting became especially pronounced. Temperature and pressure fluctuations combined with matrix effects and stationary phase degradation may all lead to retention time variations. Many times alignment [36,37] or warping [38] is chosen as a solution to correct for retention time changes. Alignment, while not particularly time consuming, can impart unfortunate artifacts to the data as the peaks are warped and time-shifted. The tile-based F-ratio software is designed to address this challenge without explicitly requiring alignment [29,30].

The other challenge, how to minimize the false positive discoveries, is tied to determining a useful F-ratio threshold value under which further analysis is deemed fruitless. Statistics has come up with two schools of thought on the topic of threshold determination [39,40]. One has the analyst consult a tabular solution for the applied statistic, in this case the F-test. Based on the degrees of freedom and number of samples, the analyst determines an F-critical value and judges that all features with calculated F-ratios greater than F-critical do significantly differ between/among classes. The other school of thought, less well known outside statistics circles, suggests that the tabulated values for the textbook distribution may not accurately apply to the experiment at hand. This group is known as the "frequentists." The frequentist camp approaches the problem of determining the threshold with the expectation that every experiment includes many underlying random and statistical errors which may not be appropriately represented by a Bayesian Distribution.

To address this challenge of determining an appropriate threshold to prudently guide hit list data mining, combinatorial null distribution analysis is coupled with tile-based F-ratio analysis [30,31], leveraging the large volume of related data that may be tested by GC × GC–TOFMS to experimentally determine the distribution of potential false positives. By rearranging the sample classes to nullify the class distinguishing variance, the effects of non-meaningful variation in the dataset may be estimated for the true class comparison. To increase the extensibility of the approach, it is desirable to determine all possible null comparisons in a rigorous and automated fashion. With the recent improvements to the computational performance of the software, it is now possible to glean the benefits of using the complete set of null distributions. By algorithmically determining all possible null combinations it is possible to utilize the resulting Fisher Distributions to determine a rigorous threshold.

Herein tile-based F-ratio analysis is evaluated and validated by application to the already thoroughly analyzed fermenting and respiring yeast benchmark dataset from the Mohler et al. study [26]. These data were completely identified and quantified previously and provide an insightful opportunity to compare the tile-based F-ratio analysis [29–31], which in turn had evolved from pixel-based F-ratio analysis [26–28]. The novelty of the study herein is based upon the simultaneous demonstration of improved F-ratio analysis software performance for a complex metabolomics dataset, which is manifested as a decisive improvement in the ranking of true positive hits above false positive hits, in concert with the demonstration that the null distribution threshold accurately identifies this optimized transition from true positive hits to false positive hits in the hit list. In addition, quantification by ChromaTOF (LECO software) is utilized in lieu of PARAFAC to demonstrate the possibilities of rapid peak find-