Journal of Proteomics 129 (2015) 83–92

Contents lists available at ScienceDirect

Journal of Proteomics

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

Comparative analysis of statistical methods used for detecting differential expression in label-free mass spectrometry proteomics☆

Sarah R. Langley [a,b,*], Manuel Mayr [b]

[a] Division of Brain Sciences, Imperial College Faculty of Medicine, London, UK
[b] King's British Heart Foundation Centre, King's College London, London, UK

ARTICLE INFO

Article history:
Received 9 March 2015
Received in revised form 10 July 2015
Accepted 13 July 2015
Available online 18 July 2015

Keywords:
Label-free mass spectrometry
Spectral counts
Differential expression
Statistical methodology

ABSTRACT

Label-free LC-MS/MS proteomics has proven itself to be a powerful method for evaluating protein identification and quantification from complex samples. For comparative proteomics, several methods have been used to detect the differential expression of proteins from such data. We have assessed seven methods used across the literature for detecting differential expression from spectral count quantification: Student's t-test, significance analysis of microarrays (SAM), normalised spectral abundance factor (NSAF), normalised spectral abundance factor-power law global error model (NSAF-PLGEM), spectral index (SpI), DESeq and QSpec. We used 2000 simulated datasets as well as publicly available data from a proteomic standards study to assess the ability of these methods to detect differential expression in varying effect sizes and proportions of differentially expressed proteins. At two false discovery rate (FDR) levels, we find that several of the methods detect differential expression within the data with reasonable precision, others detect differential expression at the expense of low precision, and finally, others which fail to identify any differentially expressed proteins. The inability of these seven methods to fully capture the differential landscape, even at the largest effect size, illustrates some of the limitations of the existing technologies and the statistical methodologies.

Significance: In label-free mass spectrometry experiments, protein identification and quantification have always been important, but there is now a growing focus on comparative proteomics. Detecting differential expression in protein levels can inform on important biological mechanisms and provide direction for further study. Given the high cost and labour intensive nature of validation experiments, statistical methods are important for prioritising proteins of interest. Here, we have performed a comparative analysis to investigate the statistical methodologies for detecting differential expression and provide a reference for future experimental designs.

This article is part of a Special Issue entitled: Computational Proteomics.

1. Introduction

Protein detection and quantification have vastly improved in recent years with the technological advances of mass spectrometry. Liquid chromatography tandem mass spectrometry (LC-MS/MS) has become the method of choice for quantitative proteomics and can now assess protein samples in a bottom-up format with reasonable throughput. There are several methods for tagged or isotope labelled quantification, including isobaric tags for relative and absolute quantitation (iTRAQ) [1], tandem mass tags (TMT) [2] and stable isotope labelling by amino acids in cell culture (SILAC)[3]. These methods offer multiplexing capability at the requirement of more complex protocols and expensive reagents. However, SILAC is unsuitable for clinical samples and the tagged methods have the limitation that co-isolation of multiple precursor ions can interfere with accurate quantitation. Instead, label-free methods aim to provide relative quantification without isotopic labelling and are becoming increasingly popular in proteomics [4–6].

For label-free proteomics, one can quantify proteins by using their spectral counts as an approximation of protein abundance. Spectral counts are simply the total number of spectra per identified protein and can be easily calculated from the detected peptides by LC-MS/MS; within a protein, they can be taken as an semi-quantitative approximation as a protein with higher abundance in one group should have more identified spectra than the protein with lower abundance in another. Several methods have been proposed and applied which take advantage of the relationship between the spectral counts and protein abundance to detect differential expression. The primary goal of a differential expression analysis is to detect as many truly differentially expressed proteins as possible (reducing the number of false negatives or type II errors) while controlling for the number of false positives (type I errors). As label-free methods can quantify hundreds to thousands of proteins, multiple testing corrections must be applied to differential expression

☆ This article is part of a Special Issue entitled: Computational Proteomics.

* Corresponding author at: Duke-NUS Graduate Medical School Singapore, 8 College Road, 169857 Singapore, Republic of Singapore.
E-mail addresses: sarah.r.langley@duke-nus.edu.sg, s.langley@imperial.ac.uk (S.R. Langley).

http://dx.doi.org/10.1016/j.jprot.2015.07.012
1874-3919/

analyses to control the number of false positives or type I errors. One approach is to control the false discovery rate (FDR), which is the expected proportion of false positives within a set of significantly differentially expressed proteins. For example, if one had 100 proteins which are detected as differentially expressed at a 5% FDR, five of them are expected to be false positives. This is a separate FDR measure than the one associated with protein inference and identification.

In this study, we chose seven methods for identifying significant differences in spectral count based protein expression. These methods were chosen from the literature and included methods originally proposed for differential expression analysis in microarrays and RNA-seq as well as those specific to proteomics. We included the significance analysis of microarrays (SAM) [7] and the normalised spectral abundance factor coupled with a power law global error model [8]; both methods were designed for gene expression microarray data and have been used for the analysis of label-free MS proteomics [9]. The spectral index (SpI) [10] and QSpec [11] methods were included as methods which were developed specifically for spectral count quantification and have been used in several studies [12,13]. Others have now taken advantage of the methods developed for RNA-seq experiments and applied them in spectral count proteomic studies [14,15], so we have also included the DESeq method [16]. Finally, we included the *t*-test and normalised spectral abundance factor (NSAF) [17] coupled with the *t*-test. The *t*-test is one of the most commonly used statistical tests and has been used to detect differential protein expression [18].

To evaluate these methods, we used 2000 simulated datasets as well as data with a known spike-in difference from the CPTAC standards assessment [19,20]. We investigated the ability of these seven methods to identify differential expression with respect to several different measures; 1) effect sizes, or the percentage of abundance difference, 2) proportion sizes, or the percentage of proteins within a dataset that are differentially expressed and 3) at two levels of multiple testing corrections. Within this evaluation, we provide insight into the performance of these methods with respect to these measures and suggestions for their use in future proteomic studies.

2. Materials and methods

2.1. Simulated data

Real data from two previously published studies were used as the basis for the simulated data (Fig. 1). The first dataset was from LC-MS/MS study investigating the proteomic changes resulting from the addition of an exogenous matrix metallopeptidase within a population of three cases and three controls [21]. The second was a shotgun proteomic analysis of hibernating arctic squirrels within a population of four cases and four controls [9]. 2000 datasets were simulated — 1000 based on the data from the matrix metallopeptidase study (denoted D1) and 1000 based on the data from the arctic squirrel study (denoted D2). In D1, each of the 1000 datasets consisted of simulated counts from 606 proteins and in D2, each of the 1000 datasets consisted of simulated counts from 3538 proteins. Spectral count data can be modelled as a Poisson distribution where the probability of observing a count, n, with respect to the expected count, γ, is given in Eq. (1).

$$f(n,\gamma) = \frac{\gamma^n e^{-\gamma}}{n!} \tag{1}$$

In our simulations, we set γ to the average spectral count of an individual protein from a real dataset and used it to derive a set of Poisson distributed random deviates to simulate the spectral counts for a given protein. This was to preserve the relationship between the spectral count abundance and protein length, as the number of amino acids is used by several of the statistical methods. To incorporate effect sizes into the simulated data, we randomly sampled 20 simulated proteins and added additional counts to one group as follows

$$\overline{SC}_{i,j} = SC_{i,j} \times (1+p) \tag{2}$$

where $SC_{i,j}$ is the simulated spectral count from the *jth* sample of protein i and p is one of 0.2 (20%), 0.5 (50%), 0.8 (80%), 1 (100%), and 2 (200%). In both D1 and D2, one hundred datasets at each effect level were simulated, resulting in 500 datasets from each. The set of 500 (100

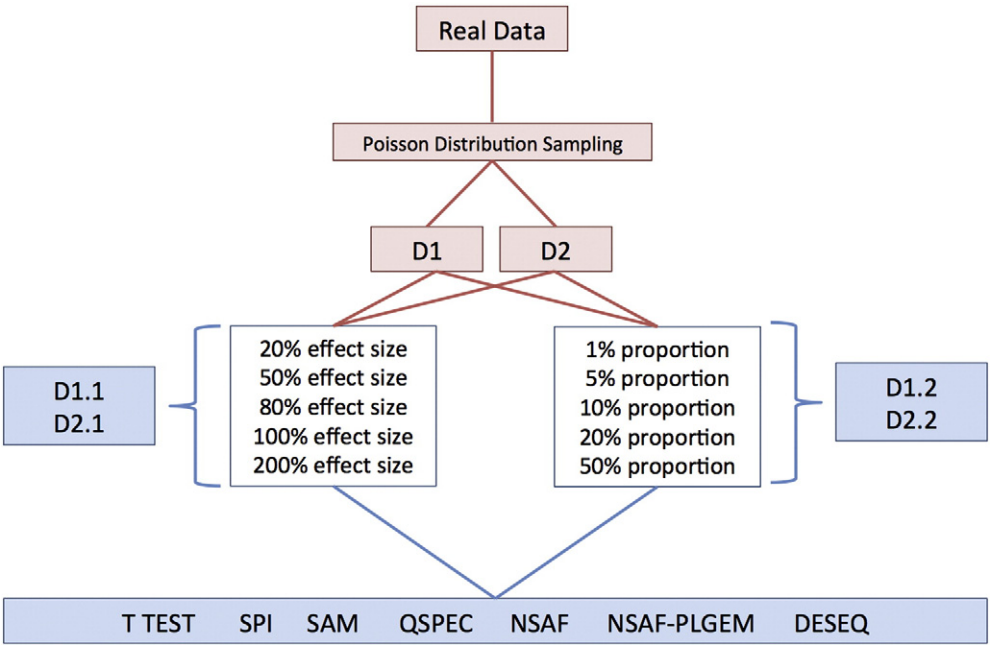

Fig. 1. Simulation data scheme. Overview of the simulated data generation with different effect sizes and different proportions of differentially expressed proteins. *T* TEST — Student's *t*-test; SPI — spectral index; SAM — significance analysis of microarrays; QSPEC — QSpec; NSAF — normalized spectral abundance factor; NSAF-PLGEM — normalized spectral abundance factor-power law global error model; DESEQ — DESeq.