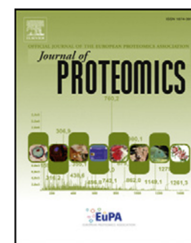
Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot



An effect size filter improves the reproducibility in spectral counting-based comparative proteomics☆

Josep Gregori[a,b], *Laura Villarreal*[a], *Alex Sánchez*[b,c], *José Baselga*[a], *Josep Villanueva*[a,*]

[a]*Vall d'Hebron Institute of Oncology (VHIO), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain*
[b]*Statistics Department, University of Barcelona (UB), Barcelona, Spain*
[c]*Statistics and Bioinformatics Unit, Vall d'Hebron Institut de Recerca, Barcelona, Spain*

ARTICLE INFO

Available online 11 June 2013

Keywords:
Effect size
Feature filters
Poisson
Quasi-likelihood
edgeR
QSpec

ABSTRACT

The microarray community has shown that the low reproducibility observed in gene expression-based biomarker discovery studies is partially due to relying solely on p-values to get the lists of differentially expressed genes. Their conclusions recommended complementing the p-value cutoff with the use of effect-size criteria. The aim of this work was to evaluate the influence of such an effect-size filter on spectral counting-based comparative proteomic analysis. The results proved that the filter increased the number of true positives and decreased the number of false positives and the false discovery rate of the dataset. These results were confirmed by simulation experiments where the effect size filter was used to evaluate systematically variable fractions of differentially expressed proteins. Our results suggest that relaxing the p-value cut-off followed by a post-test filter based on effect size and signal level thresholds can increase the reproducibility of statistical results obtained in comparative proteomic analysis. Based on our work, we recommend using a filter consisting of a minimum absolute $\log_2$ fold change of 0.8 and a minimum signal of 2–4 SpC on the most abundant condition for the general practice of comparative proteomics. The implementation of feature filtering approaches could improve proteomic biomarker discovery initiatives by increasing the reproducibility of the results obtained among independent laboratories and MS platforms.

Biological significance

Quality control analysis of microarray-based gene expression studies pointed out that the low reproducibility observed in the lists of differentially expressed genes could be partially attributed to the fact that these lists are generated relying solely on p-values. Our study has established that the implementation of an effect size post-test filter improves the statistical results of spectral count-based quantitative proteomics. The results proved that the filter increased the number of true positives whereas decreased the false positives and the false discovery rate of the datasets. The results presented here prove that a post-test filter applying a reasonable effect size and signal level thresholds helps to increase the reproducibility of statistical results in comparative proteomic analysis. Furthermore, the implementation of feature filtering approaches could improve proteomic biomarker discovery initiatives by increasing the reproducibility of results obtained among independent laboratories and MS platforms.

This article is part of a Special Issue entitled: Standardization and Quality Control in Proteomics.

© 2013 Elsevier B.V. All rights reserved.

☆ This article is part of a Special Issue entitled: Standardization and Quality Control in Proteomics.

* *Corresponding author at:* Vall d'Hebron Institut d'Oncologia, Psg. Vall d'Hebron 119-129, Barcelona, Spain. Tel.: +34 934894175.
E-mail address: jvillanueva@vhio.net (J. Villanueva).

1874-3919/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.05.030

1. Introduction

The gene expression microarray (MA) community has pioneered the development and implementation of statistical tools for comparative analysis in the *Omics* field [1,2]. *Omics* technologies are in general challenged by a high dimensionality problem, whereby a high number of statistical tests are carried out at the same time on very few samples. The high number of tests that are carried out simultaneously requires a p-value adjustment to control the false discovery rate (FDR). The low number of replicates, combined with a high number of statistical tests easily leads to low reproducibility in the results. Shi et al. addressed this issue in a systematic manner using a large dataset elaborated within the studies of the MicroArray Quality Control Consortium (MAQC) [3]. The MAQC studied the possible sources of the limitations of microarray (MA) studies that precluded obtaining reproducible results across laboratories and platforms. Their conclusions underscored the influence of batch effects and the limitation introduced by using only the p-value criteria to get ordered lists of differentially expressed genes (DEG) as the main weaknesses in MA studies [4,5]. In a recent work, we established the existence of batch effects in the practice of quantitative proteomics, and we proposed a correction of batch effects for spectral counting-based label-free comparative proteomics [6].

The MAQC project pointed out that the observed low reproducibility of DEG lists could be partially attributed to the fact that they relied solely on the p-values. Their conclusions recommended using a non-stringent p-value cut-off in favor of effect size criteria, like a fold-change, in order to obtain reproducible lists across laboratories and platforms [3,4]. This recommendation points to the use of a post-test filter excluding features that show low p-values but poor effect size. There has been a great interest in feature filtering by the gene expression community, since only 40% of the thousands of probes in a typical array are expected to bear any biological information. The consequence of having a high number of features in gene expression comparative analysis is the need to perform a multiple testing adjustment. The caveat of this technique is that—given that it yields smaller p-values—the probability of false negatives is increased and hence informative genes may be lost during data analysis. A common filter used is a non-specific (independent of the class labels) pre-test filter devised to remove non-informative probe sets [7]. The pertinence and side effects of these filters are still under debate [8]. Although filters are fully independent from the test, they may cause a change in the distribution of null p-values, causing a bias in the p-value adjustment with FDR control [9].

The concept of non-informative genes in MA analysis has no translation to LC–MS/MS-based proteomics. On the one hand, only half of the probe sets in an array are expected to bear biological information. On the other hand, the number of features involved in a LC–MS/MS-based proteomic experiment is generally much smaller than in a gene expression experiment, and no assumption on the percentage of informative features can be done a priori. Nevertheless, the use of post-test filters could be helpful to improve the reproducibility of proteomic studies, mainly by restraining the number of false positives. There is a precedent in spectral counting-based proteomics for this type of filters. The QSpec method uses a fold change filter to implicitly alleviate the number of false positives produced by the test [10].

Based on the MAQC conclusions on MA analysis, we decided to investigate the effects of post-test feature filtering on statistical tests used for Spectral Counting-based label-free proteomics. The filters applied are based on effect size and signal strength. In this work five common statistical tests used in Spectral Counting have been evaluated: Poisson Generalized Linear Model (GLM), quasi-likelihood GLM, QSpec, edgeR and the t-test on the square root transform of the SpC. The Poisson GLM is a log linear model based on the Poisson distribution and it is indicated when the number of replicates ranges from 1 to 3, and the expected biological variability is low. Its limitation is that the variability explained by the model equals the mean expression of the dataset, and it is not able to explain extra sources of variability [11]. The quasi-likelihood GLM is a distribution free log linear model that estimates dispersion from the data and requires a higher number of replicates, above 3, to estimate the dispersion parameter with enough precision [11,12]. This test is able to accommodate both under- and over-dispersion (lower and higher variability than the Poisson distribution). QSpec is a GLM of mixed effects that was developed to accommodate for a high biological variability (overdispersion) [10]. Besides, it uses a hierarchical Bayes to estimate the model parameters globally from all replicates and across proteins, allowing for obtaining good results with few replicates. The authors claim that the QSpec method may be used with as few as two replicates. With no replicates, QSpec assumes the Poisson distribution model. edgeR is a method that uses empirical Bayes estimation and negative binomial distribution-based testing. It was developed for SAGE and RNA-seq to accommodate for overdispersion [13]. To estimate the model parameters for each protein, the method combines information from that protein with that borrowed from the whole set of proteins. This is intended to avoid unstable parameter estimates due to small sample sizes. Finally, the t-test or the ANOVA on the square root transform of the SpC works well when the number of replicates is greater than four and significant features are not expected at low signals. From the dispersion point of view, both QSpec and edgeR limit with the Poisson distribution at low dispersion values and are not able to accommodate for underdispersion.

In this work, we have implemented an effect size filter following the recommendations of MAQC, which were given to improve the reproducibility of MA studies, adapting them to quantitative proteomics. We show the effects of post-test filtering on five representative statistical tests used in spectral counting-based differential proteomics by means of a series of spiking experiments. We have used a series of statistical metrics (number of true positives (TP), false positives (FP), positive predictive value (PPV), sensitivity and the false discovery rate (FDR)) for monitoring the influence of an effect size filter on the experimental results. Although simple, the spiking experiments model quite well what we observe in our comparative studies on cancer cell line secretomes [6]. Furthermore, we have performed simulations to examine the distribution of sensitivities and FDRs, before and after the post-test filter, when variable fractions of differentially expressed proteins (DEP) and low signal DEP were being evaluated.

Download English Version:

https://daneshyari.com/en/article/7636782

Download Persian Version:

https://daneshyari.com/article/7636782

Daneshyari.com