

Metriculator: quality assessment for mass spectrometry-based proteomics

Ryan M. Taylor¹, Jamison Dance², Russ J. Taylor³ and John T. Prince^{1,*}

¹Department of Chemistry and Biochemistry, ²Department of Computer Science and ³Department of Information Systems, Brigham Young University, Provo, UT 84602, USA

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Quality control in mass spectrometry-based proteomics remains subjective, labor-intensive and inconsistent between laboratories. We introduce Metriculator, a software designed to facilitate long-term storage of extensive performance metrics as introduced by NIST in 2010. Metriculator features a web interface that generates interactive comparison plots for contextual understanding of metric values and an automated metric generation toolkit. The comparison plots are designed for at-a-glance determination of outliers and trends in the datasets, together with relevant statistical comparisons. Easy-to-use quantitative comparisons and a framework for integration plugins will encourage a culture of quality assurance within the proteomics community.

Availability and Implementation: Available under the MIT license at <http://github.com/princelab/metriculator>.

Contact: jtp@chem.byu.edu

Received on July 10, 2013; revised on August 9, 2013; accepted on August 25, 2013

1 INTRODUCTION

As omics-level experiments increase in size and complexity, assessing the quality of a dataset can be a laborious undertaking. This is particularly true of mass spectrometry (MS)-based proteomics, where the spectrometer and associated chromatography exhibit variable—and sometimes erratic—performance. Researchers would like to analyze more samples and in greater depth (i.e. fractionation), but maintaining high quality across the set—and knowing that the set is of high quality—is a mounting challenge in proteomics.

Typically, highly trained technicians spend significant time adjusting capillary plumbing, working to achieve stable nano-electrospray and calibrating and tuning the mass spectrometer. The quality of a large analysis is then assessed by visual inspection of the 2D or 3D ion trace. Assessing reproducibility is particularly challenging when datasets involve runs collected over a period of weeks or months. Furthermore, human assessment of quality is both time-consuming and a potential source of bias (Danziger *et al.*, 2011). Software tools to aid in quality assessment are needed and can improve confidence in published proteomic datasets (Kinsinger *et al.*, 2012).

An extensive set of quality/performance metrics was introduced by NIST to begin to assess data quality in MS proteomics

(Rudnick *et al.*, 2010). These 284 metrics include measures of chromatographic performance, ion source stability, ion signal intensity and data-dependent sampling efficiency. These measures can significantly augment manual interpretation of data quality, but their utility depends on contextual comparisons between datasets. Comparing metrics over time is the key to leveraging them for quality assessment. Subsequent work has attempted to expand metrics across vendor platforms, and to provide curated metrics, such as those demonstrated by QuaMeter (Ma *et al.*, 2012), indicating the value of metrics to the proteomics community.

A similar suite of software designed to monitor lock mass and quality control evaluation at the instrument and identification level, called SIMPATIQCO, was recently released (Pichler *et al.*, 2012). It provides a similar web-based interface designed to assist instrument operators in monitoring quality control samples, yet it lacks interactive graphing and comparison capabilities, which enable applicability to any questions of performance differences, as well as augmentation to existent workflows. Additionally, there is a commercial product, MassQC, which provides a utility for submitting metric information for longitudinal comparison, but does not provide automation. There remains a need for an open-source configurable method for tracking and comparing performance metrics, as well as integration into a proteomics workflow.

2 SOFTWARE

Here we introduce Metriculator, an easily installable database backed web service that generates, stores and compares metrics across datasets for quality control purposes. We also provide archival features to facilitate automatic metric generation of the NIST metrics and workflow integration. This package is meant to serve as a framework for an automated workflow customizable by each research group. We chose Ruby to enable users to easily extend the framework through our integration plug-in setup; Ruby is easy to learn (Prince *et al.*, 2008), boasts a large number of off-the-shelf utilities for web programming, and is gaining traction in the scientific community (SciRuby, 2013).

2.1 Implementation

Metriculator is cross-platform, tested on *NIX systems and Windows, and only requires an installation of Ruby (version ≥ 1.9) to function. The interface is built on Rails, the popular web framework, and provides interactive graphs through the HighCharts library (Highsoft Solutions AS, 2012).

*To whom correspondence should be addressed.

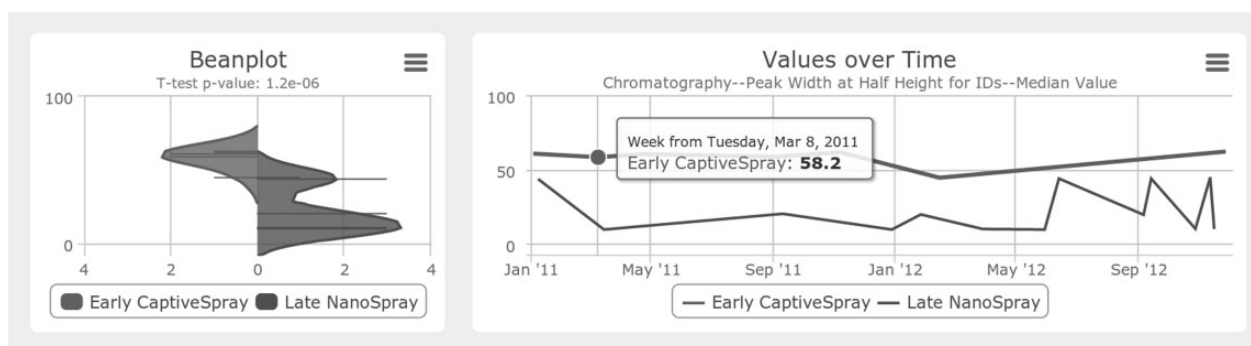


Fig. 1. Visualization Plot, representative of those generated by the software, showing the combination of a bean plot with the associated time plot for a single metric, comparing 7 values with 13 values in a second dataset

2.2 Metric generation

Metriculator automates generation of the 284 NIST performance metrics from LC-MS run *.RAW files and stores them in a database to ensure that relevant meta-information can be compared over time even if raw data are lost. The NIST metrics are generated by the NIST MSQC software, which is platform specific for Windows, as detailed by NIST (Rudnick *et al.*, 2010).

2.3 Web interface

The browser interface is designed for easy access to the stored metric information. Through it, a user can access any of the meta-information about a run and can generate a comparison of all metric values between two lists of msruns. These comparisons provide graphs designed to provide at-a-glance evaluation of the metric information. The website also provides for email-based alerts, customizable via a QC_alerts configuration file, which specifies a notification threshold, deviations from the historic mean for each metric, to trigger an alert email, notifying technicians of instrument problems immediately. This utility could easily be expanded to hook into any notification systems.

2.4 Visualization

The graphs incorporate a time-rendered plot of the datapoints for each set of msruns, as well as a visual comparison of the two populations by beanplot (Kampstra, 2008). Beanplots provide a compact visual summary and comparison between two distinct populations without sacrificing visualization of potentially interesting individual datapoints. They consist of two vertically plotted density plots, with a univariate plot of individual datapoints. Each plot provides at-test comparison, as well as an immediate visual summary of any trends or significant differences in the comparison sets of msruns. These plots are generated dynamically via Javascript and are interactive to enable simple identifications of anomalous metric values (Figure 1).

3 MS-ARCHIVER—INTEGRATION WITH A WORKFLOW

Metriculator is ideally integrated into a workflow to ensure that metrics are generated for every run. Metriculator allows

for automated data off-loading/backup and metric generation on run completion, to reduce loading of data acquisition systems. Communication between computer nodes is accomplished by a simple file system-based queue; the automation framework only requires access to a shared storage location, common to most laboratories. Complete automation can be achieved through the use of a cascading set of settings files in the archival directory and use of the integration capabilities provided.

In its current scope, Metriculator fills a niche role not covered by existing Laboratory Information Management Systems and analysis software (e.g. CPAS). Although other software provides some utility, the ease of use and plotting capabilities of Metriculator provide more intuitive investigation of metric datapoints, as well as an extensible framework for pipeline management. An open customizable code-base allows others to expand on the software to suit their needs through integrating their own automation tools to the integration plugin framework provided by Metriculator, thereby encouraging metric adoption in the proteomics community.

Conflict of Interest: none declared.

REFERENCES

- Danziger, S. *et al.* (2011) Extraneous factors in judicial decisions. *Proc. Natl Acad. Sci. USA*, **108**, 6889–6892.
- Highsoft Solutions. (2012) *Highcharts JS*. <http://www.highcharts.com/products/highcharts> (31 May 2013, date last accessed).
- Kampstra, P. (2008) Beanplot: a boxplot alternative for visual comparison of distributions. *J. Stat. Softw.*, **28**, 1–9.
- Kinsinger, C.R. *et al.* (2012) Recommendations for mass spectrometry data quality metrics for open access data (corollary to the Amsterdam Principles). *J. Proteome Res.*, **11**, 1412–1419.
- Ma, Z.-Q. *et al.* (2012) QuaMeter: multi-vendor performance metrics for LC-MS/MS proteomics instrumentation. *Anal. Chem.*, **84**, 5845–5850.
- MASSQC (2011). <http://www.massqc.com> (31 May 2013, date last accessed).
- Pichler, P. *et al.* (2012) SIMPATIQCO: a server-based software suite which facilitates monitoring the time course of LC-MS performance metrics on Orbitrap instruments. *J. Proteome Res.*, **11**, 5540–5547.
- Prince, J.T. and Marcotte, E.M. (2008) msPire: mass spectrometry proteomics in Ruby. *Bioinformatics*, **24**, 2796–2797.
- Rudnick, P.A. *et al.* (2010) Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. *Mol. Cell. Proteomics*, **9**, 225–241.
- SciRuby Project (2013). <http://sciruby.com/> (31 May 2013, date last accessed).