Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

OXFORD

Subject Section

MsQuality – an interoperable open-source package for the calculation of standardized quality metrics of mass spectrometry data

Thomas Naake [1], Johannes Rainer,[2] and Wolfgang Huber [1]*

[1] Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, 69117, Germany
[2] Institute for Biomedicine (Affiliated to the University of Lübeck), Eurac Research, Viale Druso 1, 39100 Bolzano, Italy

*corresponding author.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Obtaining high-quality data from mass spectrometry (MS) experiments can be challenging due to various factors that can impact the accuracy and reproducibility of the data. Quality control (QC) techniques are needed to guarantee that the data sets are of high quality.
Results: The `MsQuality` R-package calculates and assesses various quality metrics for mass spectrometry-derived spectral data at the per-measurement level. It calculates around 40 low-level quality metrics based on the controlled vocabulary of the mzQC quality metrics defined by HUPO-PSI. The package helps to identify low-quality measurements and track data quality, ultimately improving the quality and reproducibility of mass spectrometry data for robust scientific discoveries.
Availability: MsQuality is implemented in R and is available through Bioconductor at https://bioconductor.org/packages/MsQuality.
Contact: wolfgang.huber@embl.org
Supplementary information: Supplementary data are available at *Bioinformatics* online.

Mass spectrometry (MS) is a versatile analytical technique that has been adopted in a variety of disciplines, including proteomics, metabolomics, and lipidomics. MS enables the identification and quantification of a wide range of molecules. Obtaining high-quality data from mass spectrometry experiments can be a challenging task, as numerous factors can impact the accuracy and reproducibility of the obtained data. To ensure that MS data is of the highest quality, it is imperative to implement appropriate quality control (QC) techniques to guarantee that the data sets are reliable (Köcher *et al*., 2011).

Here, we introduce the `MsQuality` R-package, which provides functionality to calculate, assess, and track quality metrics for mass spectrometry-derived spectral data at the per-measurement level. The package allows to compute about 40 of the mzQC quality metrics defined by the Human Proteome Organization-Proteomics Standards Initiative (HUPO-PSI, hupo-psi.github.io/mzQC). These are calculated on low-level MS data such as retention times and m/z and associated intensity values.

The package enables tracking and quantification of data quality using multiple metrics even on a large scale and helps to identify measurements that are of low quality, such as those with a high number of missing values, ahead-of-time termination of chromatographic runs, or low instrument sensitivity. Following the definitions from Bittremieux *et al*. (2017), `MsQuality` focuses on the calculation of inter-experiment metrics, which can typically be obtained by summation from an intra-experiment metric. Examples for intra-experiment metrics are the chromatogram of the total ion current (TIC) over the retention time. Inter-experiment metrics, on the other hand, facilitate the comparison of multiple MS runs or experiments, e.g. via longitudinal analysis of quality metrics, such as the fractions of the total retention time required to accumulate a given percentile of the TIC.

1 Usage scenario and implementation

`MsQuality` offers easy-to-use means of evaluating data quality on a per-measurement basis, including the identification of low-quality measurements (e.g. poor intensities, low signal-to-noise ratio), biases and outliers, variations in calibration, and batch and confounding effects within datasets (Fig. 1 a and b). Utilizing community standards

for data representation in mass spectrometry defined by HUPO-PSI, `MsQuality`'s metric calculations enable data quality comparison, storage, and facile exchange.

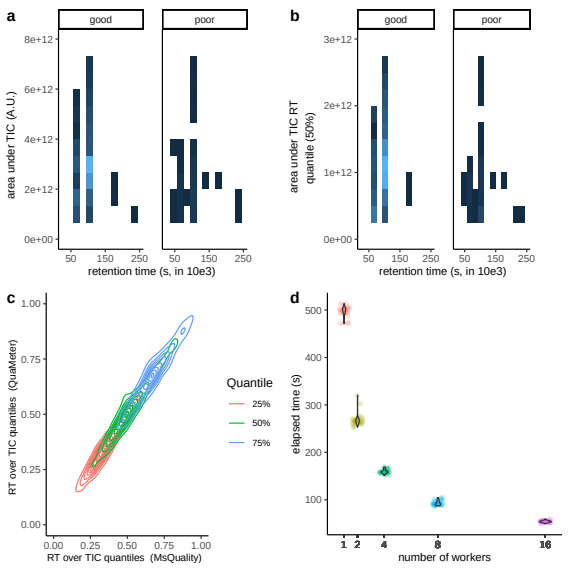

Fig. 1. Examples of `MsQuality` functionality. Metrics are based on MS1 spectra and one data point is obtained per MS1 spectra. (a) Area under TIC: The area under the total ion chromatogram. (b) Area under TIC RT quantiles: The area under the total ion chromatogram of the retention time quantiles. (c) Comparison of quality metrics calculated by `MsQuality` and QuaMeter: RT over TIC quantiles. For (a), (b), and (c), the data points are displayed as 2D densities and stratified for high-quality and low-quality measurements as classified in Amidan et al. (2014) or by the quantiles 25%, 50%, and 75%. Brighter areas correspond to high 2D density areas. (d) Execution time for the calculation of quality metrics of the data set of Amidan et al. (2014) under parallelization (1, 2, 4, 8, and 16 workers). A.U. arbitrary units.

The versatility of `MsQuality` in calculating metrics extends to a wide range of applications, including small-scale studies and long-term acquisition of mass spectrometry data. The utility of `MsQuality` is demonstrated through two case studies, utilizing a previously published 180 cancer cell line dataset obtained by flow injection analysis (Cherkaoui *et al*., 2022) and a liquid chromatography(LC)-MS dataset of the same QC sample (Amidan *et al*., 2014) for long-term quality control. The tool provides consistent results when compared to other data quality tools, such as QuaMeter (Ma *et al*., 2012) (Fig. 1 c) or MatrixQCvis (Naake *et al*., 2022). Correlating the `MsQuality` metrics to pre-calculated QuaMeter metrics (Amidan *et al*., 2014) showed that 75% of the analyzed metrics showed Pearson correlation coefficients over 0.81 and Spearman correlation coefficients over 0.87 (see the Supplementary Data for further details).

`MsQuality` is implemented as an open-source R package, relying on the established `Spectra` and `MsExperiment` packages (Rainer *et al*., 2022) to provide and represent the MS data. By building on these packages `MsQuality` thus supports a large variety of data input formats as well as analyses of very large experiments through the use of data representations with low memory footprint. Native parallelization enables a fast and scalable calculation of quality metrics (Fig. 1 d, see the Supplementary Data for further details).

Finally, `MsQuality` requires little programmatic interaction by using only one function call after the instantiation of `Spectra` or `MsExperiment` objects, thus, is user-friendly to calculate the quality metrics.

2 Conclusion

The software tool, `MsQuality`, contributes to the expanding list of tools that utilize the `Spectra`/`MsExperiment` framework to address various stages in the analysis pipeline of mass spectrometry data. The implementation of `MsQuality`'s metric calculation is designed to be user-friendly and streamlined and requires little programmatic interaction, facilitating reproducible calculations and evaluations of data quality metrics.

By building upon an extensive ecosystem for mass spectrometry data, centered around the `Spectra` and `MsExperiment` packages (Rainer *et al*., 2022), `MsQuality` enables researchers to create seamless analysis workflows for rapid, efficient, and standardized evaluation of MS data quality, ultimately leading to more robust scientific discoveries in mass spectrometry workflows.

3 Acknowledgements

We acknowledge feedback from the SMART-CARE consortium on usability of `MsQuality` and all developers and maintainers of the R/Bioconductor packages `MsQuality` is built upon. We would like to thank Nicola Zamboni for his valuable assistance in understanding the files and pointing to the required files of the study by Cherkaoui et al. (2022).

3.1 Author contributions statement

T.N. conceptualized the R package. T.N. and J.R. implemented the algorithms as an R package. T.N. analysed the results. T.N., J.R., and W.H. wrote and reviewed the manuscript.

3.2 Funding

This work was supported by the Bundesministerium für Bildung und Forschung [grant agreement no. 161L0212E].

Conflict of Interest: none declared.

References

Amidan, B.G. *et al* (2014) Signatures for mass spectrometry data quality. *Proteome Research*, **13**, 2215–2222.

Bittremieux, W. *et al* (2017) Computational quality control tools for mass spectrometry proteomics. *Proteomics*, **17**, 1–11.

Cherkaoui, S. *et al* (2022) A functional analysis of 180 cancer cell lines reveals conserved intrinsic metabolic programs. *Molecular Systems Biology*, **18**, e11033.

Köcher, T. *et al* (2011) Quality control in LC-MS/MS. *Proteomics and Systems Biology*, **11**, 1026–1030.

Ma, Z.-Q. *et al* (2012) QuaMeter: Multivendor Performance Metrics for LC–MS/MS Proteomics Instrumentation. *Analytical Chemistry*, **84**, 5845–5850.

Naake, T. *et al* (2022) MatrixQCvis: shiny-based interactive data quality exploration for omics data. *Bioinformatics*, **38**, 1181–1182.

Rainer, J. *et al* (2022) A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R. *Metabolites*, **12**, 173.