

MetaboScape as a quality-focused software solution for exploring metabolic profiling datasets

Nikolas Kessler¹, Aiko D. Barsch¹, Konstantin Schwarze¹, Matthew R. Lewis¹

¹Bruker Daltonics GmbH & Co. KG, Fahrenheitstraße 4, 28359 Bremen, Germany

Introduction

Metabolic profiling is routinely employed in biomedical research demanding high quality data at various scales of clinical and population phenotyping. In recent years, numerous publications have focused on quality control in metabolomics, describing increased interest and cooperation within the field [1,2]. MetaboScape[®] is a quality-focused software solution for the processing of untargeted metabolic profiling data that integrates the field's best practices to support the user's comprehensive investigation and control of metabolomics data quality. We assessed quality control (QC) measures on a batch of more than 1000 urine sample analyses acquired on a UHPLC-HR-QTOF system. The study design featured regularly interleaved pooled QC samples for investigation of technical variation and use in data filtering, within-batch correction, and quality assessment.

Methods

Three distinct human urine samples and three pairwise mixtures were diluted 1:3 with water and centrifuged. The supernatant was aliquoted for chromatographic separation using a linear reversed-phase gradient with a 15 minute cycle time. LC-MS data (ESI positive mode) were acquired on an Impact II QTOF-MS (Bruker Daltonics). Every seventh injection corresponded to a QC sample which was an equal-parts mixture of each sample in the study (see Figure 1 for the experimental setup) [3]. The ion source region was cleaned and the detector was tuned prior to sample analysis which was conducted in two distinct intended batches.

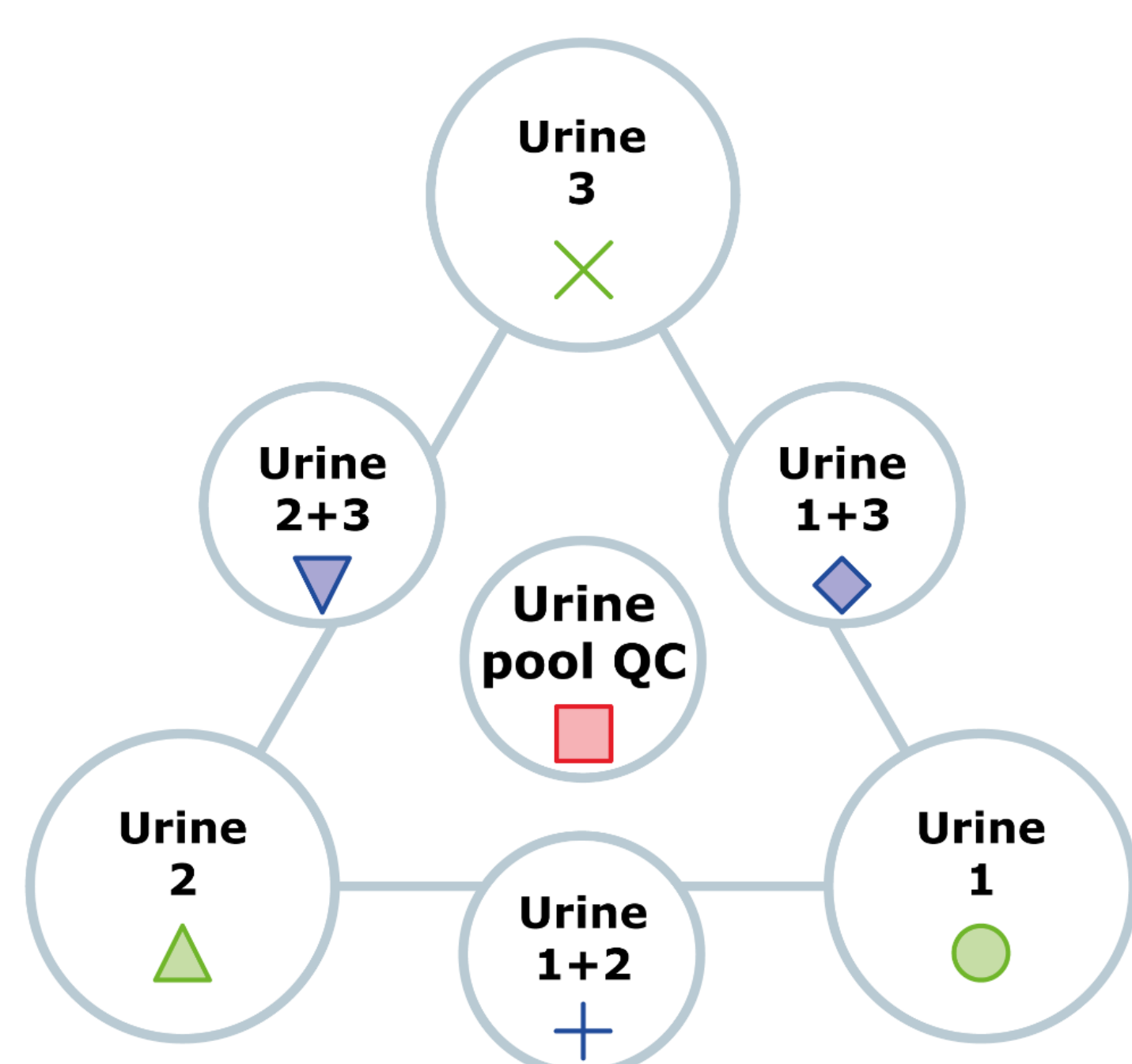


Fig. 1 – Legend of original and pooled samples. The pool of all three original samples is used for as PQC.

Data Processing and Filtering using Pooled Quality Control (QC) Samples

The complete data set was processed with T-ReX[®] 3D in default configuration for LC-ESI-MS data. T-ReX allows to filter for features that were found in a chosen number of analyses. It also allows to filter for features that were found in at least a chosen fraction of a certain group of samples. Using the latter, it is easily possible to focus on features that were found in all QC samples.

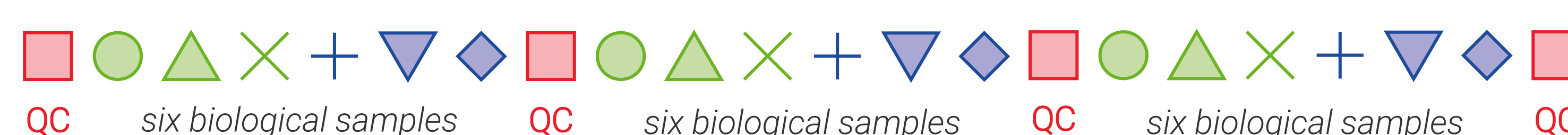


Fig. 2 – The experimental design considers interspersed QC samples. Having QCs in the beginning and at the end, allows the within-batch correction (WBC) to interpolate. Additionally, QCs allow for meaningful filtering of features.

Within-Batch Correction (WBC) and Highlighting of QC Outliers

The WBC in MetaboScape makes use of QC samples to estimate run-order dependent intensity drifts. For every detected feature, a LOESS correction function with a large bandwidth (to avoid overfitting) is calculated, based on the QC intensities. This non-linear, smooth estimate of intensity drift is used to adjust feature intensities of all study samples. The WBC will only interpolate, and therefore it is mandatory that QC samples are run at the start and end of each batch.

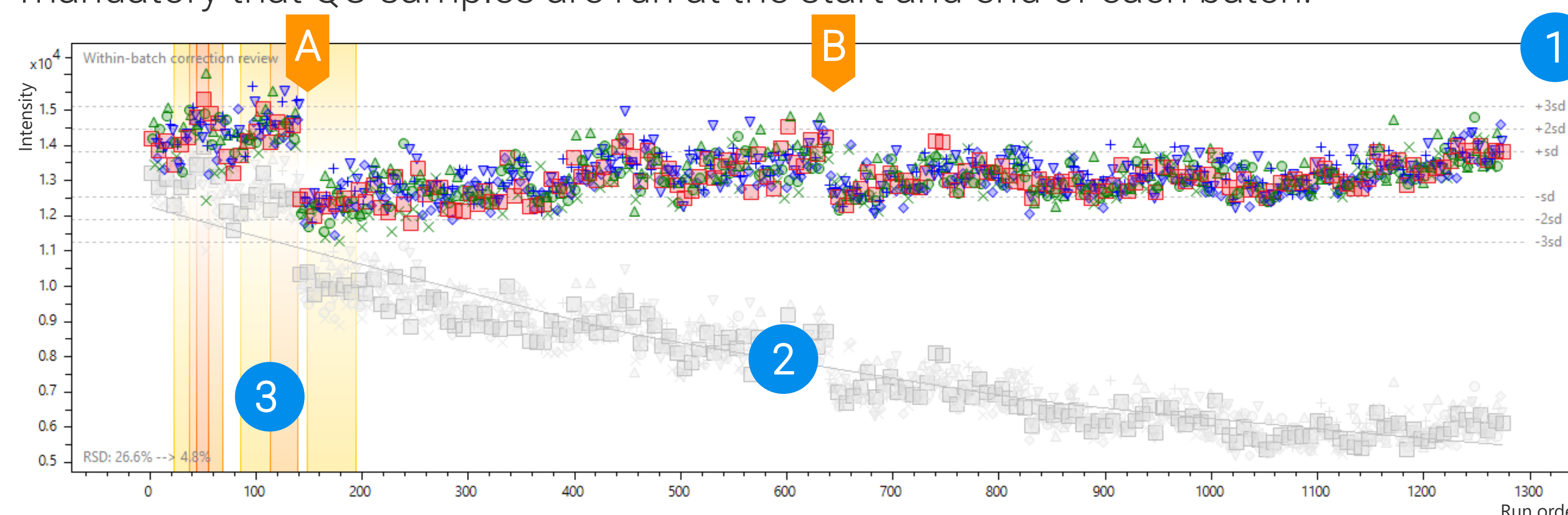


Fig. 3 – The WBC review plot for a strongly decomposing feature putatively annotated as L-Valine, shown before (grey) and after (color) within-batch correction. Legend: see Fig. 1. WBC improved the inter-QC RSD from 26.6% to 4.8%. (1) The dotted lines mark the one, two, and three standard deviations (sdev) ranges, based on QC intensities. (2) The correction function, calculated using the raw QC intensities, is shown as a grey line fitted to the uncorrected data. (3) Outliers are highlighted according to Gika *et al.* [4] in red (one QC outside the three sdev-range), orange (two consecutive PQCs outside the two sdev-range), and yellow (five consecutive QCs outside one sdev-range). (A) A change in LC solvents led to severe disturbance of the batch visible at injection 140. (B) A planned 24h break in the experiment also produced a discontinuity in the data.

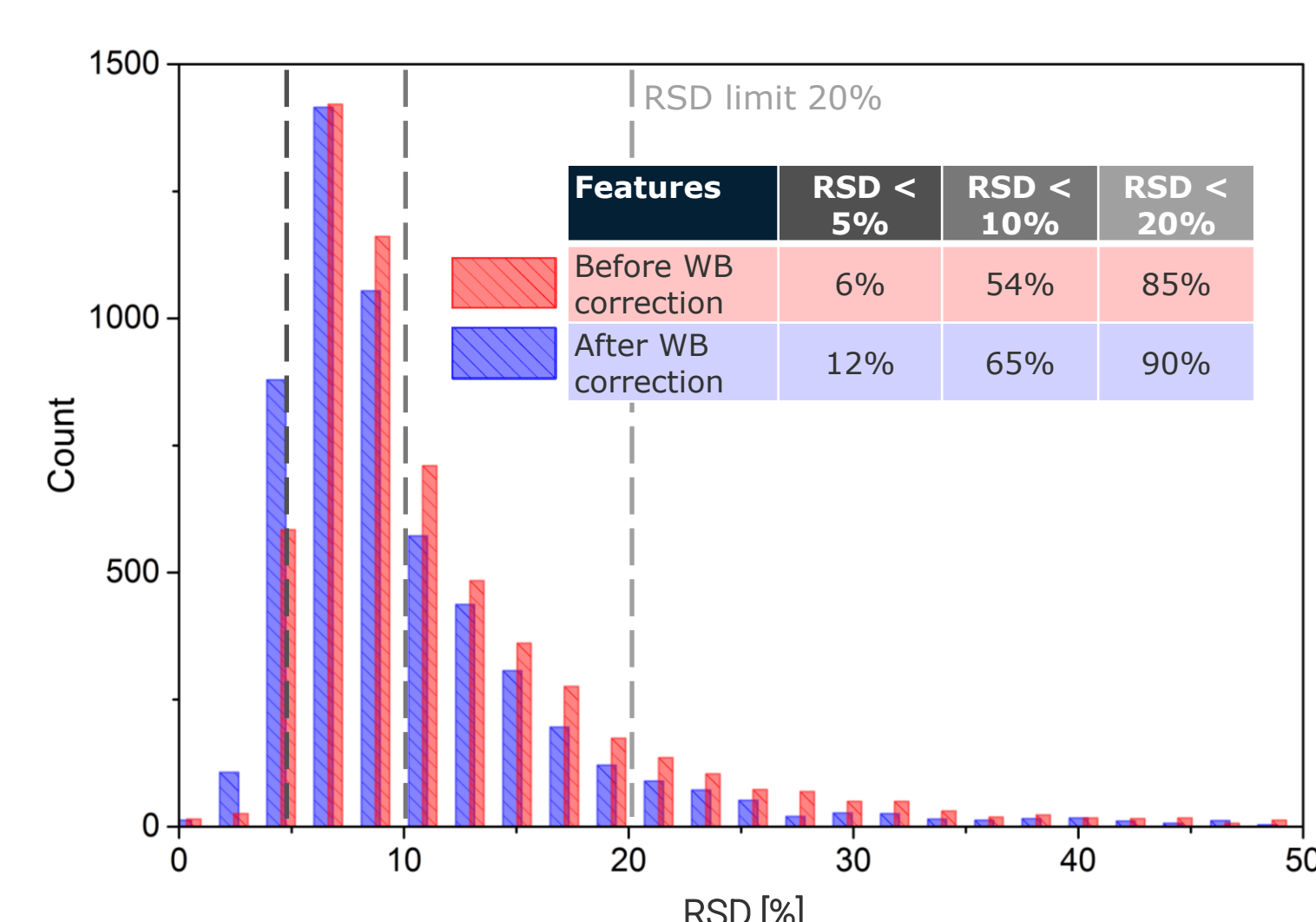


Fig. 4 – The histogram of RSDs before (red) and after (blue) WBC shows a clear improvement of the coefficients of variance for the intensities of all features found in the biological samples. The dotted lines mark RSD thresholds of 5%, 10% and 20% respectively.

Ratio of total and technical variation

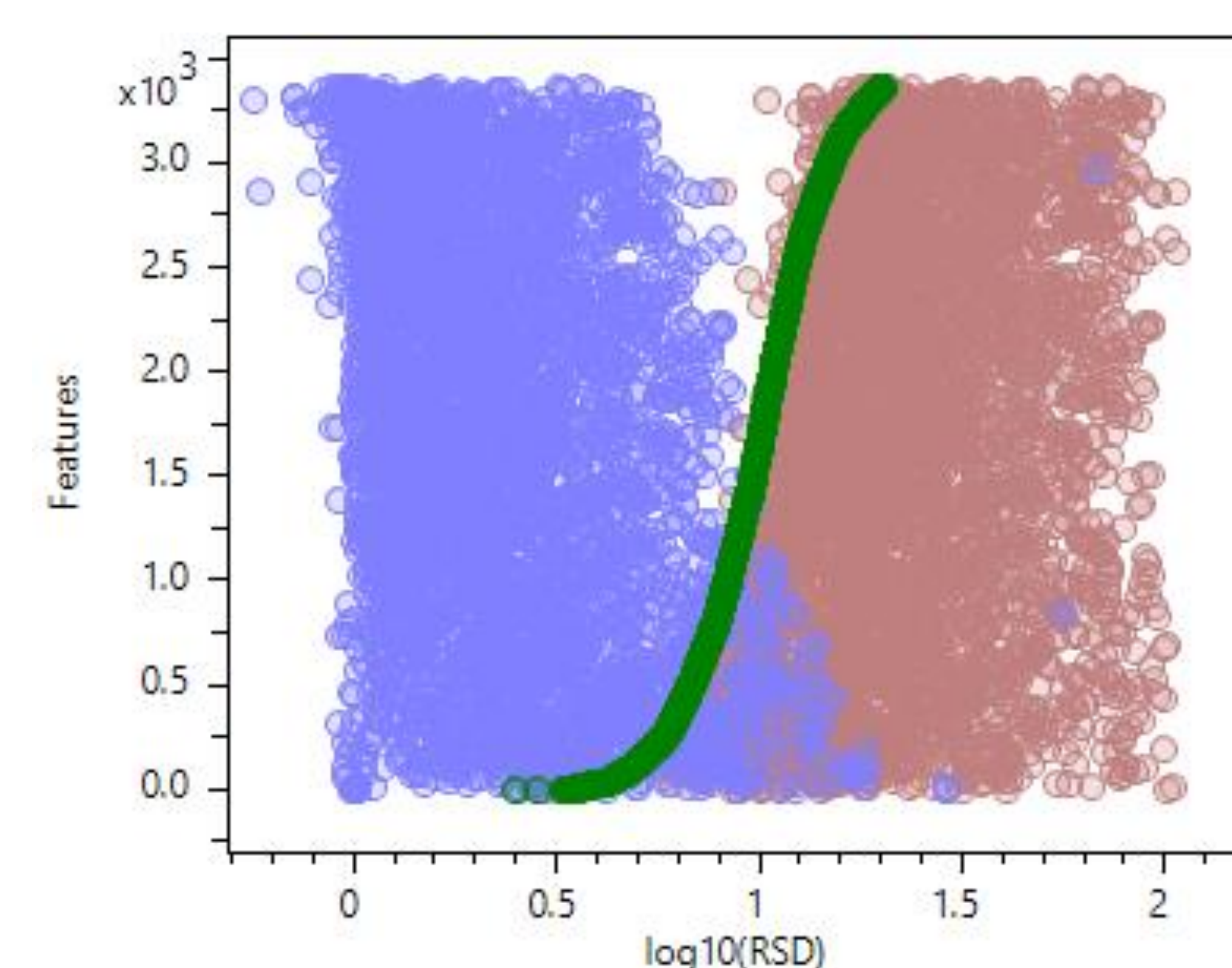


Fig. 5 – The interactive RSD plot sorts features by RSD of feature intensity in QCs (green) and plots the respective biological sample RSD values (red) and the quotients of RSD_{Sample} and RSD_{QC} (blue).

Display of Annotation Quality

Beyond the reporting of raw and processed data, the quality of metabolite annotation is also considered and presented in a concise but comprehensive manner with a five-tiered Annotation Quality (AQ) score: m/z, RT, isotope pattern, MS/MS and CCS. Users define which deviations are acceptable or excellent and the AQ symbol comprises that for each annotation.

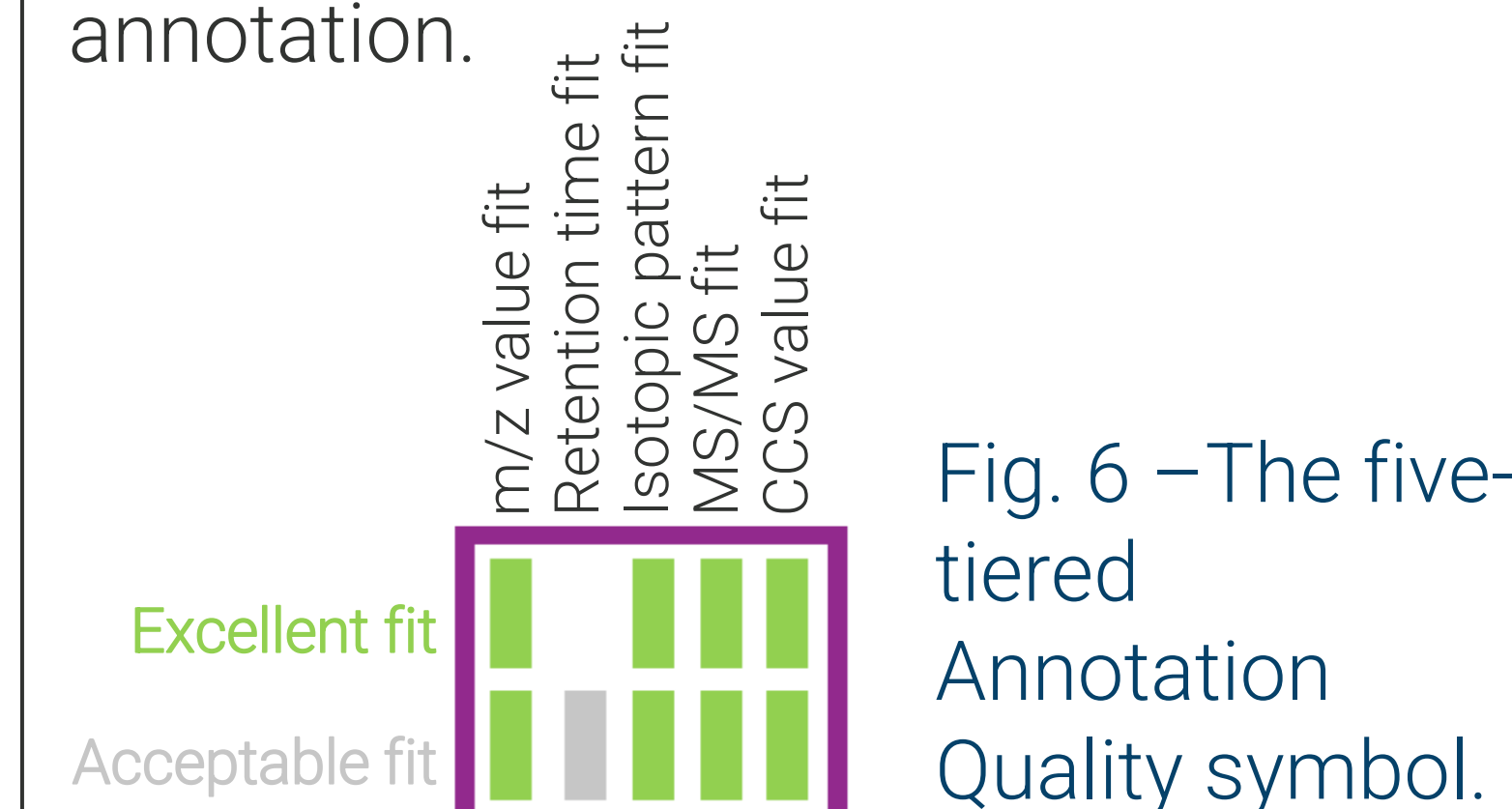


Fig. 6 – The five-tiered Annotation Quality symbol.

Custom Solutions via REST API

MetaboScape provides programmatic access to Feature Tables via a REST API. That allows users to write scripts in Python, R, Java, and many other languages to – for example – retrieve the feature matrix, create annotations, or assign flags to features.

References

- [1] Berger *et al.* *Metabolomics*. 2019 DOI: 10.1007/s11306-018-1460-7
- [2] Evans *et al.* *Metabolomics*. 2020 DOI: 10.1007/s11306-020-01728-5
- [3] Lewis *et al.* *Anal. Chem.* 2016 DOI: 10.1021/acs.analchem.6b01481
- [4] Gika *et al.* *J Chromatogr B Analyt Technol Biomed Life Sci.* 2016 DOI: 10.1016/j.jchromb.2015.10.045

Conclusion

- A thorough experimental design including pooled quality control samples aids in quality assessment.
- MetaboScape allows the use of QCs for quality assessment, filtering, and within-batch correction.
- Interactive visualizations help in quality review. QC outliers are highlighted, to reveal batch disturbances.
- Annotation quality is reported in a comprehensive but concise manner.

MetaboScape