

Proteomics Quality Control: Quality Control Software for MaxQuant Results

Chris Bielow,^{*,†,‡} Guido Mastrobuoni,[†] and Stefan Kempa^{*,†,‡}

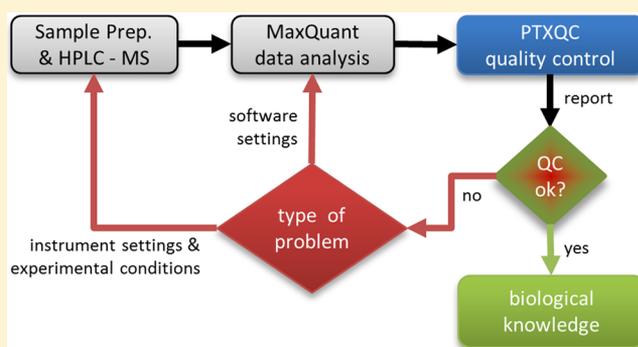
[†]Max-Delbrück-Centrum for Molecular Medicine Berlin, Robert-Rössle-Straße 10, 13125 Berlin, Germany

[‡]Berlin Institute of Health, Kapelle-Ufer 2, 10117 Berlin, Germany

S Supporting Information

ABSTRACT: Mass spectrometry-based proteomics coupled to liquid chromatography has matured into an automatized, high-throughput technology, producing data on the scale of multiple gigabytes per instrument per day. Consequently, an automated quality control (QC) and quality analysis (QA) capable of detecting measurement bias, verifying consistency, and avoiding propagation of error is paramount for instrument operators and scientists in charge of downstream analysis. We have developed an R-based QC pipeline called Proteomics Quality Control (PTXQC) for bottom-up LC–MS data generated by the MaxQuant¹ software pipeline. PTXQC creates a QC report containing a comprehensive and powerful set of QC metrics, augmented with automated scoring functions. The automated scores are collated to create an overview heatmap at the beginning of the report, giving valuable guidance also to nonspecialists. Our software supports a wide range of experimental designs, including stable isotope labeling by amino acids in cell culture (SILAC), tandem mass tags (TMT), and label-free data. Furthermore, we introduce new metrics to score MaxQuant's Match-between-runs (MBR) functionality by which peptide identifications can be transferred across Raw files based on accurate retention time and *m/z*. Last but not least, PTXQC is easy to install and use and represents the first QC software capable of processing MaxQuant result tables. PTXQC is freely available at <https://github.com/cbielow/PTXQC>.

KEYWORDS: PTXQC, quality control, MaxQuant



INTRODUCTION

The importance of quality control (QC) and quality assessment (QA) has been long acknowledged. Data quality is a cornerstone of solid research, demanding repeatability and reproducibility.² Ideally, small deviations in performance are observable, and their origin can be tracked down.

Adverse effects of missing QC can be found in early proteomics research,³ which could have been prevented if proper QC was in place.⁴

In 2009, the Amsterdam Principles⁵ called for the development of universally applicable quality metrics to ensure that only high-quality data is used in publications and released to public repositories. In 2012, a corollary was published, detailing potential metrics.⁶ Additionally, the National Institute of Standards and Technology (NIST) proposed a set of 46 QC metrics in 2010.⁷

Since then, many QC packages have been developed. NIST provide their own pipeline called MSQC^{7,8} for Microsoft Windows, which can read Thermo and Agilent TOF data and uses the Open Mass Spectrometry Search Algorithm (OMSSA)⁹ or SpectraST¹⁰ as the identification engine. Unfortunately, the output of MSQC is text-based, i.e., no visualization is provided, and development of OMSSA has been discontinued. A similar approach, extending the NIST metrics,

is pursued in QuaMeter,¹¹ relying on pepXML or mzIdentML file formats, which are not supported by all software packages. Another tool named Metriculator¹² uses the NIST pipeline as the backend and provides plots and tracking of samples via a web interface. Recently, QC workflows were introduced for OpenMS/KNIME, along with an XML-based file exchange format named qcML.¹³ SIMPATIQCO¹⁴ is being actively developed and extracts QC metrics like injection times, peptide spectral matches over retention time (RT), protein coverage, etc. directly from Thermo Raw or Agilent Wiff files, running on a dedicated server. Identification requires a local Mascot server or MS Amanda.¹⁵ SIMPATIQCO also offers a qcML export. Amidan et al.¹⁶ reported a machine learning approach to automatically classify standard QC samples that determines overall instrument performance. The classifier should be trained on a per-laboratory basis to account for high lab-to-lab variability. Another tool, which collects and plots QC metrics like chromatograms, source current, and injections times

Special Issue: Large-Scale Computational Mass Spectrometry and Multi-Omics

Received: August 21, 2015

Published: December 14, 2015

Table 1. Overview of Quality Metrics and Plots

file source (.txt)	abbreviation in plots	data source	heatmap score basis	scoring function ^a
Parameters	PAR	•General parameters settings (MaxQuant version, modifications, ppm tol., FASTA database, FDR cutoff, etc.)	NA	
Summary	SM	•MS ² identification rate	(1) Distance to “great” threshold	•LinRef
ProteinGroups	PG	•Protein Intensity (MS ¹ , iTRAQ reporter, [LFQ]) •Fraction of contaminants •User-defined contaminants •{SILAC only} Ratio distributions •PCA plot	NA (not suitable for Raw file based heatmap, since an experimental group might correspond to more than one Raw file)	
Evidence	EVD	•Peptide Intensity •Number of protein and peptides per condition (w and w/o matched) •MBR RT alignment •MBR RT matching •Charge distributions •IDs over RT •MS ¹ decalibration •MS ¹ recalibration error •Contaminants •RT peak width distribution •Twin sequence fraction (oversampling estimation)	(2) Intensity threshold (3) Count threshold (4) Interfile pair distance (5) Intrafile group distance (6) Deviation from prototype (7) Equal counts per RT bin (8) Proximity to max. tol. (9) Centeredness around 0 (10) Summed intensity (11) Deviation from prototype (12) % of single MS/MS per Peak	•LinRef •LinRef •AlignDist ^c •MatchDist •MedianDist ^b •Uniform •CenteredRef ^c •GaussDev •LinRef •BestKS ^b •LinRef
Msms	MSMS	•Missed cleavage •Missed cleavages variance •MS ² fragment mass error	(13) Fraction of MC > 0 (14) Deviation from prototype (15) Centeredness around 0	•Percent •MedianDist ^b •Centered
MsmsScans	MSMSscans	•TopN over RT •TopN •% identified by TopN •Ion Injection time	(16) Equal saturation over RT (17) Reaching highest N consistently (18) Equal ID rate for all N (19) Fraction of scans > time threshold	•Uniform •MaxN •Uniform •Percent

^aSee Supporting Information for details. ^bThe quality function computes scores per Raw file using other Raw files as reference. All other functions will return an absolute score, which depends only on the Raw file itself. ^cThe quality function relies on parameter settings in MaxQuant, which must be matched in PTXQC. If the mqpar.xml file is present, these settings are extracted automatically.

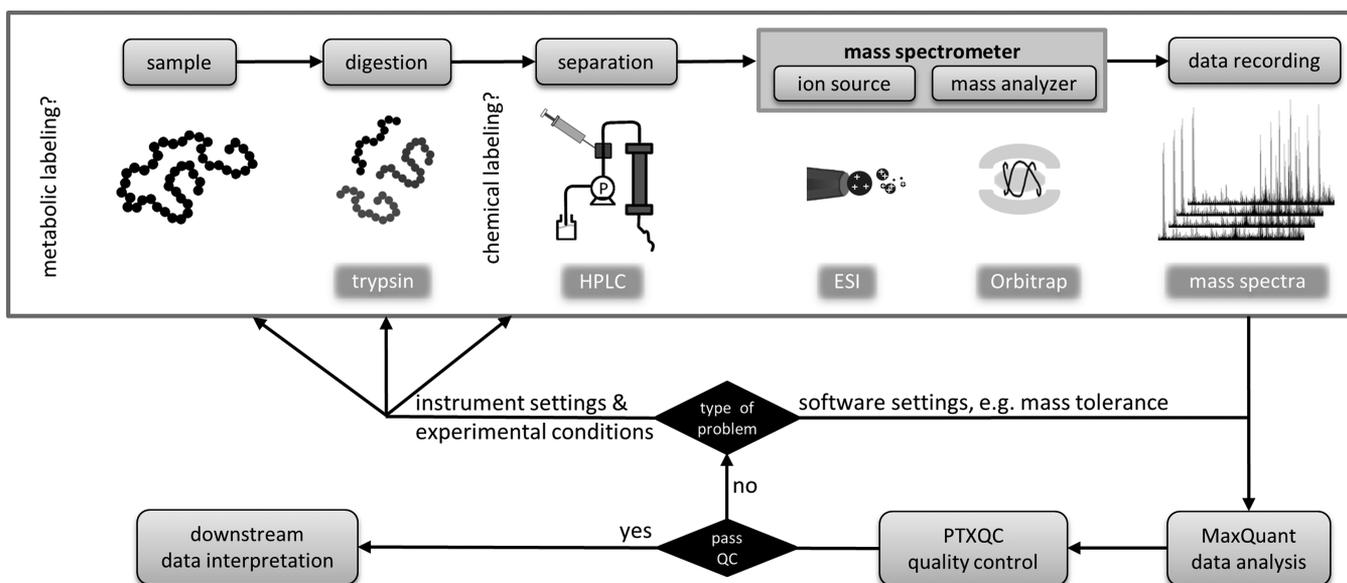


Figure 1. Experimental and software workflow for bottom-up shotgun proteomics experiments. First, the protein sample is digested, typically using trypsin, to yield peptides. Subsequently, the sample is subjected to HPLC, separating the peptides by their physicochemical properties. The eluent is then ionized using electrospray ionization, and the mass/charge ratio of the peptides is measured. The quality of the resulting spectra is influenced by all preceding steps. Spectra are then submitted to MaxQuant for analysis. The resulting output is assessed by PTXQC, and upon passing the quality criteria, it is cleared for downstream analysis. If quality is not satisfactory, then either remeasurement is required or (preferably) MaxQuant parameters are adapted to remove the bias detected by PTXQC.

directly from single Thermo Raw files, is RawMeat (<http://vastsci.com/rawmeat/>). However, since RawMeat performs no

actual data processing (e.g., peptide identification), only limited conclusions can be drawn. Extensive data visualization

capabilities including QC plots for MS/MS-based proteomics data are offered by various software packages written in R¹⁷ (see ref 18 for an exhaustive review).

To our knowledge, there is no published QC software capable of processing MaxQuant¹ results. However, MaxQuant has a large user base within the proteomics community and would thus benefit greatly from QC software to ensure unbiased downstream analysis.

In principle, raw data could be checked using any QC software as described above, even before running MaxQuant. Unfortunately, some of the above tools are not actively maintained, offer only a command line interface, or produce only text-based results. Additionally, some QC tools (e.g., MSQC) are meant to compare performance only among dedicated QC samples, i.e., they cannot be used for samples of biological interest. However, the major drawback of using such an external QC is the lacking guarantee that MaxQuant will deliver the same performance. This might simply be due to deviating parameter settings (e.g., MS² search tolerance), internal algorithm specifics (e.g., possibility for calibration of RT and *m/z*, Andromeda¹⁹ search engine, false discovery rate (FDR) model), or the support for special experimental designs (e.g., phosphorylation enrichment). In addition, MaxQuant features algorithms, such as mass recalibration and second peptide search, which might enable data recovery to an extent that is not possible with other tools. Thus, it is paramount to perform QC checks on the results of MaxQuant and report a set of comprehensive metrics.

We developed Proteomics Quality Control (PTXQC), which is capable of reading MaxQuant output and generating a comprehensive report using a wide range of QC metrics. In total, PTXQC reports up to 24 quality metrics (Table 1), of which 19 can be automatically scored using dedicated scoring functions. The scores are collated to create an overview chart (heatmap) that displays up to 19 scores per Raw file for a compressed overview of the whole experiment. The user can subsequently follow up on detailed quality metric plots of interest in the remainder of the report.

METHODS

A typical shotgun LC–MS experiment in bottom-up proteomics encompasses the following main steps: digestion, separation by high-performance liquid chromatography, and subsequent acquisition of mass spectra (MS) and tandem mass spectra (MS²) data by a mass spectrometer (Figure 1). Subsequently, a processing suite for proteomics data (here, MaxQuant) is used to identify and quantify proteins, typically comparing different biological conditions. Intermediate results (e.g., peptide–spectrum matches) are commonly available as well. Subsequently, it is highly recommended to carry out a QC assessment (e.g., using PTXQC). If quality is satisfactory, then the data is cleared for downstream analysis. Upon rejection, previous steps of the pipeline require optimization. Depending on the severity of the detected problem, a change in MaxQuant software parameters (e.g., calibration tolerance or alignment window) might suffice to pass the QC. Ultimately, QC failure can trigger a complete remeasurement of the samples (e.g., upon unsatisfactory protein digestion).

A good approximation for overall performance of the pipeline is the number of quantified proteins per sample. However, this reflects only the average performance of the pipeline as a whole. If one could benchmark individual stages, then not only QC but also optimization becomes possible.

Thus, QC tools that report metrics on individual steps of the shotgun proteomics pipeline can be used to identify poorly performing parts and identify targets for optimization (Figure 1).

PTXQC makes use of two distinct but related concepts, namely, quality metrics and quality scores. Quality metrics (such as digestion performance) are shown in the report using different kinds of plots, usually detailing the performance of multiple Raw files concurrently. On the basis of the data underlying these metrics, PTXQC computes a quality score using a scoring function (see below). The scores represent the basis for the overview heatmap, which is presented at the beginning of the report. In summary, quality metrics offer a visual guide to user to judge quality, whereas scores computed from the underlying data represent a mathematically more rigid way to automatically flag data sets as failed or successful.

To conveniently visualize the metrics, PTXQC makes use of different types of plots, multiplotting, and color schemes. If thresholds are known (e.g., MS² fragment ion search tolerance), then they are added for visual guidance.

QC Metrics

PTXQC's metrics can be assigned to four categories, corresponding to steps in the experimental workflow (sample preparation, LC, MS, and general performance). An abbreviation of the data source (Table 1) is provided with every plot, allowing the user to trace the origin of the information, e.g., PG indicates the MaxQuant's protein groups table, EVD points to the evidence table, etc.

In the following paragraphs, we introduce three novel and powerful QC metrics exclusively found in PTXQC, namely, custom contaminants and metrics for RT alignment and transfer of spectrum identifications across Raw files. Details on common metrics like digestion efficiency, charge distribution, and ion injection times can be found in the Supporting Information.

Customizable Contaminant Search

While MaxQuant supports customizable contaminant lists, it is sometimes not desirable to modify this file, especially when multiple operators utilize the same MaxQuant installation. On the other hand, flagging a protein posthoc as contaminant is possible only by manually editing the MaxQuant output. Thus, PTXQC offers configurable lists of custom protein contaminants, supplied as a regular expression applied on the protein name or description. If a larger set of proteins with nonoverlapping names is sought, then the user can employ custom FASTA files amended with protein name tags during the MaxQuant analysis or provide a more complex regular expression. The latter allows the PTXQC analysis to be run without rerunning MaxQuant. We compute two abundance measures for contaminants from the evidence table: one based on intensity and the other based on spectral counts. PTXQC reports the sum-of-intensity/proportion of peptides matching the regular expression compared to all peptides. Non-unique peptides and hits to the reverse database are discarded in advance.

Retention Time Alignment and ID-Transfer

MaxQuant's match-between-runs (briefly described in refs 20 and 21) will align the retention times across Raw files using 3D peaks with identical peptide IDs as landmarks. The alignment function is nonlinear and can correct retention time differences up to a certain extent (by default, 20 min). In a second step,

MaxQuant will transfer MS² identifications across Raw files using corrected retention times and an accurate mass, thus assigning a peptide ID to hitherto unlabeled 3D peaks. For samples where MS² coverage was not sufficient to identify all peptides, MBR can significantly increase the number of annotated 3D peaks, therefore providing more data for downstream quantification of proteins. The MaxQuant developers recommend using MBR only on samples with comparable LC gradient conditions.

In the remainder of this article, we will refer to peptide identifications as genuine if the identification was obtained from an MS² spectrum and passes the FDR filtering, whereas we call an identification transferred if the corresponding 3D peak is annotated via MBR. Additionally, a peptide sequence implicitly includes modifications (e.g., carbamidomethyl), i.e., two identical sequences are regarded as unequal if they have different modifications.

Retention Time Alignment. MaxQuant's RT alignment function can be reconstructed from the evidence table. Note that MaxQuant's retention time correction is reported relative to the first Raw file, even though files were aligned using a guide tree for the alignment.²⁰ Thus, reported shifts may exceed the given 20 min tolerance since RT shifts can accumulate when walking the alignment tree. We found the shape of the alignment function not to be feasible for assessing the success of the alignment quantitatively. Instead, we introduce two new metrics: the first metric is aimed at alignment quality (and is thus an inter-Raw file metric) and the second is aimed at the actual transfer of identifications between Raw files (constructed as an intra-Raw file metric; see the [ID Transfer](#) section below).

In order to estimate the alignment quality, PTXQC compares the residual retention time difference of two RT-aligned 3D peaks with identical identifications across Raw files (i.e., using corrected retention times). For example, after alignment, a peptide with sequence DFINGAR with charge state 2, genuinely identified both in files A and B, should have a very similar corrected RT in both files. Such pairs of peptides, genuinely identified in both the reference Raw file and another Raw file, with identical sequence and charge state, are called ID-pairs. For each Raw file, we compute the RT difference of every ID-pair using the calibrated retention time. Ideally, most differences are within MaxQuant's matching tolerance (see [ID Transfer](#) section below). The reasoning is as follows: only if ID-pairs, i.e., landmarks, are aligned well can we expect the subsequent ID transfer to be successful. If ID-pairs are not well-aligned for a certain stretch in RT, then every ID-transfer within this stretch will be a random hit and thus a false positive. PTXQC plots results for each Raw file ([Figure 4](#)) and reports the alignment score "EVD: MBR Align" in the heatmap as the percentage of ID-pairs that are within the matching tolerance.

The alignment metric also estimates the maximally required RT alignment window (in rare cases, more than 20 min is needed), allowing the user to make a data-based decision on how to change MaxQuant parameters to obtain a better alignment.

For experimental designs using a prefractionation strategy, PTXQC picks one reference file per fraction and compares it to all proximal Raw files (i.e., the immediate fraction neighbors). This is required since the overlap between distant fractions will usually be small or empty and MaxQuant uses only proximal fractions for transferring IDs. See [Figure S1](#) for an example.

ID Transfer. After retention times have been calibrated using genuine MS² identifications, MaxQuant transfers peptide

IDs from any Raw file to any other Raw file (if MaxQuant's "match-from-and-to" setting was unchanged). Further restrictions apply for fractionated samples (see above). An unidentified 3D peak (target) receives an annotation if a genuinely identified counterpart from another Raw file (source) has a similar calibrated retention time (0.7 to 2 min deviation by default, depending on the MaxQuant version) and if its m/z matches the theoretical m/z of the peptide to be transferred (within 4.5–7 ppm by default, depending on the MaxQuant version).

MaxQuant reports the RT difference between the source identification and its target in the "match-time-difference" column of the evidence table. However, small values do not indicate that this matching is correct, since any unannotated 3D peak with similar RT and m/z is a putative target candidate. Therefore, a robust alignment is paramount for the ID-transfer.

To gauge the correctness of the transfer step, the PTXQC metric compares all transferred identifications to the genuine identifications within each Raw file. If the transfer was correct, then no identification should occur more than once. In particular, a genuine ID (locally confirmed by MS²) and a transferred ID in the same Raw file indicate that the matching targeted the wrong 3D peak (generating a false positive) because the same peptide was already identified genuinely. Alternatively, the MaxQuant feature finding algorithm accidentally split a 3D peak into two separate entities, where only one was identified by MS². In this case, the genuine and transferred IDs will have similar corrected retention times.

PTXQC assigns every 3D peak into one of three classes: "Single", "Group – in width", and "Group – out width". The "Single" class covers all 3D peaks whose peptide sequence and charge state is unique for the Raw file at hand. The other two classes represent 3D peaks that are part of a group, i.e., that have siblings with identical sequence and charge in the same Raw file. Within each peak group, PTXQC uses the retention time deviations to decide if the group is valid ("in width") or invalid ("out width"). The threshold to decide if a peak group is "in width" is the median RT peak width of the respective Raw file. If the RT span of the group is larger than the typical RT peak, then the evidence is considered segmented and it is assigned to the out-width class, i.e., it is unlikely that the out-width group represents a split 3D peak but, rather, two (or more) entirely different 3D peaks.

Depending on which subset of peaks is used to assign the three classes, different conclusions can be drawn. Considering only genuine 3D peaks and assigning them to a class, we can determine the intrinsic segmentation of a Raw file. The proportion of out-width peaks is usually very small, since a peptide usually elutes only once from the LC column. If we consider only the subset of 3D peaks that were identified via MBR plus all genuine 3D peaks that have the same identification, then we can draw conclusions about the success of the ID-transfer. PTXQC reports the fraction of singlets plus in-width group as the quality score for ID-transfer (see yellow arrows inserted into [Figure 5](#)). The out-width fraction can rise considerably, depending on the success of the alignment, resulting in a lower score. Finally, if we consider all 3D peaks (irrespective of if they are genuine or transferred) and assign a class to them, then we obtain an overall view on the segmentation issue.

We emphasize that the plain number of transferred IDs per Raw file is a rather inaccurate indicator for correct ID transfers since (a) samples with high genuine MS² coverage and good

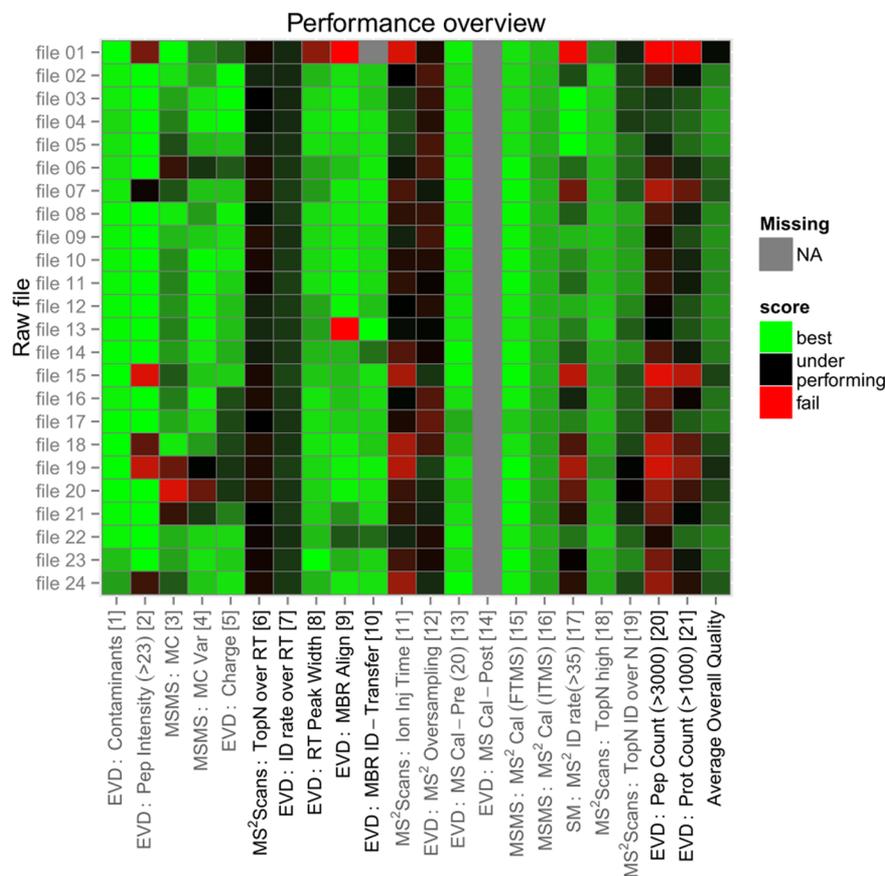


Figure 2. Heatmap overview of a TMT-labeled data set. Columns denote the metric; rows correspond to Raw files. The color gradient for each cell ranges from best (green), to underperforming (black), and, finally, fail (red). Column names are sorted and color-coded (gray or black, alternating) by the four main steps in the analytical workflow.

alignments to other samples are expected to yield few ID transfers and (b) high-complexity samples with low genuine MS² coverage and bad alignment to other samples are expected to yield many false positive ID transfers.

The MBR-FDR calculation mentioned in Geiger et al.²¹ is a valid alternative to our ID-transfer scoring function as described above. However, MBR-FDR values are not calculated by default and cannot be activated within the MaxQuant application. This feature needs to be manually enabled using the “matchBetweenRunsFdr” entry in the MaxQuant XML configuration file. Subsequently, the “match.q.value” column of the evidence table will contain *q*-values for matched evidence. However, the validity of this intrinsic MaxQuant metric critically depends on alignments of good quality.

If either alignment or matching is unsatisfactory, then MBR should be disabled partially or completely.

QC Scores

For 19 out of 24 quality metrics supported by PTXQC, we have devised a set of mathematical equations (Table 1 and Supporting Information) that allows one quality score to be computed per Raw file and metric. The remaining five metrics remain unscored since they are based on the protein groups table where a 1:1 relationship between the experimental groups and Raw files cannot be guaranteed. Each quality score ranges between zero and one. The exact mathematical formula is listed in the Supporting Information. A heatmap summarizes up to 19 quality metrics per Raw file, with quality scores represented by a color gradient. The columns (metrics) are ordered according

to the analytical flow (Figure 1). Each row represents one Raw file. Green tiles indicate good quality, red indicates failure, and black marks indicate intermediate performance.

The majority of scores (16 of 19) are reference-less, i.e., their value depends only on the particular Raw file at hand. Thus, all Raw files can potentially achieve very good performance, i.e., there is no relative scaling. The three remaining quality metrics are scored relative to the data available in the study: “Charge distribution”, “RT peak width over time”, and “Missed cleavages variance”. These metrics depend on the data at hand, and target values are hard to formalize. In particular, the charge distribution should be similar across all Raw files, whereas the exact share of doubly charged peptides is of secondary concern. The scoring function is therefore selecting the most representative Raw file and penalizes deviations from this reference. A similar argument applies for the variance of missed peptide cleavages. Recent research shows that missed cleavages do not negatively influence protein quantification if all samples share the same degree of digestion.²² The degree of digestion itself is additionally represented by a “Missed Cleavage” score. Finally, the RT peak width strongly depends on the LC setup. The quality score penalizes Raw files that deviate strongly in their RT peak width distribution compared to a representative Raw file of the same study.

RESULTS

In this section, we will provide an example of the overview heatmap and in-depth examples of the three metrics described in the Methods section. Please refer to the Supporting

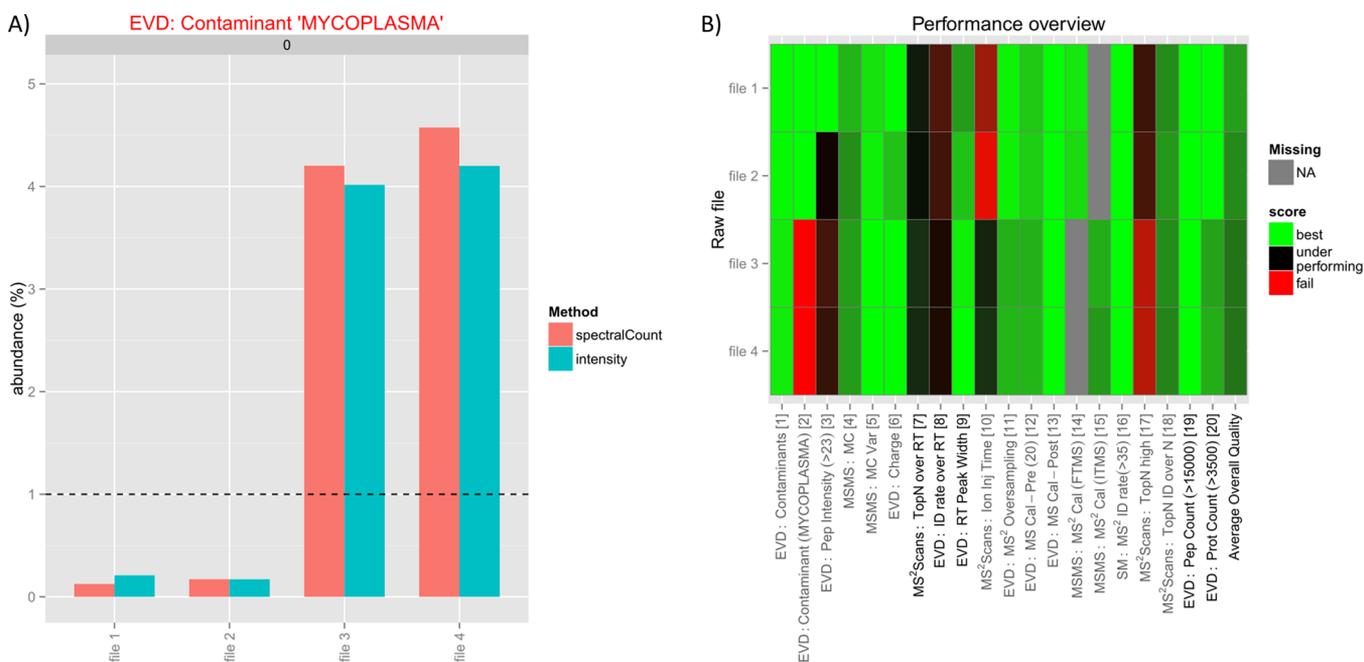


Figure 3. A custom database containing proteins from *Mycoplasma hyorhinis* was included during the MaxQuant analysis of an in-house human QC data set. PTXQC was configured to search for mycoplasma proteins. (A) Summary of the relative abundance (red) and spectral counts (blue) of proteins (or protein descriptions) containing the string “MYCOPLASMA”. The first two Raw files (file 1, file 2) serve as negative controls, in addition to two Raw files with known contamination (file 3, file 4), as confirmed by both intensity and spectral counting. The default threshold of 1% is plotted by PTXQC as a horizontal dashed line for visual guidance. Exceeding the threshold will report the respective Raw file as failed in the overview heatmap. (B) Corresponding heatmap summarizing the whole study. The second column shows the scores for the mycoplasma query. This column is present only if a custom contaminant query is requested via the PTXQC configuration file.

Information for a detailed description of the remaining metrics, example figures, [full QC reports](#), and a description of the data sets.

Overview Heatmap

We use a prefractionated, TMT-labeled data set²³ consisting of 24 Raw files as the basis for the heatmap shown in [Figure 2](#). The MaxQuant result folder was obtained from the Pride Archive,²⁴ ID PXD000427. Expected protein and peptide counts per Raw file were adapted from 3500 and 15 000 to 1000 and 3000, respectively, via PTXQC’s YAML configuration file (see below) to account for the reduced protein content due to prefractionation.

Sample preparation quality is shown in the first five columns of the heatmap: The first column, “EVD: Contaminants”, represents common laboratory contaminants (e.g., keratins), as annotated by MaxQuant. For the TMT-labeled data set, most samples contain low amounts of contaminants; only the last two fractions show a minor increase. Peptide intensity (column 2) is as expected, except for fractions 1, 15, 18, and 19. Fractions with low overall intensity show a poor ion injection time (column 11), MS² identification rate (column 17), and number of identified peptides (column 20). Digestion was very thorough with few missed cleavages (“MSMS: MC”, column 3), except for fractions 6 and 19–21. Not surprisingly, the same files are also negatively indicated in the MC variation column since they deviate from the majority of files with good digestion (column 4).

LC performance is shown in columns 6 to 10. Along the LC gradient, peptides do not seem to elute uniformly over time (“MSMS: TopN over RT”, column 6), also affecting the number of successful identifications over time (“EVD: ID rate over RT”, column 7). The first fraction shows unusual RT peak

width (column 8). The alignment step of Match-between-runs has failed for fractions 1 and 13 (column 9); fraction 1 simply shares no landmarks with its immediate neighbors (fraction 2); thus, MaxQuant could not align them. Fraction 13 should align very well, but it was unintentionally labeled as fraction 3 in the MaxQuant configuration. Hence, MaxQuant (and PTXQC) cannot find any landmarks for alignment. However, PTXQC’s ID transfer metric (column 10) suggests that fraction 13 behaved very well. The reason is simply that all IDs transferred to fraction 13 (from fraction 2 and 4) have sequences that are not expected at such a late fractionation stage. Hence, transferred IDs are singlets, not conflicting with genuine MS² IDs from fraction 13. Fraction 1, on the other hand, shows an NA score since it received no transferred IDs.

Instrument performance is reflected in columns 11–19: MS¹ calibration was very good on the instrument-level already (“EVD: MS Cal-Pre”, column 13, with 20 ppm tolerance). MaxQuant’s internal mass recalibration cannot be scored (“EVD: MS Cal-Post” is NA, column 14) since mass deltas for chemically modified peptides (such as TMT and iTRAQ) are reported incorrectly by MaxQuant (see [Supporting Information](#) for details). The instrument mostly reached its TopN limit (“MSMS: TopN high”, column 18).

General parameters, reflecting overall performance, are shown last: Not surprisingly, the overall protein and peptide counts per Raw file vary widely (columns 20 and 21), with the richest fraction containing at most 900 proteins.

In summary, a few fractions show extremely low peptide abundance, which causes dependent metrics such as ion injection time and MS/MS ID rate to underperform. MBR across neighboring fractions has worked very reliably and should remain enabled, on average contributing 36% increased

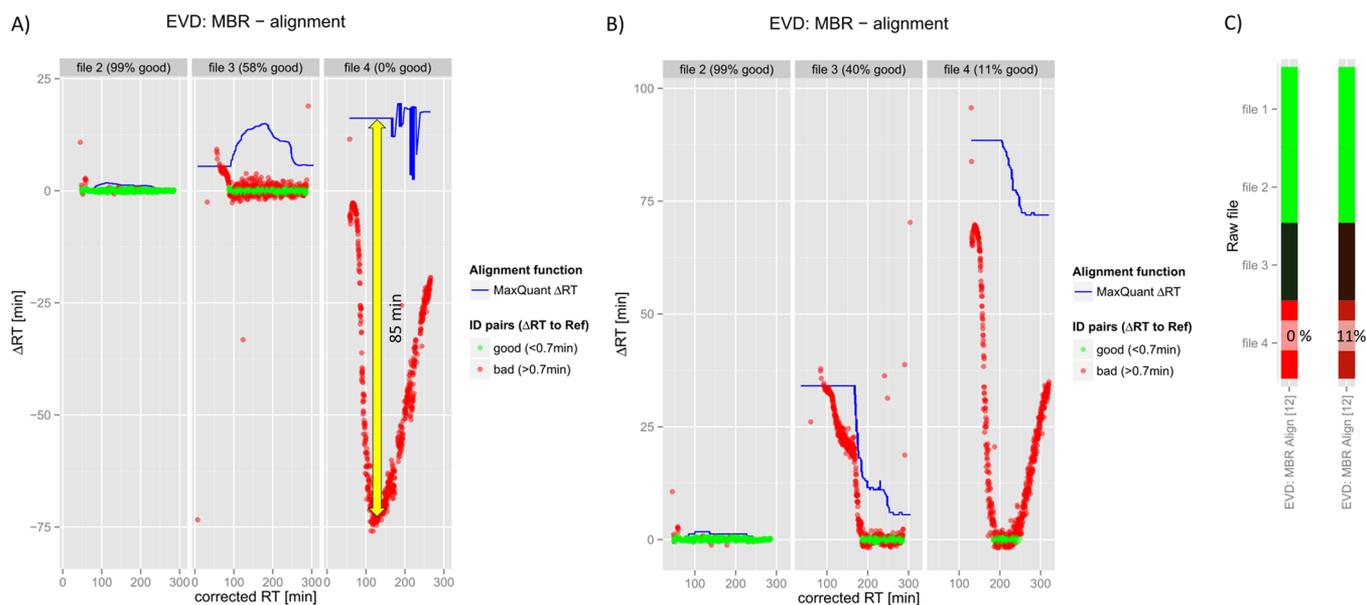


Figure 4. Retention time correction using Match-between-runs. Alignment performance is judged using the residual RT difference (Δ RT) of identical genuine 3D peak pairs after alignment with respect to a reference file (file 1). Each ID-pair is represented by a dot: green dots indicate that the underlying 3D peaks are successfully aligned, with a residual RT difference of less than 0.7 min. Red dots indicate that the alignment was unable to bring the 3D peaks close enough in RT (>0.7 min). The RT correction function of MaxQuant is shown in blue. The fraction of good pairs is given in the panel title, e.g., 99% of the pairs between the reference (file 1) and file 2 are successfully aligned. (A) Four Raw files of human QC samples with varying degrees of alignment success (decreasing). MaxQuant's RT alignment tolerance window was set to the default of 20 min. The horizontal yellow arrow indicates the required RT alignment tolerance (~ 85 min). (B) The same files as in (A) but with a larger RT alignment tolerance of 100 min. Note the increased fraction of good ID-pairs for file 4 (11%) due to a small region between 200 and 250 min that was now successfully aligned. (C) Side-by-side representation of the MBR alignment scores for the analyses in A (left column) and B (right column) as shown in the heatmap. The actual heatmap has many more columns; we show only the column of interest, "EVD: MBR Align". File 3 shows a trend toward being colored red (due to the score decreasing from 58 to 40%); file 4 shows a slight improvement (from 0 to 11%).

ID counts per fraction. If resources permit, then MaxQuant should be rerun with a corrected fraction assignment for fraction 13, which received 13 wrong peptide assignments (from 34 PSMs) in addition to the genuinely identified 1438 PSMs. Conversely, real fractions 2, 3, and 4 also received false positive identifications from fraction 13 (since it was labeled as fraction 3). For subsequent studies of similar sample complexity, we would recommend combining low-abundance fractions with neighboring fractions prior to LC-MS measurement, reducing the number of sample injections, and avoiding most problems mentioned here.

Custom Contaminant Detection (Mycoplasma)

To demonstrate the flexibility of PTXQC's custom contaminant approach, we searched for *Mycoplasma hyorhinitis* in a study using HEK293 cell lines. Mycoplasma contamination should be avoided at all costs since infection can alter cell metabolism and physiology. Infection of tissue culture cell lines was first described over 50 years ago, and to this day, it remains a persistent problem since it is subtle and hard to detect unless specific measures are taken. Sources of contamination range from animal-derived media products and laboratory personnel to cross-contamination, with estimated cell culture contamination rates up to 35%.²⁵ LC-MS data can serve as a basis for a confirmatory experiment. Creating a suitable protein database is straightforward. Since mycoplasma contamination can usually be attributed to a few mycoplasma strains (here, *M. hyorhinitis*), the choice of strains is paramount for successful detection. We advise against using a mycoplasma database containing all strains. Instead, the search should be restricted to likely candidate strains. Adding a full-blown database will unnecessa-

rily increase the peptide search space and most likely reduce the number of successfully identified peptide spectra at a fixed FDR.²⁶ For example, the UniRef90²⁷ database contains a remarkably high number (27 535) of mycoplasma protein clusters.

Figure 3 shows a plot with results from a mycoplasma query. For this analysis, we included an unmodified *M. hyorhinitis* FASTA database during the MaxQuant run and instructed PTXQC to search for protein hits containing the string "mycoplasma". Two samples can be clearly identified as being contaminated by *M. hyorhinitis*, contributing almost 5% of the total sample content. These files should be excluded from downstream analysis. Furthermore, the source of contamination needs to be tracked down and eliminated.

Retention Time Alignment

MaxQuant's Match-between-runs represents a valuable mechanism to boost the protein coverage and increase the number of quantifiable proteins. However, it should be used only under comparable column conditions for all samples involved. However, the exact degree of comparability is hard to quantify by manual analysis. Using a set of four files from an in-house HEK293 QC study, we demonstrate the sensitivity of our alignment and ID-transfer metrics. Files 1 and 2 were measured on the same day, file 3, the following day, and file 4, under different column conditions a few months earlier.

Figure 4 shows the alignment plot and the corresponding scores for the RT alignment. The RT calibration function reported by MaxQuant is normalized with respect to file 1. File 2 aligns perfectly, whereas file 3 can be only partially aligned. File 4 used a different column, resulting in a failed alignment.

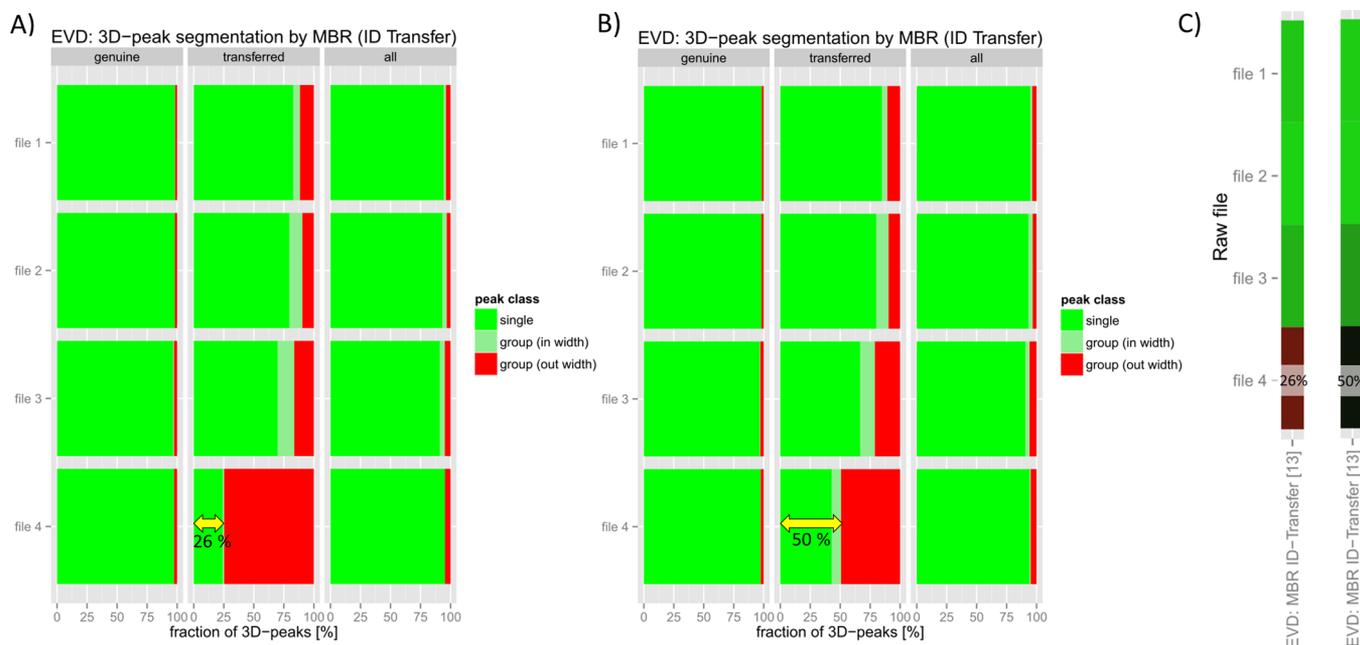


Figure 5. ID-transfer performance of Match-between-runs. Per Raw file (rows), three different aspects of evidence are shown (columns): “genuine” uses only 3D peaks that have genuine MS² identifications, “transferred” ignores 3D peak groups that are purely genuine, and “all” considers all evidence (genuine + transferred). Each stacked bar contains three peak classes, together summing to 100% of peaks: single, group (in width), and group (out width). (A) Four Raw files of human QC samples. Files 1 and 2 were measured on the same day, file 3, the following day, and file 4, under different column conditions (aging) a few months earlier. MaxQuant’s RT alignment tolerance was set to the default of 20 min. Most IDs transferred to file 4 are false positives (large red bar in the “transferred” column). The overall effect is not drastic (“all” column) since most IDs in file 4 are genuine and only few IDs were transferred to file 4. (B) The same files as in (A) but with a larger RT alignment tolerance of 100 min. Note the decreased contribution of the “group (out-width)” for file 4, indicating fewer false positive matches. (C) Side-by-side representation of the MBR ID-transfer scores for the analyses in A (left column) and B (right column) as shown in the heatmap. The actual heatmap has many more columns; we show only the column of interest, “EVD: MBR ID-Transfer”. The first three files show almost no change, whereas file 4 shows an improvement (dark red to black).

All ID-pairs between files 1 and 4 show a large residual delta RT (Δ RT) after alignment, reaching up to 75 min, which is much larger than MaxQuant’s target value of <0.7 min (the default for MaxQuant v1.5, which was used here). However, MaxQuant, by default, searches only within a 20 min RT tolerance window for an optimal alignment. The required width of the RT tolerance window can be gauged visually using the maximum vertical distance between the ID-pairs (red/green) and the alignment function (blue). Here, we can estimate the maximal RT difference between the two columns to be around 85 min (see yellow arrow, Figure 4a), indeed suggesting very different column conditions considering the total gradient length of 300 min.

In summary, the data indicate that an optimization of MaxQuant parameters could rescue the failed alignment. This is a promising solution since it avoids a costly remeasurement of data on one hand and loss of information by disabling the failed MBR on the other hand. One minor drawback is a potential increase in MaxQuant processing time due to the larger search space for alignment. However, for this data set, we did not observe a longer runtime when increasing the RT tolerance window settings from 20 to 100 min. Indeed, the new setting yields a partially successful alignment (Figure 4b). However, overlap of only 11% of ID-pairs in file 4 still renders this alignment unacceptable. Performance of file 3 even decreases from 58 to 40%. Figure 4c shows alignment scores extracted from the heatmap plot corresponding to Figure 4a,b. In conclusion, only files 1 and 2 should be aligned. Files 3 and 4

do not align well to any other file and should be excluded from MBR.

Larger data sets with distinct subsets of column conditions usually benefit from introducing an artificial fraction assignment during the configuration of MaxQuant. By assigning non-neighboring fractions to each subset (e.g., fraction 1 for all samples with LC-column A; fraction 5 for all samples with LC-column B, etc.), MaxQuant will apply MBR only within the groups of similar column conditions. On one hand, samples in the same group align well and benefit mutually from transferred IDs. On the other hand, false positive ID-transfers due to failed alignments are avoided.

ID Transfer

Our second metric visualizes the performance of the second step during MaxQuant’s Match-between-runs, i.e., the transfer of identifications across Raw files after retention times have been aligned. Failed alignments usually lead to (a) few identifications being transferred and (b) a large proportion of false positive annotations. In order to quantify the ID-transfer performance, we inspect the increase in segmentation by Match-between-runs (i.e., increase of identically annotated 3D peaks with widely different retention times within one Raw file).

Figure 5 shows the ID-transfer corresponding to the analysis conducted above for the RT alignment (Figure 4). Figure 5a clearly shows false positive identification transfers to file 4 (red bar), affecting $100\% - 26\% = 74\%$ of all transferred IDs. The overall impact (“all” column, row 4) is not severe: file 4 received only 507 transferred IDs (vs 26 298 genuine IDs

locally identified). Consequently, 507 peptide quantifications (~2%) are certainly based on wrong identifications since we already know that alignment completely failed for file 4 (compare to Figure 4a). Thus, 74% for out-width peaks is an underestimation and, conversely, singlets are overestimated (i.e., false positive ID transfers that are unique to the target Raw file and cannot be detected as false positives since they have no genuine counterpart to create an out-width group).

Using a partially successful alignment (100 min RT tolerance; compare to Figure 4b), we also find slightly improved ID-transfer performance (compare Figure 5, panel A vs B, increase from 26 to 50%).

In conclusion, files 1 and 2 should be included during MBR (requiring a change in MaxQuant's match-from-and-to settings and a subsequent rerun of MaxQuant) since they show sufficient gain in IDs and almost no segmentation. After the restriction of MBR to files 1 and 2, the (already good) ID-transfer scores for files 1 and 2 increase further (from 90 to 97% for both files; data not shown) since false positive transfers from files 3 and 4 are avoided.

In most data sets that we examined (data not shown), we find MBR to increase the segmentation, i.e., create out-width groups. The severity, however, depends highly on the alignment quality. This is not surprising since a good alignment will prevent false ID transfers. In general, both the alignment score and ID-transfer score should be >95% for each Raw file participating in MBR.

Report Configuration

PTXQC is capable of extracting parameters from the MaxQuant configuration file (mqpar.xml) automatically, thus reducing the user's configuration effort to a minimum. Other parameters, e.g., individual target thresholds for the number of identified proteins, can be configured via a configuration file in YAML format [<http://www.yaml.org>]; see Table 2 for an example. The default configuration has sensible defaults for high-complexity samples acquired on an LTQ-Velos Orbitrap and a long nano-LC gradient of 4 h. The user is free to specify

Table 2. Shortened PTXQC Configuration File in YAML Format^a

```
### configuration file for PTXQC
PTXQC:
  UseLocalMQPar: yes
  ReportFilename:
    extended: yes
File:
  Evidence:
    enabled: yes
  SpecialContaminants:
    cont_MYCO:
      - MYCOPLASMA
      - '1'
  ProteinGroups:
    enabled: yes
  RatioPlot:
    LabelIncThresh_num: 4.0
```

^aTo disable all plots based on proteinGroups.txt, the parameter "File → ProteinGroups → enabled" should be changed from yes to no. A detailed manual of parameters and their values is provided with the PTXQC package.

new default settings for different setups (e.g., fractionated samples, long/short LC gradients) and apply them depending on the data set at hand.

Additionally, the configuration file allows only a subset of metrics to be enabled, permitting the evaluation of incomplete MaxQuant result folders or reducing the size of reports. Input file names are automatically shortened or renamed to allow the figure axis annotation within plots to be compact. If desired, the user can modify the name mapping and assign new file names globally.

DISCUSSION

We have introduced PTXQC, a tool that greatly facilitates and automates QC checks of proteomics data. The QC tool was developed to compare samples from the same batch but also from different batches to allow for comparisons of multiple parameters in an easy and structured way. This ultimately became necessary when working with large sample batches regardless if SILAC was used or label-free quantitation was applied. Differences in the sample input, digestion efficiency, or machine performance ultimately influence identification and quantification of peptides. Thus, besides the (desired) biological changes in protein abundance, technical limitations can introduce significant bias into the data. If no QC is applied, then it is hard to find the origin of the variance within the data. At the same time, a positive QC increases the confidence of the experimental results and marks an important step before data publication.

We have introduced two new metrics (RT alignment and ID-transfer) to judge the Match-between-runs functionality of MaxQuant. Also, visualization and scoring of contaminations, as demonstrated on the example of mycoplasma, have proven to be useful in our day-to-day routine. PTXQC has additional convenience functionality (e.g., detecting mass calibration issues), which is described in the Supporting Information.

We believe PTXQC is useful to a wide audience and can drastically shorten the number of data evaluation/remeasurement iterations since quality can be checked directly without scripting/programming experience. The heatmap provides a wealth of information on a single page at the beginning of the report, allowing underperforming parts of the pipeline to be quickly tracked or failed samples to be detected. The underlying quality scores are automatically exported to a text file and can be readily used for automated annotation of data sets and to trigger notifications. Since PTXQC supports QC measures for many checkpoints along the shotgun proteomics pipeline, it is also suitable for performance optimization.

Additionally, but not less important, a structured QC is necessary for every proteomics platform to ensure a constant level of performance. For example, we use PTXQC to monitor instrument performance over time using a human cell line standard.

Future extensions of PTXQC include support for qcML, an XML-based reporting format for QC and addition of other quality metrics, such as reporter-ion fragmentation efficiency.

SOFTWARE INFORMATION

Runtime

If the MaxQuant result folder is placed on a local spinning hard disk, then a small number of samples is processed on the order of minutes on a standard desktop PC. A larger study comprising 350 Raw files, featuring a MaxQuant result folder

of 25 GB, was processed in 75 min. On average, each Raw file requires about 15 s for processing.

System Requirements

PTXQC will run on any modern operating system (Windows, Linux, or MacOSX) where the R software¹⁷ can be installed and usually requires less than 2 GB of RAM. For larger studies with more than 100 Raw files, we recommend a 64 bit operating system with at least 8 GB of RAM.

MaxQuant Support

PTXQC was designed to support a wide range of MaxQuant versions, starting from MaxQuant 1.0.13 to the current version, 1.5. Recent versions of MaxQuant provide additional functionality (e.g., Match-between-runs). PTXQC will automatically detect their presence and incorporate the data into the report. Note that PTXQC can currently read only MaxQuant txt files. Output from software packages other than MaxQuant would require appropriate reformatting into a MaxQuant-like CSV format to enable processing by PTXQC.

Target Audience

PTXQC is designed for a wide audience (including technicians operating the instrument, biologists providing the sample, or bioinformaticians conducting downstream analysis) and can be run from within R (all operating systems) or using a convenient drag-and-drop functionality (Windows only), requiring basic computer skills only.

Software Availability and Documentation

The software is available open source under a GPL license at <https://github.com/cbielow/PTXQC>, along with documentation (for users and developers) and the sample data used here. PTXQC is actively used in our lab, ensuring future maintenance. We welcome suggestions and contributions from the community.

Sample Data

All data used in this article was obtained from (PXD000427) or uploaded to (PXD003133, PXD003134) the PRIDE archive.²⁴ For more information, see [Supporting Information 1](#).

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jproteome.5b00780](https://doi.org/10.1021/acs.jproteome.5b00780).

Complete reports for the data sets generated by PTXQC (ZIP)

Summary of data sets, including PRIDE archive identifiers, Figure S1, and a detailed description of all metrics and scoring functions (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(C.B.) E-mail: chris.bielow@mdc-berlin.de.

*(S.K.) E-mail: stefan.kempa@mdc-berlin.de. Tel: +49 30 9406 3114. Fax: +49 30 9406 49164.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank Olga Vvedenskaya for critically reading the manuscript prior to submission. C.B. was supported by the

HepatomaSys project (grant no. 0316172B), funded by the German Federal Ministry of Education and Research (BMBF). G.M. and S.K. gratefully acknowledge funding by BMBF and the Senate of Berlin via the Berlin Institute for Medical Systems Biology.

■ ABBREVIATIONS

iTRAQ, isobaric tag for relative and absolute quantitation; SILAC, stable isotope labeling by amino acids in cell culture; QA, quality analysis; QC, quality control; NIST, National Institute of Standards and Technology; OMSSA, Open Mass Spectrometry Search Algorithm; PTXQC, Proteomics Quality Control; LFQ, label-free quantification; RT, retention time; FDR, false discovery rate; MC, missed cleavages; MBR, Match-between-runs; ppm, parts per million; RSD, relative standard deviation; TMT, tandem mass tag; PCA, principal component analysis

■ REFERENCES

- (1) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (2) Tabb, D. L. Quality assessment for clinical proteomics. *Clin. Biochem.* **2013**, *46*, 411–420.
- (3) Petricoin, E. F., III; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002**, *359*, 572–577.
- (4) Baggerly, K. A.; Morris, J. S.; Coombes, K. R. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* **2004**, *20*, 777–785.
- (5) Rodriguez, H.; Snyder, M.; Uhlén, M.; Andrews, P.; Beavis, R.; Borchers, C.; Chalkley, R. J.; Cho, S. Y.; Cottingham, K.; Dunn, M. Recommendations from the 2008 international summit on proteomics data release and sharing policy: The Amsterdam principles. *J. Proteome Res.* **2009**, *8*, 3689–3692.
- (6) Kinsinger, C. R.; Apffel, J.; Baker, M.; Bian, X.; Borchers, C. H.; Bradshaw, R.; Brusniak, M.-Y.; Chan, D. W.; Deutsch, E. W.; Doman, B.; et al. Recommendations for Mass Spectrometry Data Quality Metrics for Open Access Data (Corollary to the Amsterdam Principles). *J. Proteome Res.* **2012**, *11*, 1412–1419.
- (7) Rudnick, P. A.; Clauser, K. R.; Kilpatrick, L. E.; Tchekhovskoi, D. V.; Neta, P.; Blonder, N.; Billheimer, D. D.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; et al. Performance Metrics for Liquid Chromatography-Tandem Mass Spectrometry Systems in Proteomics Analyses. *Mol. Cell. Proteomics* **2010**, *9*, 225–241.
- (8) Paulovich, A. G.; et al. Interlaboratory Study Characterizing a Yeast Performance Standard for Benchmarking LC-MS Platform Performance. *Mol. Cell. Proteomics* **2010**, *9*, 242–254.
- (9) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3*, 958–964.
- (10) Lam, H.; Deutsch, E.; Eddes, J.; Eng, J.; King, N.; Yang, S.; Roth, J.; Kilpatrick, L.; Neta, P.; Stein, S. SpectraST: An open-source MS/MS spectramatching library search tool for targeted proteomics, 54th ASMS Conference on Mass Spectrometry, Seattle, Washington, May 28–June 1, 2006.
- (11) Ma, Z.-Q.; Polzin, K. O.; Dasari, S.; Chambers, M. C.; Schilling, B.; Gibson, B. W.; Tran, B. Q.; Vega-Montoto, L.; Liebler, D. C.; Tabb, D. L. QuMeter: multivendor performance metrics for LC-MS/MS proteomics instrumentation. *Anal. Chem.* **2012**, *84*, 5845–5850.
- (12) Taylor, R. M.; Dance, J.; Taylor, R. J.; Prince, J. T. Metricator: quality assessment for mass spectrometry-based proteomics. *Bioinformatics* **2013**, *29*, 2948–2949.

(13) Walzer, M.; et al. qcML: An Exchange Format for Quality Control Metrics from Mass Spectrometry Experiments. *Mol. Cell. Proteomics* **2014**, *13*, 1905–1913.

(14) Pichler, P.; Mazanek, M.; Dusberger, F.; Weilnböck, L.; Huber, C. G.; Stingl, C.; Luider, T. M.; Straube, W. L.; Köcher, T.; Mechtler, K. SIMPATIQQCO: a server-based software suite which facilitates monitoring the time course of LC–MS performance metrics on Orbitrap instruments. *J. Proteome Res.* **2012**, *11*, 5540–5547.

(15) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *J. Proteome Res.* **2014**, *13*, 3679–3684.

(16) Amidan, B. G.; Orton, D. J.; LaMarche, B. L.; Monroe, M. E.; Moore, R. J.; Venzin, A. M.; Smith, R. D.; Sego, L. H.; Tardiff, M. F.; Payne, S. H. Signatures for Mass Spectrometry Data Quality. *J. Proteome Res.* **2014**, *13*, 2215–2222.

(17) R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2014.

(18) Gatto, L.; Breckels, L. M.; Naake, T.; Gibb, S. Visualization of proteomics data using R and Bioconductor. *Proteomics* **2015**, *15*, 1375–1389.

(19) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10*, 1794–1805.

(20) Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513–2526.

(21) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics* **2012**, *11*, M1111.014050.

(22) Chiva, C.; Ortega, M.; Sabidó, E. Influence of the Digestion Technique, Protease, and Missed Cleavage Peptides in Protein Quantitation. *J. Proteome Res.* **2014**, *13*, 3979–86.

(23) Licker, V.; Turck, N.; Kövari, E.; Burkhardt, K.; Côte, M.; Surini-Demiri, M.; Lobrinus, J. A.; Sanchez, J.-C.; Burkhard, P. R. Proteomic analysis of human substantia nigra identifies novel candidates involved in Parkinson's disease pathogenesis. *Proteomics* **2014**, *14*, 784–794.

(24) Vizcaino, J. A.; et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **2013**, *41*, D1063–9.

(25) Drexler, H. G.; Uphoff, C. C. Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention. *Cytotechnology* **2002**, *39*, 75–90.

(26) Noble, W. S. Mass spectrometrists should search only for peptides they care about. *Nat. Methods* **2015**, *12*, 605–608.

(27) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015**, *31*, 926–932.