Supporting Information

For “QuaMeter: Multi-Vendor Performance Metrics for Shotgun Proteomics Instrumentation”

Ze-Qiang Ma[1], Kenneth O. Polzin[2], Surendra Dasari[1], Matthew C. Chambers[1], Birgit Schilling[3], Bradford W. Gibson[3], Bao Q. Tran[4,5], Lorenzo Vega-Montoto[1], Daniel C. Liebler[6], David L. Tabb[1,5,6*]

1 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232-8575
2 Chemical and Physical Biology Admissions Program, Vanderbilt University Medical Center, Nashville, TN 37232-0301
3 Buck Institute for Research on Aging, Novato, California 94945
4 Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232-2358
5 Mass Spectrometry Research Center, Vanderbilt University Medical Center, Nashville, Tennessee 37232-8575
6 Department of Biochemistry, Vanderbilt University Medical Center, Nashville, TN 37232-6350
* Corresponding author: (phone) 615-936-0380; (fax) 615-343-8372; (email) david.l.tabb@vanderbilt.edu

Abstract

LC-MS/MS based proteomics studies rely on stable analytical system performance that can be evaluated by objective criteria. The National Institute of Standards and Technology (NIST) introduced the MSQC software to compute diverse metrics from experimental LC-MS/MS data, enabling the quality analysis and quality control (QA/QC) of proteomics instrumentation. In practice, however, several attributes of the MSQC software prevent its use for routine instrument monitoring. Here we present QuaMeter, an open-source tool that improves MSQC in several aspects. QuaMeter can directly read raw data from instruments manufactured by different vendors. The software can work with a wide variety of peptide identification software for improved reliability and flexibility. Finally, QC metrics implemented in QuaMeter are rigorously defined and tested. The source code and binary versions of QuaMeter are available under Apache 2.0 License at http://fenchurch.mc.vanderbilt.edu.

Data Sources

Thermo Fisher LTQ-XL Data Set

This data constitutes of 280 routine BSA runs at the Jim Ayers Institute for Precancer Detection and Diagnosis at Vanderbilt University. The data set was previously used to test the Pepitome software and the experimental description is published by Dasari et al. (Dasari, S.; Chambers, M. C.; Martinez, M. A.; Carpenter, K.; Ham, A.-J.; Vega-Montoto, L.; Tabb, D. L. J. Proteome Res. 2012.). The files average 11917 MS/MS scans each. All files were searched using MyriMatch against a RefSeq BOVINE database or using Pepitome to match the NIST BSA spectral library (http://peptide.nist.gov).

Thermo Fisher LTQ-Orbitrap Data Set

This data set was also collected at the Jim Ayers Institute for Precancer Detection and Diagnosis at Vanderbilt University. Experimental settings were exactly the same as above except 10x BSA peptide mixtures were used instead of 1x BSA. All samples were analyzed on a Thermo Fisher LTQ-Orbitrap mass spectrometer. A total of 53 files were used in this manuscript. Spectra were searched using MyriMatch against a RefSeq BOVINE database. The files average 3417 MS/MS scans.

Thermo Fisher LTQ-Velos Data Set

This data set was previously used to test the ScanRanker software and the experimental details were described in the original publication (Ma, Z.-Q.; Chambers, M. C.; Ham, A.-J. L.; Cheek, K. L.; Whitwell, C. W.; Aerni, H.-R.; Schilling, B.; Miller, A. W.; Caprioli, R. M.; Tabb, D. L. J. Proteome Res. 2011, 10, 2896-2904.). Five technical replicates were collected for a yeast lysate on a Thermo Fisher LTQ-Velos instrument. The files average 38466 MS/MS scans each. Spectra were identified using MyriMatch against a yeast database (http://www.yeastgenome.org) downloaded on March 2009. All files were also searched by Pepitome against the NIST yeast spectral library (http://peptide.nist.gov).

Bruker Daltonics HCT Ultra Data Set

Stock bovine serum albumin (BSA) solution prepared in 100mM ammonium bicarbonate buffer was digested overnight with sequencing grade Trypsin (Promega) at enzyme-to-substrate ratio of 1:50 at 37°C. LC-MS/MS analysis was carried out on an Eksigent 1D-nanopump coupling to a Bruker HCT Ultra iontrap mass spectrometer. The mobile phases were water and acetonitrile with 0.1% formic acid as an additive. 2uL of working BSA solution of 100fmol/uL was load by a FAMOS autosampler with a 10uL sample loop onto a 3cm, 360/100 OD/ID trap column of 5um Jupiter C18 particles with loading aqueous buffer of 0.1% formic acid at flow rate of 1uL/min and separated on a 15cm 360/75um OD/ID PicoFrit emitter column from New Objective packed with 3um Jupiter C18 particles. Both columns were in house packed. Tryptic peptides eluted during a gradient from 2% to 50% acetonitrile at flow rate of 250nL/min. Different LC-gradients were applied throughout the data collection. LC-MS/MS data was acquired in positive ionization mode with scan segments of 1 precursor ion scan (m/z=375-2000) in standard enhanced and 3 subsequent tandem MS scans of three most abundant ions in ultra scan mode. Scan average was set to 2 and ion charge control (ICC) was 200,000. Singly charge ions were excluded from tandem MS and a 1 minute dynamic exclusion was activated for each peptide after two MS tandem acquisitions. Instrument raw files were converted to mzML format by the MSConvert tool in

ProteoWizard. Since Bruker data extraction library does not write precursor spectrum reference information in mzML files, which is required for running QuaMeter, a Perl script is created to add precursor spectrum references to MS/MS scans. The latest previous MS1 scan is assumed as the precursor of neighboring MS/MS scans. 24 files were collected with averagely 3467 MS/MS scan each. All spectra were searched using MyriMatch against a RefSeq BOVINE protein database and identifications were filtered by IDPicker.

AB SCIEX QSTAR Elite Data Set

Predigested, tryptic beta-galactosidase solutions (*E. coli*) were obtained from AB SCIEX and used as quality control samples. Samples were analyzed by reverse-phase nano-HPLC-ESI-MS/MS using an Eksigent nano-LC 2D HPLC system (Eksigent, Dublin, CA) which was directly connected to a quadrupole time-of-flight (QqTOF) QSTAR Elite mass spectrometer (AB SCIEX, Concord, CAN). Briefly, peptide mixtures were loaded from the autosampler (using partial loop fill methods) onto a guard column (C18 Acclaim PepMap100, 300 µm I.D. x 5 mm, 5 µm particle size, 100 Å pore size, Dionex, Sunnyvale, CA) and washed with the loading solvent (0.1 % formic acid, flow rate: 20 µL/min) for 5 min. Subsequently, samples were transferred onto the analytical C18-nanocapillary HPLC column (C18 Acclaim PepMap100, 75 µm I.D. x 15 cm, 3 µm particle size, 100 Å pore size, Dionex, Sunnyvale, CA) and eluted at a flow rate of 300 nL/min using the following gradient: 2-30% solvent B in A (from 0-15 min), 30-80% solvent B in A (from 15-17 min) and at 80% solvent B in A (from 17-20 min), with a total runtime of 52 min (including mobile phase equilibration). Solvents were prepared as described below for the TripleTOF 5600. Mass spectra (ESI-MS) and tandem mass spectra (ESI-MS/MS) were recorded in positive-ion mode with a resolution of 12,000-15,000 full-width half-maximum. For collision induced dissociation tandem mass spectrometry (CID-MS/MS), the mass window for precursor ion selection of the quadrupole mass analyzer was set to ± 1 m/z. The precursor ions were fragmented in a collision cell using nitrogen as the collision gas. Advanced information dependent acquisition (IDA) was used for MS/MS collection, including QSTAR Elite (Analyst QS 2.0) specific features, such as “Smart Collision” and “Smart Exit” (fragment intensity multiplier set to 2.0 and maximum accumulation time at 2.5 sec) to obtain MS/MS spectra for up to seven most abundant precursor ions following each survey scan. Dynamic exclusion features were based on value M not m/z and were set to exclusion mass width 50 mDa and exclusion duration of 60 sec. All 23 files were searched using MyriMatch against a UniProt *E.coli* database and identifications passing 5% FDR in IDPicker analysis were confident IDs.

AB SCIEX TripleTOF 5600 Data Set

Predigested, tryptic beta-galactosidase solutions (*E. coli*) were obtained from AB SCIEX and used as quality control samples. Samples were analyzed by reverse-phase HPLC-ESI-MS/MS using an Eksigent Ultra Plus nano-LC 2D HPLC system (Dublin, CA) which was directly connected to a new generation quadrupole time-of-flight (QqTOF) TripleTOF 5600 mass spectrometer (AB SCIEX, Concord, CAN) in direct injection mode. The autosampler was operated in full injection mode overfilling a 1 µl loop with 3 µl analyte for optimal sample delivery reproducibility. Briefly, after injection, peptide mixtures were transferred onto the analytical C18-nanocapillary HPLC column (C18 Acclaim PepMap100, 75 µm I.D. x 15 cm, 3 µm particle size, 100 Å pore size, Dionex, Sunnyvale, CA) and eluted at a flow rate of 300 nL/min using the following gradient: at 5% solvent B in A (from 0-13 min), 5-35% solvent B in A (from 13-29 min), 35-80% solvent B in A (from 29-31 min) and at 80% solvent B in A (from 31-37 min), with a total runtime of 58 min including mobile phase equilibration. Solvents were prepared as follows, mobile phase A: 2% acetonitrile/98% of 0.1% formic acid (v/v) in water, and mobile phase B: 98% acetonitrile/2% of 0.1% formic acid (v/v) in water. Mass spectra and tandem mass spectra were recorded in positive-ion and “high-sensitivity” mode, with a resolution of ~35,000 full-width half-

maximum in MS1 mode and ~15,000 in MS/MS mode. The nanospray needle voltage was 2,400 V in HPLC-MS mode. After acquisition of ~ 5 to 6 samples, TOF MS spectra and TOF MS/MS spectra were automatically calibrated during dynamic LC-MS & MS/MS autocalibration acquisitions injecting 25 fmol beta-galactosidase. For collision induced dissociation tandem mass spectrometry (CID-MS/MS), the mass window for precursor ion selection of the quadrupole mass analyzer was set to ± 1 m/z. The precursor ions were fragmented in a collision cell using nitrogen as the collision gas. Advanced information dependent acquisition (IDA) was used for MS/MS collection on the TripleTOF 5600 (Analyst TF 1.5) to obtain MS/MS spectra for the 20 most abundant precursor ions following each survey MS1 scan (allowing for 50 msec acquisition time per each MS/MS). Dynamic exclusion features were based on value M not m/z and were set to an exclusion mass width of 50 mDa and an exclusion duration of 15 sec. All 60 files were searched using MyriMatch against a UniProt *E.Coli* database and processed by IDPicker.

MyriMatch Configuration

PrecursorMzToleranceRule = "avg" for LTQ-XL and LTQ-Velos, “mono” for LTQ-Orbitrap, QSTAR Elite and TripleTOF
AvgPrecursorMzTolerance = 1.5 m/z
MonoPrecursorMzTolerance = 10 ppm for LTQ-Orbitrap, 100 ppm for QSTAR Elite and 50 ppm for TripleTOF
MonoisotopeAdjustmentSet = [-1,2]
FragmentMzTolerance = 0.5 m/z for LTQ-XL, LTQ-Velos and LTQ-Orbitrap, 0.4 m/z for QSTAR Elite and 50ppm for TripleTOF
StaticMods = "C 57.0215"
DynamicMods = "M ^ 15.9949 (Q * -17.026"
MinTerminiCleavages = 1
CleavageRules = "Trypsin/P"
MaxMissedCleavages = 2
MaxDynamicMods = 2
DecoyPrefix = "rev_"
NumChargeStates = 3
OutputFormat= "pepXML"
SpectrumListFilters = "peakPicking false 2-"
TicCutoffPercentage = 0.98
FragmentationAutoRule = true
MaxResultRank = 5
MinPeptideMass = 0 Da
MaxPeptideMass = 10000 Da
MinPeptideLength = 5
MaxPeptideLength = 75
UseSmartPlusThreeModel = false
ProteinSampleSize = 100
ComputeXCorr = true
UseMultipleProcessors = true

Pepitome Configuration

PrecursorMzToleranceRule = "avg"
MonoPrecursorMzTolerance = "10 ppm"
AvgPrecursorMzTolerance = "1.5 mz"
FragmentMzTolerance = "0.5 mz"
SpectrumListFilters = "peakPicking true 2-;chargeStatePredictor false 3 2 0.9"
RecalculateLibPepMasses = false
CleanLibSpectra = true
LibTicCuttoffPercentage = 0.98f
LibMaxPeakCount = 100
MonoisotopeAdjustmentSet = "0"
TicCutoffPercentage = 0.98
MaxPeakCount = 150
CleavageRules = "trypsin"
MaxMissedCleavages = 2
MinTerminiCleavages = 1
MinPeptideLength = 5
DynamicMods = "C % 57.021"
MaxDynamicMods = 3
StaticMods = ""
MaxResultRank = 2
FASTARefreshResults = false

IDPicker Configuration

Maximum FDR = 0.05
Minimum distinct peptides =2
Minimum additional peptides = 1
Minimum spectra per protein = 2

QuaMeter Configuration

RawDataPath = ../mzMLs/ # where to find the raw files for each idpDB
RawDataFormat = mzML # the file extension to expect for the raw files; e.g. mzML, mzXML, raw
Instrument = LTQ # if set to LTQ, average masses are used, else monoisotopic masses
ScoreCutoff = 0.05 # IDPicker FDR cutoff
ChromatogramMzLowerOffset = 1.0mz # the lower bound of the window for building chromatograms; can be in m/z or ppm
ChromatogramMzUpperOffset = 1.0mz # the upper bound of the window for building chromatograms; can be in m/z or ppm
ChromatogramOutput = false # if true, creates an mz5 file with the chromatograms (best viewed with SeeMS)

Metrics Definition

Table S-1. Description of metrics. Table adapted from original NIST MSQC publication (Rudnick, P. A. et al. Mol. Cell Proteomics 2010, 9, 225 - 241). Th, thomsons; ID, identification; IDed, identified; max, maximum; Betw/in, between/within; Med, median; pctile, percentile; tryp, tryptic; Pep, peptide; interQ, interquartile; fract, fraction.

Code	Category	Metric group	Metric	Units	Optimal	Purpose/use	Description
C-1A	Chromatography	Fraction of repeat peptide IDs with divergent RT	-4 min	Fraction	↓	Estimates very early peak broadening	Fraction of all peptides identified at least 4 min earlier than max MS1 for ID
C-1B	Chromatography	Fraction of repeat peptide IDs with divergent RT	+4 min	Fraction	↓	Estimates very late peak broadening	Fraction of all peptides identified at least 4 min later than max MS1 for ID
C-2A	Chromatography	Interquartile retention time period	Period (min)	min	↑	Longer times indicate better chromatographic separation	Time period over which 50% of peptides were identified
C-2B	Chromatography	Interquartile retention time period	Pep ID rate	Peps/min	↑	Higher rates indicate efficient sampling and identification	Rate of peptide identification during C-2A
C-3A	Chromatography	Peak width at half-height for IDs	Median value	s	↓	Sharper peak widths indicate better chromatographic resolution	Median peak widths for all identified unique peptides (s)
C-3B	Chromatography	Peak width at half-height for IDs	Interquartile distance	s	↓	Tighter distributions indicate more peak width uniformity	Measure of the distribution of the peak widths; small values indicate consistency
C-4A	Chromatography	Peak widths at half-max over RT deciles for IDs	First decile	s	↓	Estimates peak widths at the beginning of the gradient	Median peak width for identified peptides in first RT decile (early)
C-4B	Chromatography	Peak widths at half-max over RT deciles for IDs	Last decile	s	↓	Estimates peak widths at the end of the gradient	Median peak width for identified peptides in last RT decile (late)
C-4C	Chromatography	Peak widths at half-max over RT deciles for IDs	Median value	s	↓	Estimates peak widths in the middle of the gradient	Median peak width for identified peptides in median RT decile (middle)

DS-1A	Dynamic sampling	Ratios of peptide ions IDed by different numbers of spectra	Once/twice	Ratio	↑	Estimates oversampling	Ratio of peptides identified by 1 spectrum divided by number identified by 2 spectra
DS1-B	Dynamic sampling	Ratios of peptide ions IDed by different numbers of spectra	Twice/thrice	Ratio	↑	Estimates oversampling	Ratio of peptides identified by 2 spectra divided by number identified by 3 spectra
DS-2A	Dynamic sampling	Spectrum counts	MS1 scans/full	Count	↓	Fewer MS1 scans indicates more sampling	Number of MS1 scans taken over C-2A
DS-2B	Dynamic sampling	Spectrum counts	MS2 scans	Count	↑	More MS2 scans indicates more sampling	Number of MS2 scans taken over C-2A
DS-3A	Dynamic sampling	MS1 max/MS1 sampled abundance ratio IDs	Median all IDs	Ratio	↓	Estimates position on peak where sampled for peptides of all abundances	Ratio of MS1 maximum to MS1 value at sampling for median decile of peptides by MS1 maximum intensity (1 = sampled at peak maxima)
DS-3B	Dynamic sampling	MS1 max/MS1 sampled abundance ratio IDs	Med bottom 1/2	Ratio	↓	Estimates position on peak where sampled for least abundant 50% of peptides	Ratio of MS1 maximum to MS1 value at sampling for bottom 50% of peptides by MS1 maximum intensity (1 = sampled at peak maxima)
IS-1A	Ion source	MS1 during middle (and early) peptide retention period	MS1 jumps >10×	Count	↓	Flags ESI instability	Number of times where MS1 signal greatly decreased between adjacent scans more than 10-fold (electrospray instability)
IS-1B	Ion source	MS1 during middle (and early) peptide retention period	MS1 falls >10×	Count	↓	Flags ESI instability	Number of times where MS1 signal greatly increased between adjacent scans more than 10-fold (electrospray instability)
IS-2	Ion source	Precursor m/z for IDs	Median	Th	↓	Higher median m/z can correlate with inefficient or partial ionization	Median m/z value for all identified peptides (unique ions)

IS-3A	Ion source	IDs by charge state (relative to 2+)	Charge 1+	Ratio	↓	High ratios of 1+/2+ peptides may indicate inefficient ionization	Number of 1+ peptides over 2+ peptides
IS-3B	Ion source	IDs by charge state (relative to 2+)	Charge 3+	Ratio	↓	Higher ratios of 3+/2+ peptides may preferentially favor longer peptides	Number of 3+ peptides over 2+ peptides
IS-3C	Ion source	IDs by charge state (relative to 2+)	Charge 4+	Ratio	↓	Higher ratios of 4+/2+ peptides may preferentially favor longer peptides	Number of 4+ peptides over 2+ peptides
MS1-1	MS1 signal	Ion injection times for IDs	MS1 median	ms	↓	Lower times indicate an abundance of ions	MS1 ion injection time
MS1-2A	MS1 signal	MS1 during middle (and early) peptide retention period	S/N median	None	↑	Higher MS1 S/N may correlate with higher signal discrimination	Median signal-to-noise value (ratio of maximum to median peak height) for MS1 spectra up to and including C-2A
MS1-2B	MS1 signal	MS1 during middle (and early) peptide retention period	TIC median	Counts/1,000	↑	Estimates the total absolute signal for peptides (may vary significantly between instruments)	Median TIC value for identified peptides over same time period as used for MS1-2A
MS1-3A	MS1 signal	MS1 ID max	95/5 pctile	Ratio	↑	Estimates the dynamic range of the peptide signals	Ratio of 95th over 5th percentile MS1 maximum intensity values for identified peptides (approximates dynamic range of signal)
MS1-3B	MS1 signal	MS1 ID max	Median	Counts	↑	Estimates the median MS1 signal for peptides	Median maximum MS1 value for identified peptides
MS1-4A	MS1 signal	MS1 intensity variation for peptidesa	Within series	Percent	↓	Used to monitor relative intensity differences with a series	Average of between series intensity variations for identified peptides
MS1-4B	MS1 signal	MS1 intensity variation for peptidesa	Betw/in	Ratio	↓	Used to monitor relative intensity differences with a series compared with between series	Ratio of average intensity variation between series to average intensity variation within a series (low values indicate similarity between series)
MS1-5A	MS1 signal	Precursor m/z - Peptide ion m/z	Median	Th	↓	Measures the accuracy of the identifications	Median real value of precursor errors

MS1-5B	MS1 signal	Precursor m/z - Peptide ion m/z	Mean absolute	Th	↓	Measures the accuracy of the identifications	Mean of the absolute precursor errors
MS1-5C	MS1 signal	Precursor m/z - Peptide ion m/z	ppm median	ppm	↓	Measures the accuracy of the identifications	Median real value of precursor errors in ppm
MS1-5D	MS1 signal	Precursor m/z - Peptide ion m/z	ppm interQ	ppm	↓	Measures the distribution of the real accuracy measurements	Interquartile distance in ppm of the precursor errors
MS2-1	MS2 signal	Ion injection times for IDs	MS2 median	ms	↓		MS2 ion injection time
MS2-2	MS2 signal	MS2 ID S/N	Median	Ratio	↑	Higher S/N correlates with increased frequency of peptide identification	Median S/N (ratio of maximum to median peak height) for identified MS2 spectra
MS2-3	MS2 signal	MS2 ID peaks	Median	Count	↑	Higher peak counts can correlate with more signal	Median number of peaks in an MS2 scan
MS2-4A	MS2 signal	Fraction of MS2 identified at different MS1 max quartiles	ID fract Q1	Fraction	↑	Higher fractions of identified MS2 spectra indicate efficiency of detection and sampling	Fraction of total MS2 scans identified in the first quartile of peptides sorted by MS1 maximum intensity
MS2-4B	MS2 signal	Fraction of MS2 identified at different MS1 max quartiles	ID fract Q2	Fraction	↑	Higher fractions of identified MS2 spectra indicate efficiency of detection and sampling	Fraction of total MS2 scans identified in the second quartile of peptides sorted by MS1 maximum intensity
MS2-4C	MS2 signal	Fraction of MS2 identified at different MS1 max quartiles	ID fract Q3	Fraction	↑	Higher fractions of identified MS2 spectra indicate efficiency of detection and sampling	Fraction of total MS2 scans identified in the third quartile of peptides sorted by MS1 maximum intensity
MS2-4D	MS2 signal	Fraction of MS2 identified at different MS1 max quartiles	ID fract Q4	Fraction	↑	Higher fractions of identified MS2 spectra indicate efficiency of detection and sampling	Fraction of total MS2 scans identified in the last quartile of peptides sorted by MS1 maximum intensity
P-1	Peptide identification	MS2 ID score	Median	fval	↑	Higher scores correlate with higher S/N and frequency of identification	Median peptide identification score for all peptides; higher scores generally correlate with increased MS2 S/N

P-2A	Peptide identification	Tryptic peptide counts	Identifications	Count	↑	Total identifications correlate with high levels of peptide signals, performance	Number of MS2 spectra identifying tryptic peptide ions (total “spectral counts”)
P-2B	Peptide identification	Tryptic peptide counts	Ions	Count	↑	A good overall performance measure	Number of tryptic peptide ions identified; ions differing by charge state and/or modification state are counted separately
P-2C	Peptide identification	Tryptic peptide counts	Peptides	Count	↑	A good overall performance measure	Number of unique tryptic peptide sequences identified
P-3	Peptide identification	Peptide counts	Semi/tryp peptides	Ratio		Indicates prevalence of semitryptic peptides in sample; increasing ratios may indicate changes in sample or in source	

Multi-vendor Performance

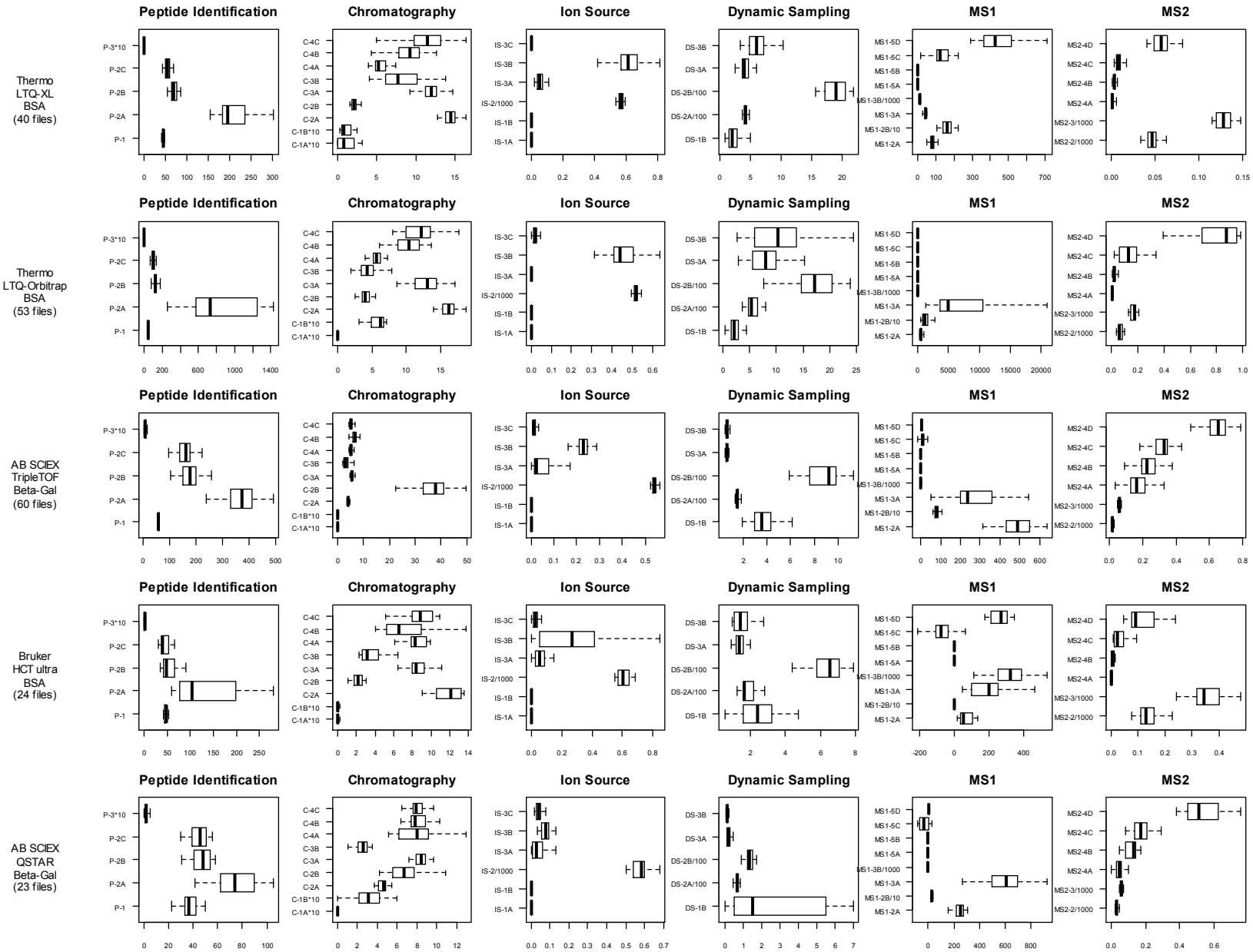
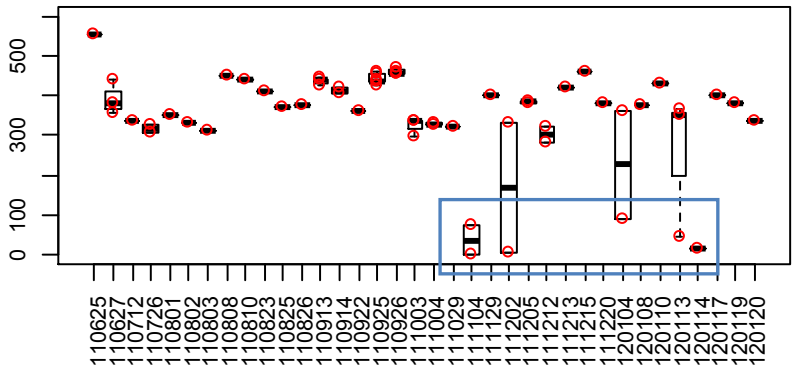

Thermo LTQ-XL BSA (40 files)
Thermo LTQ-Orbitrap BSA (53 files)
AB SCIEX TripleTOF Beta-Gal (60 files)
Bruker HCT ultra BSA (24 files)
AB SCIEX QSTAR Beta-Gal (23 files)
Peptide Identification
Chromatography
Ion Source
Dynamic Sampling
MS1
MS2
P-3*10
P-2C
P-2B
P-2A
P-1
C-4C
C-4B
C-4A
C-3B
C-3A
C-2B
C-2A
C-1B*10
C-1A*10
IS-3C
IS-3B
IS-3A
IS-2/1000
IS-1B
IS-1A
DS-3B
DS-3A
DS-2B/100
DS-2A/100
DS-1B
MS1-5D
MS1-5C
MS1-5B
MS1-5A
MS1-3B/1000
MS1-3A
MS1-2B/10
MS1-2A
MS2-4D
MS2-4C
MS2-4B
MS2-4A
MS2-3/1000
MS2-2/1000

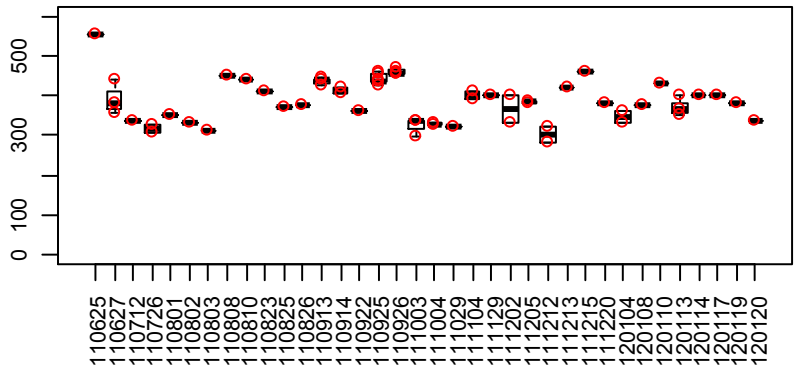
Figure S-1. QuaMeter computes QC metrics for multiple instrument platforms. Standard QC samples were analyzed on five instruments. Metrics describing the same category of performance are plotted in a single panel. Some metrics were scaled to fit the plot. It should be noted that it is not appropriate to compare metrics between instrument platforms because data sets were collected under different settings. An overview and definitions for all metrics are provided in Supplementary Table S1.

TripleTOF Outliers

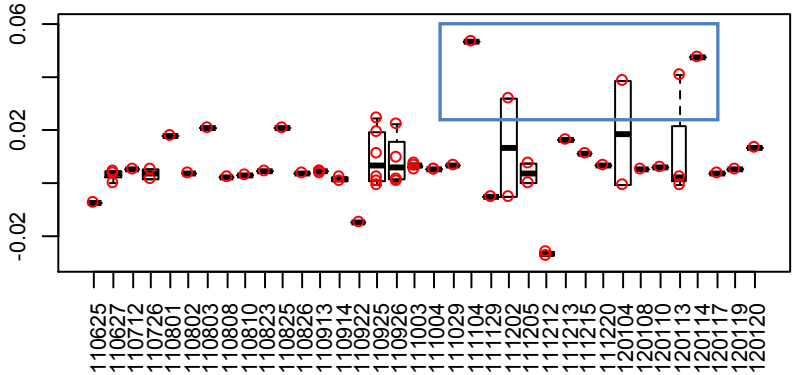
P-2A: Number of MS2 spectra identifying tryptic peptide ions
(before recalibration)



P-2A: Number of MS2 spectra identifying tryptic peptide ions
(after recalibration)



MS1-5A: Median real value of precursor errors
(before recalibration)



MS1-5A: Median real value of precursor errors
(after recalibration)

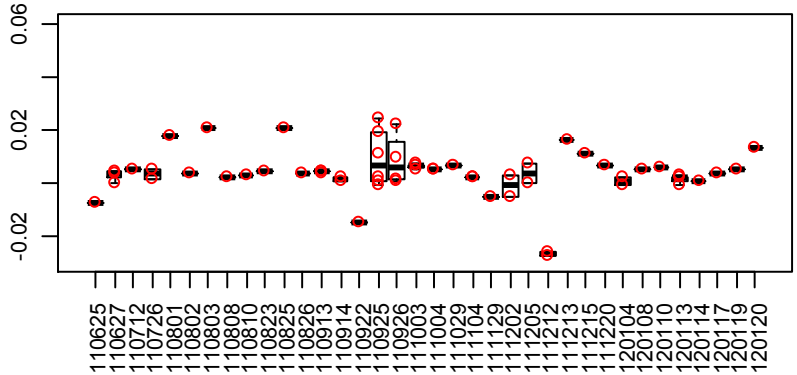


Figure S-2. QuaMeter metrics help to spot abnormal instrument performance. Metrics computed from TripleTOF data were plotted by date. Six files were recognized as outliers in early analysis that had very low number of identifications (blue box in top-left P-2A metric). A close examination of other QuaMeter metrics showed they all associated with high precursor mass accuracy errors (blue box in bottom-left MS1-5A metric; one point missing for 111104 because zero identification passed 5% FDR threshold from this file). The instrument log revealed that these files had a mass accuracy shift due to temperature variation in the laboratory (caused by air-handler failures). Recalibrating these files yielded narrow precursor errors (bottom-right panel) and comparable number of identifications as other experiments (top-right panel).

Scripts to Combine QuaMeter Output

The following scripts can be used to combine all QuaMeter metrics files in a folder to a table in text format.

Windows MS-DOS batch file:

```
@ECHO OFF
echo Filename C-1A C-1B C-2A C-2B C-3A C-3B C-4A C-4B C-4C DS-1A DS-1B DS-2A DS-2B DS-3A DS-3B IS-1A IS-1B IS-2 IS-3A IS-3B IS-3C MS1-1 MS1-2A MS1-2B MS1-3A MS1-3B MS1-5A MS1-5B MS1-5C MS1-5D MS2-1 MS2-2 MS2-3 MS2-4A MS2-4B MS2-4C MS2-4D P-1 P-2A P-2B P-2C P-3 > merged-quameter-metrics.txt
for %%x in (*.qual.txt) do (
        echo processing %%x...
        more +1 "%%x" >> merged-quameter-metrics.txt
)
PAUSE
```

Linux bash file:

```
#!/bin/bash
for x in *.qual.txt;
do
        head -1 "$x" > merged-quameter-metrics.txt
        break
done;
for x in *.qual.txt;
do
        echo processing $x...
        more +2 "$x" >> merged-quameter-metrics.txt
done;
```

R Script

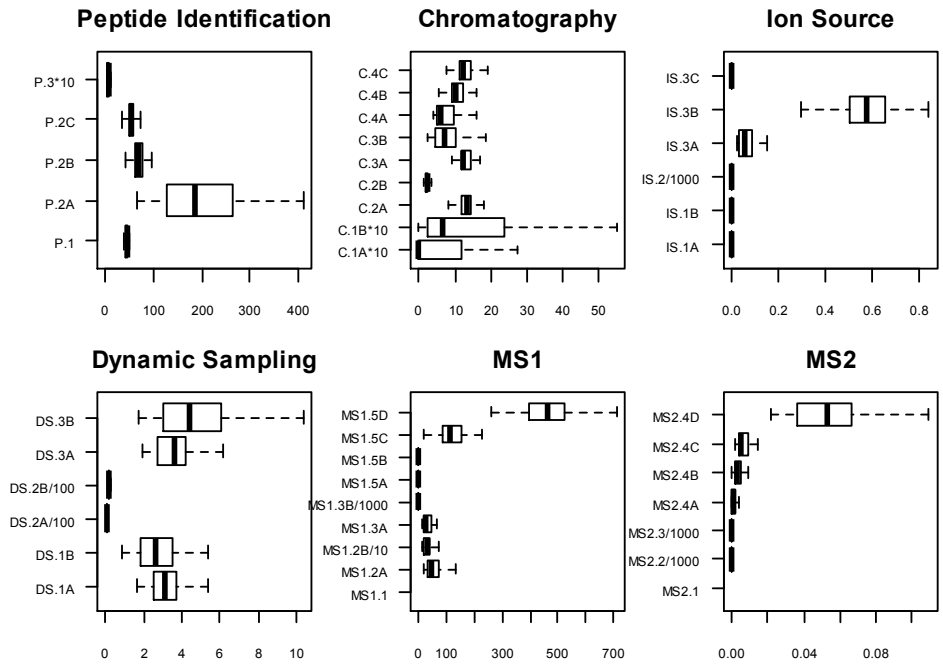
The following R script can be used to combine all metrics file in a folder to a single table and generate a boxplot to illustrate variation of each metric.

```
### read all qual.txt files and combine to one dataframe
readQualFiles <- function(qualFilePath) {
          qualFiles <- Sys.glob(paste(qualFilePath,"/*.qual.txt", sep=''))
          allData <- lapply(qualFiles, function(.file){
            dat<-read.table(.file, header=TRUE)
            dat    # return the dataframe
          })
          mergedDF <- do.call(rbind, allData)  # combine into a single dataframe
```

```
        colnames(mergedDF) <-
c("Filename","C.1A","C.1B","C.2A","C.2B","C.3A","C.3B","C.4A","C.4B","C.4C","DS.1A","DS.1B","DS.2A","
DS.2B","DS.3A","DS.3B","IS.1A","IS.1B","IS.2","IS.3A","IS.3B","IS.3C","MS1.1","MS1.2A","MS1.2B","MS1.
3A","MS1.3B","MS1.5A","MS1.5B","MS1.5C","MS1.5D","MS2.1","MS2.2","MS2.3","MS2.4A","MS2.4B","
MS2.4C","MS2.4D","P.1","P.2A","P.2B","P.2C","P.3")
        return(mergedDF)
}

qualdata <- readQualFiles("C:/datafolder/quameter")  # path to the folder where QuaMeter metrics files
stored

### scale some columns for plot
scaleColumns <- function(datatable) {
                datatable$C.1A <- datatable$C.1A * 10
                datatable$C.1B <- datatable$C.1B * 10
                # datatable$C.2A <- datatable$C.2A / 10
                # datatable$C.2B <- datatable$C.2B / 10
                datatable$IS.2 <- datatable$IS.2 / 1000
                datatable$DS.2A <- datatable$DS.2A / 100
                datatable$DS.2B <- datatable$DS.2B / 100
                datatable$MS1.2B <- datatable$MS1.2B / 10
                datatable$MS1.3B <- datatable$MS1.3B / 1000
                datatable$MS2.3 <- datatable$MS2.3 / 1000
                datatable$MS2.2 <- datatable$MS2.2 / 1000
                datatable$P.3 <- datatable$P.3 * 10

                return(datatable)
}

mains <- c("Peptide Identification","Chromatography","Ion Source", "Dynamic Sampling","MS1", "MS2")
columns <- list(c("P.1","P.2A","P.2B","P.2C","P.3"),
c("C.1A","C.1B","C.2A","C.2B","C.3A","C.3B","C.4A","C.4B","C.4C"),
c("IS.1A","IS.1B","IS.2","IS.3A","IS.3B","IS.3C"),c("DS.1A","DS.1B","DS.2A","DS.2B","DS.3A","DS.3B"),c("
MS1.1","MS1.2A","MS1.2B","MS1.3A","MS1.3B","MS1.5A","MS1.5B","MS1.5C","MS1.5D"),c("MS2.1","
MS2.2","MS2.3","MS2.4A","MS2.4B","MS2.4C","MS2.4D")) ### removed filename column
boxnames <- list(c("P.1","P.2A","P.2B","P.2C","P.3*10"),
c("C.1A*10","C.1B*10","C.2A","C.2B","C.3A","C.3B","C.4A","C.4B","C.4C"),
c("IS.1A","IS.1B","IS.2/1000","IS.3A","IS.3B","IS.3C"),c("DS.1A","DS.1B","DS.2A/100","DS.2B/100","DS.3A
","DS.3B"),c("MS1.1","MS1.2A","MS1.2B/10","MS1.3A","MS1.3B/1000","MS1.5A","MS1.5B","MS1.5C","
MS1.5D"),c("MS2.1","MS2.2/1000","MS2.3/1000","MS2.4A","MS2.4B","MS2.4C","MS2.4D")) ###
removed filename column

par(mfrow=c(2,3),mar=c(2,4,3,1) + 0.1 )
qualdata <- scaleColumns(qualdata)
m <- 1
for (j in 1:length(columns))
{
```

```
        boxplot(qualdata[,columns[[j]]],horizontal=TRUE, las=1, main=mains[m], names=boxnames[[j]],
cex.axis=0.65,outline=FALSE)
        m <- m + 1
}
```

An example plot is shown as below:



Scripts to Add NativeIDs to Search Results

QuaMeter requires the same spectral file that the identification tool was fed with. The idpDB file should have the same base name as the input spectral file. QuaMeter then associates identified peptide with their corresponding MS/MS scans using NativeIDs. If identification tools do not contain NativeIDs in their search outputs, they can be manually added to the search output before IDPicker process. Below are example Perl scripts to add NativeIDs to pepXML files. Here we use NativeID examples from Thermo RAW files. NativeIDs for other vendor formats can be found in raw files or mzML files.

In this example, Thermo Raw format files were converted to mzXML format by the MSConvert tool in ProteoWizard library. The mzXML file was used for X!Tandem and Mascot search. DTA files were generated by the mzxml2search tool in Trans Proteomic Pipeline (TPP) for Sequest search. Search results were converted to pepXML format by tandem2xml, out2xml and mascot scripts, respectively.

For Sequest and X!Tandem results, a single line Perl script can be used to add NativeIDs:

```
perl -p -i.bak -e  's/start_scan="(\d+)"/spectrumNativeID="controllerType=0 controllerNumber=1
scan=$1" start_scan="$1"/' *.pepXML
```

Scripts below can be used to process Mascot results:

```
###############################################################################
### Usage        : see below
### Purpose      : add NativeID to pepXML
### Inputs       :
### Returns      :
### Throws       : no exceptions
### Comments     :
###----------------------------------------------------------------------

=begin comments
assume the MS/MS scans were collected from Thermo instruments

=end comments
=cut

use strict;
use warnings;
use Cwd;
use Carp;
use Getopt::Long;
use File::Basename;

my $usage = << "END_USAGE";
*****************************************************

Usage:   perl $0 -i a.pepXML

*****************************************************
END_USAGE

### Initialization ****************************************************
my ($inFile,$outFile);
my $outDir = getcwd;          #default outDir is the current dir
GetOptions( 'i=s' => \$inFile,
                              'outdir=s' => \$outDir,    #overwrite outdir
                    );
die $usage if (!defined($inFile));
my $outFileBasename = basename($inFile, ".pepXML");
$outFile = $outFileBasename."-addedNativeIDs.pepXML";
my $outFilename = $outFileBasename.".pepXML";
### ---------------------------------------------------------------
open (my $inFile_FH, '<', $inFile) or croak "Cannot open the input file $inFile: $!\n";
open (my $outFile_FH, '>>', "$outDir/$outFile") or croak "Cannot open output file $outFile:$!\n";

my $scanID;
while (my $line = <$inFile_FH>) {
```

```
	if ($line =~ /start_scan="(\d+)"/) {
		# my $scanID =~ s/^0+(.)/$1/;  #remove leading zeros, e.g. Mascot output
		$scanID=$1;
		$scanID =~ s/^0+//;
		$line =~ s/start_scan/spectrumNativeID="controllerType=0 controllerNumber=1
scan=$scanID" start_scan/;
		print {$outFile_FH} $line;
		next;
	}
	print {$outFile_FH} $line;

}

close $inFile_FH;
close $outFile_FH;
rename($inFile, $inFile.".bak") or warn "Couldn't rename $inFile: $!\n";
rename($outFile, $outFilename) or warn "Couldn't rename $outFile: $!\n";
```