



MSstatsQC: Longitudinal system suitability monitoring and quality control for proteomic experiments

Dr. Eralp DOGU

Mugla University, TR

Prof. Olga Vitek Group

Northeastern University

Case Studies in Quantitative Proteomics at ASMS 2018

San Diego

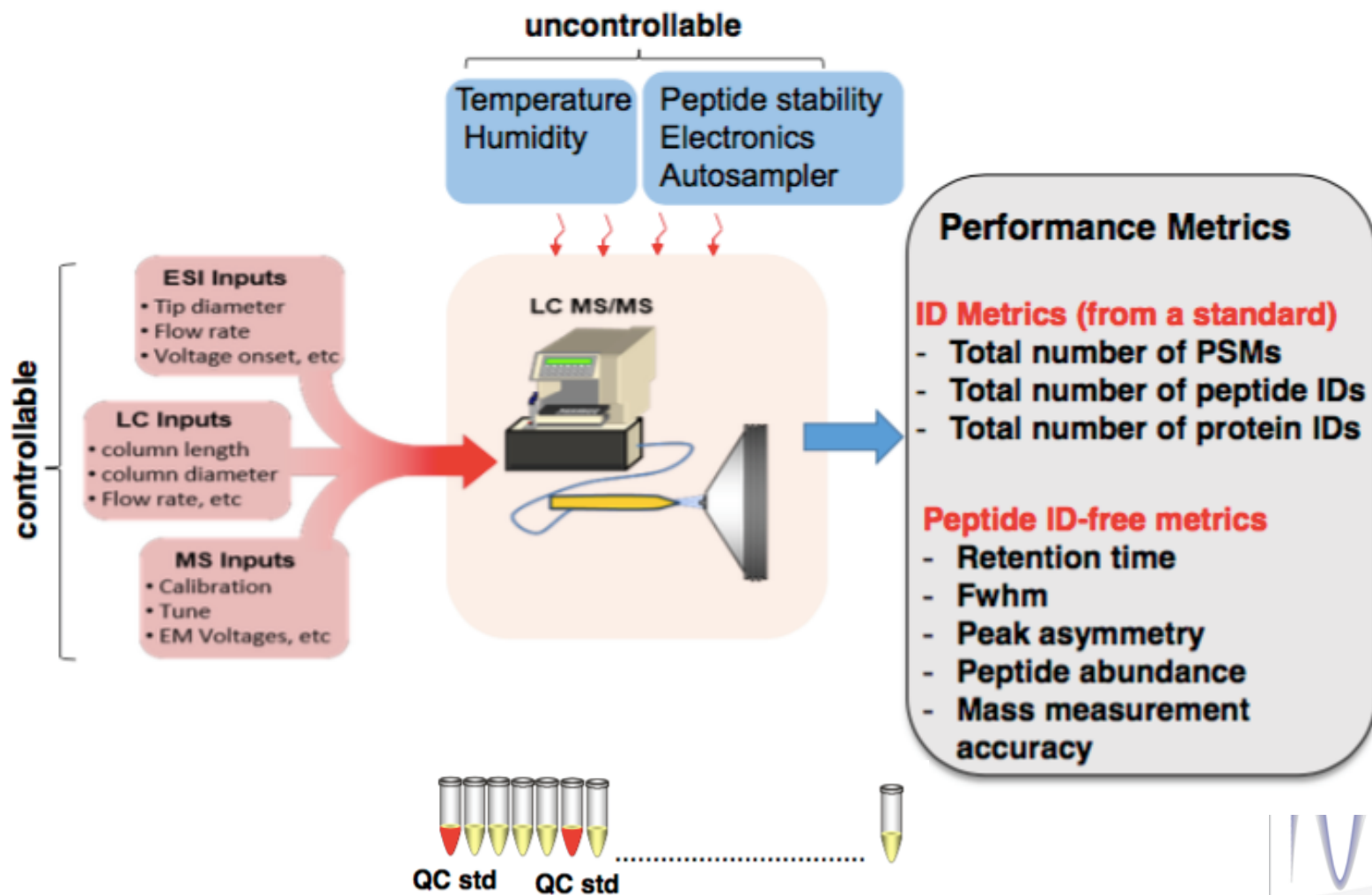


Outline

- 1. Quality assurance and definition of quality**
2. Basics of Statistical Process Control (SPC)
3. MSstatsQC
4. Case studies from CPTAC study 9.1



LC MS/MS is a process!



Outline

1. Quality assurance and definition of quality
- 2. Basics of Statistical Process Control (SPC)**
3. MSstatsQC
4. Case studies from CPTAC study 9.1

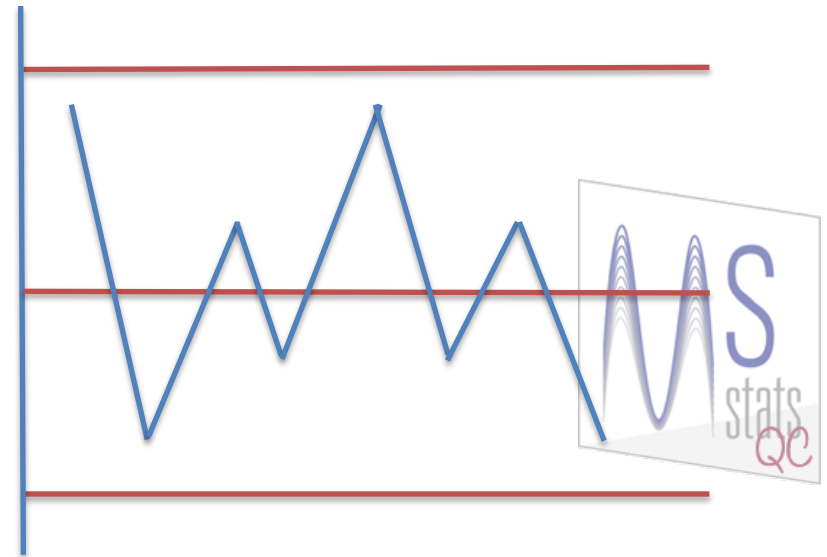


Statistical Process Control (SPC)

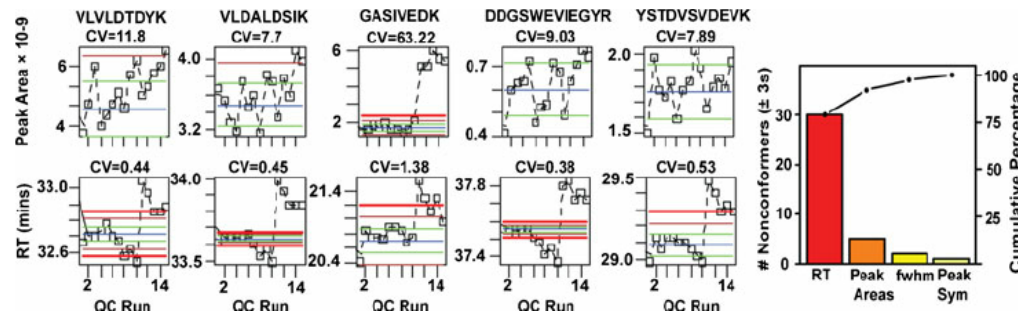
- Typically used for quality control
 - Developed in 1920s at Bell Telephone Laboratories by Walter Shewart to aid in the production of telephone components that were of uniform quality
 - Based on theory of variation
 - Long history of use within manufacturing, healthcare, food and chemical processes

➤ A key SPC tool is the control chart, which is the focus of this presentation

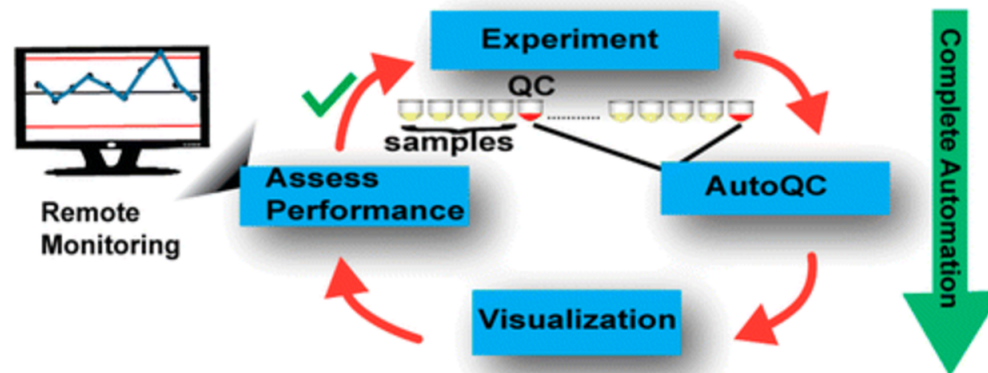
➤ Combines time-series analysis with graphical representation of data



SPC applied to mass spectrometry proteomics



SProCop



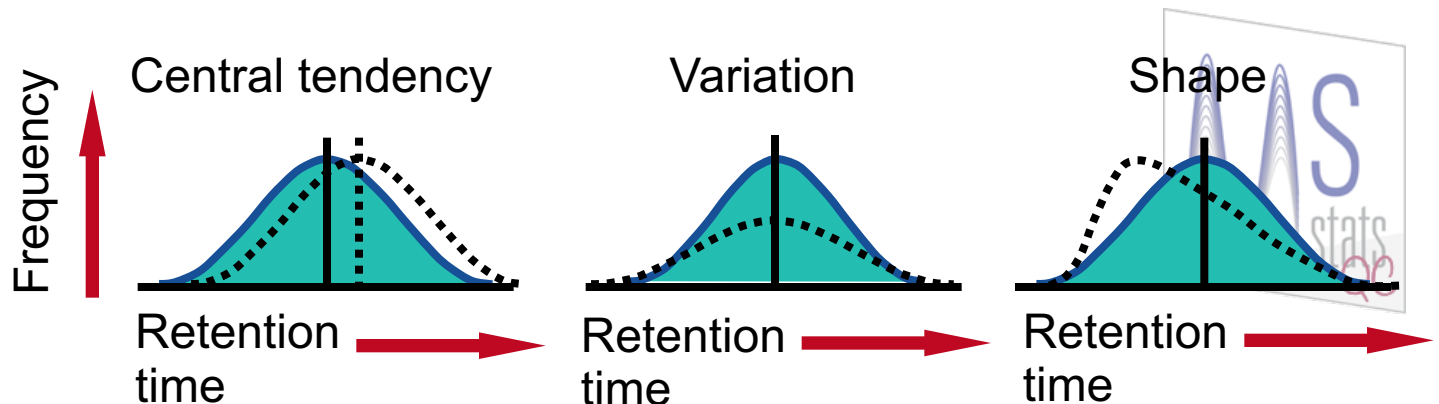
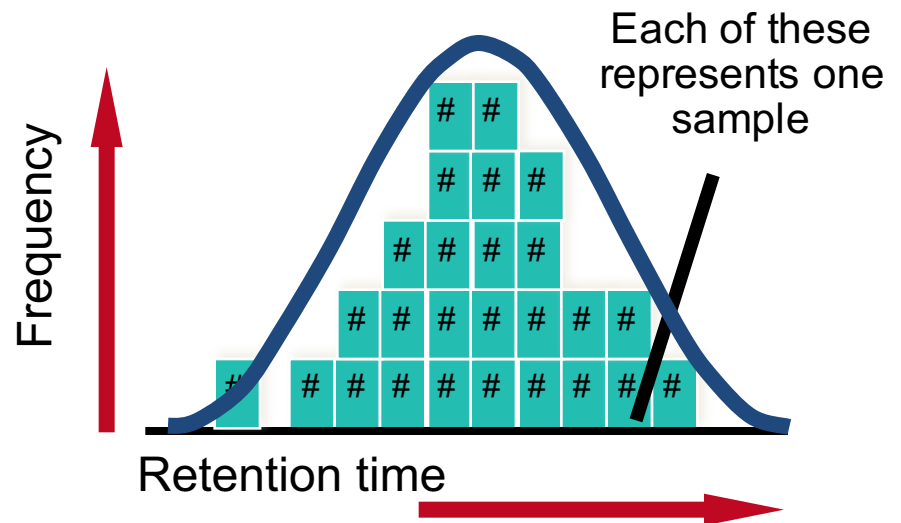
Panorama AutoQC

Bereman et. al. (2014) *J. Am. Soc. Mass Spectrom*
Bereman et. al. (2016) *J. Proteome Res.*

Sampling to set up a program

To measure the process, we take samples and analyze the sample statistics following these steps

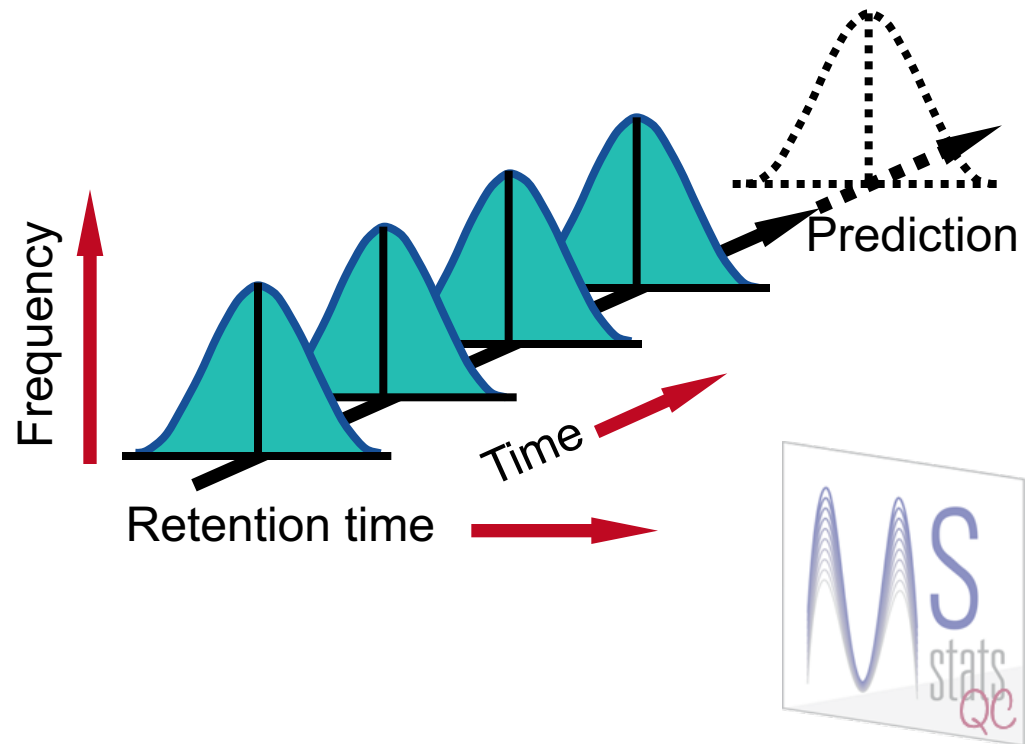
- QC or system suitability samples, vary from each other in terms of retention time
- After enough samples (guide set) are taken from a stable process, they form a pattern called a distribution
- There are many types of distributions, including the normal (bell-shaped) distribution, but distributions do differ in terms of central tendency (mean), standard deviation or variance, and shape



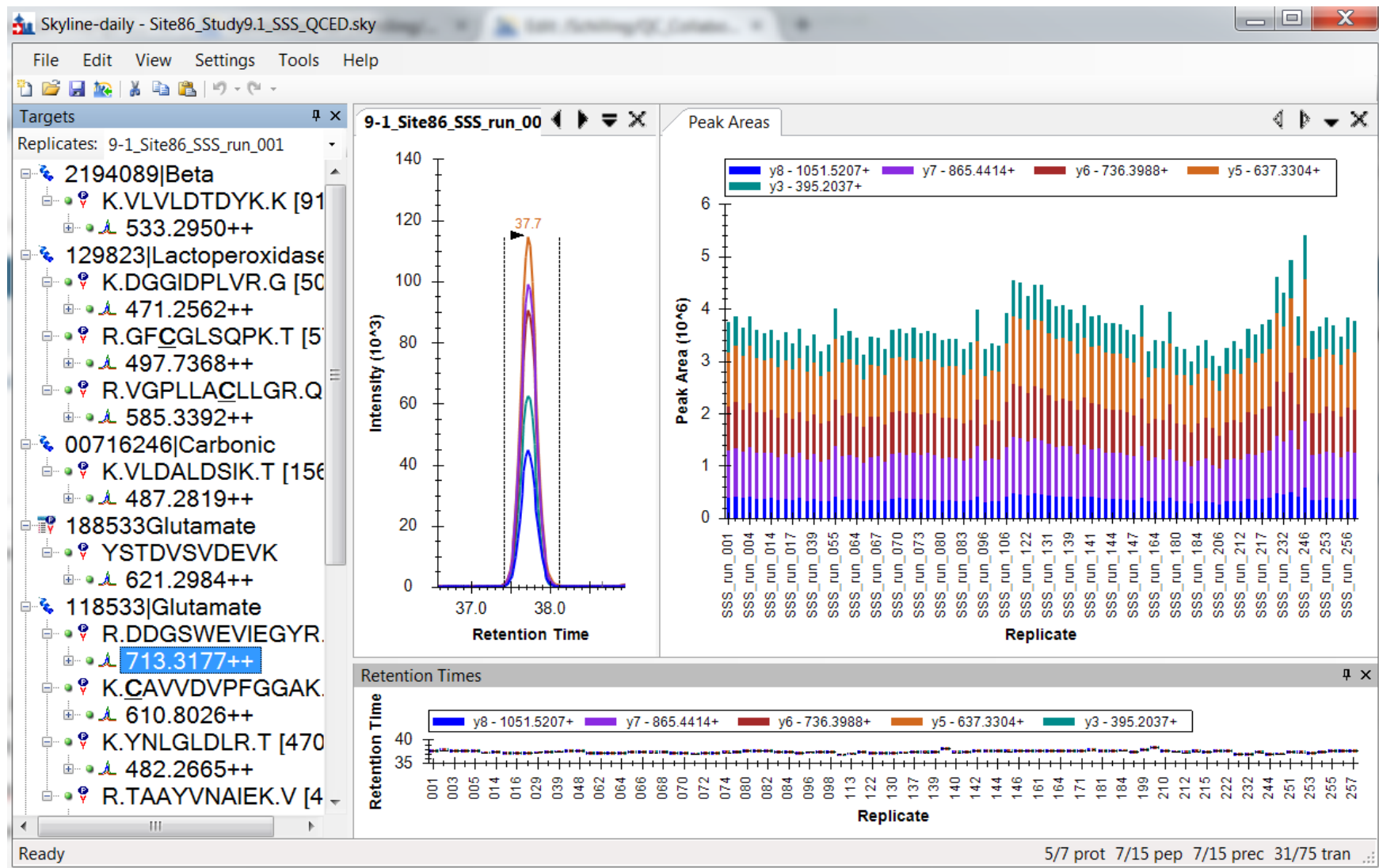
Sampling to set up a suitability program

To measure the performance, we take samples and analyze the sample statistics following these steps

- (d) If only **natural causes** of variation are present, the output of a process forms a distribution that is stable over time and is predictable



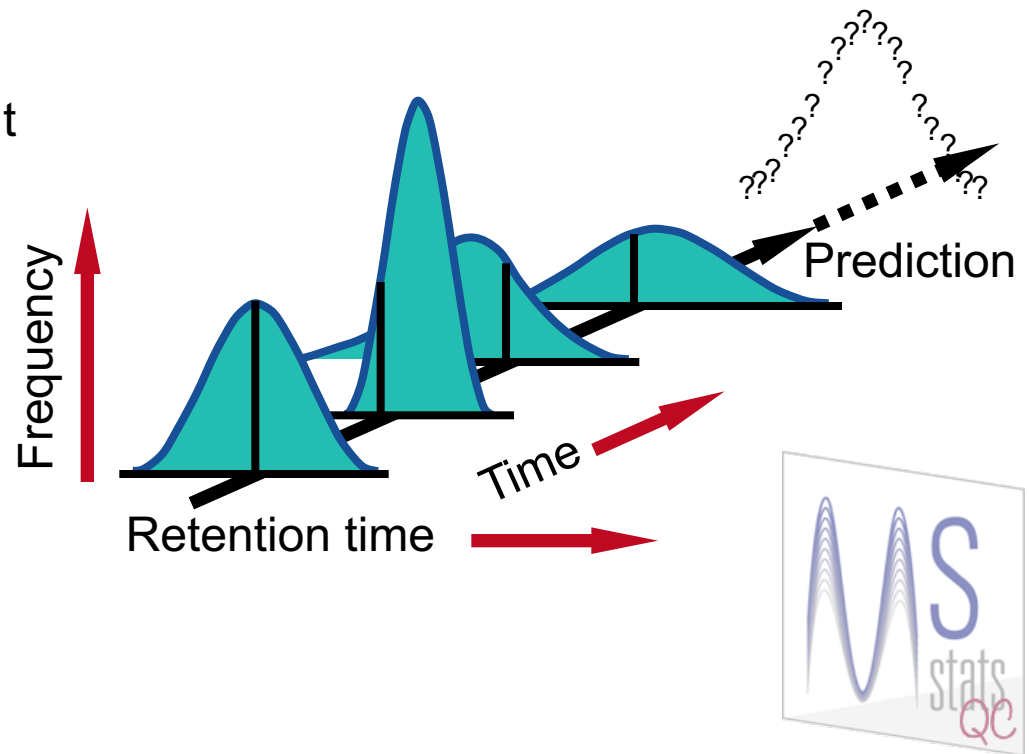
Data : CPTAC Study 9.1 Site 86



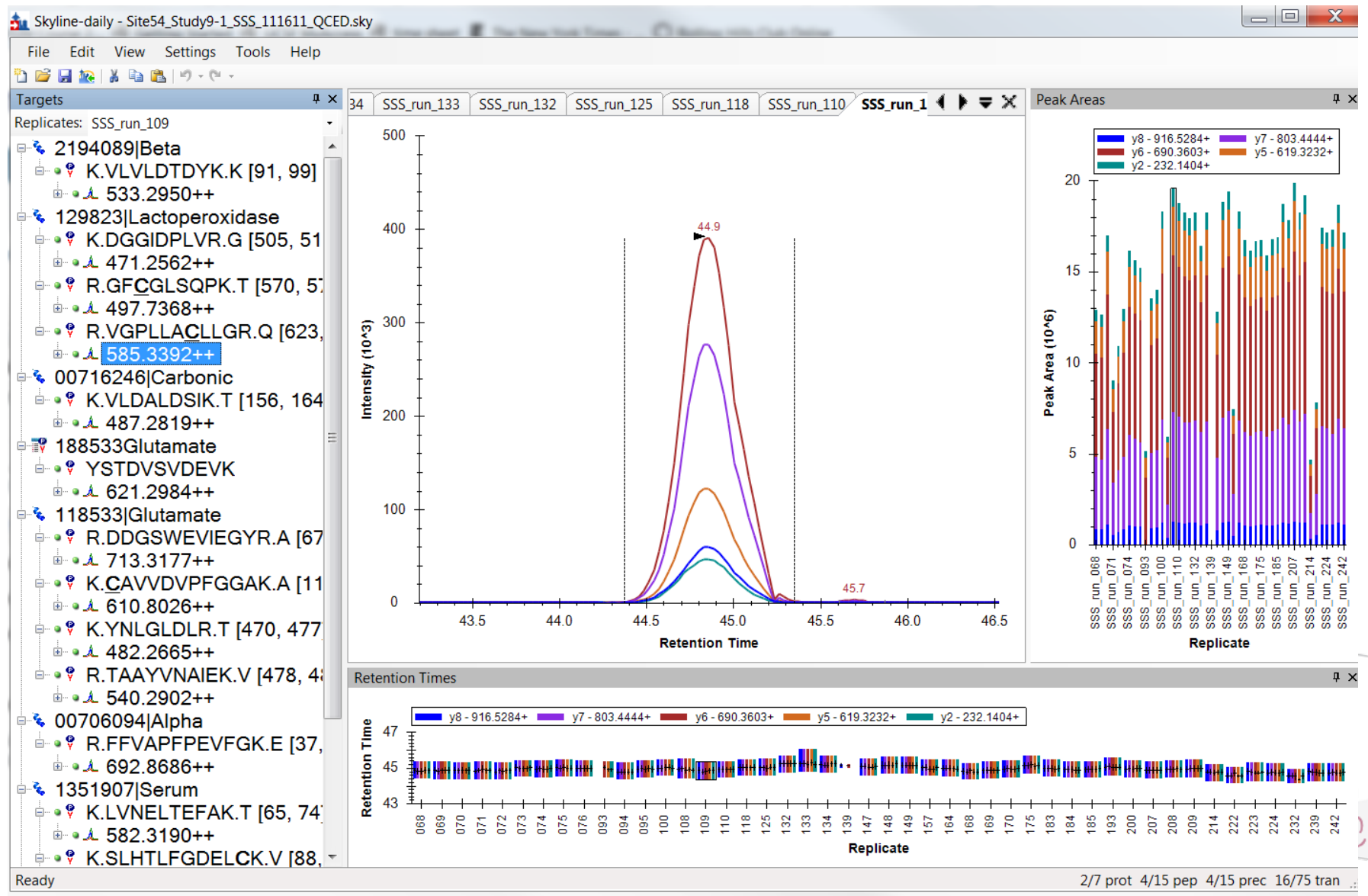
Sampling to set up a suitability program

To measure the performance, we take samples and analyze the sample statistics following these steps

- (e) If **assignable causes** are present, the process output is not stable over time and is not predictable



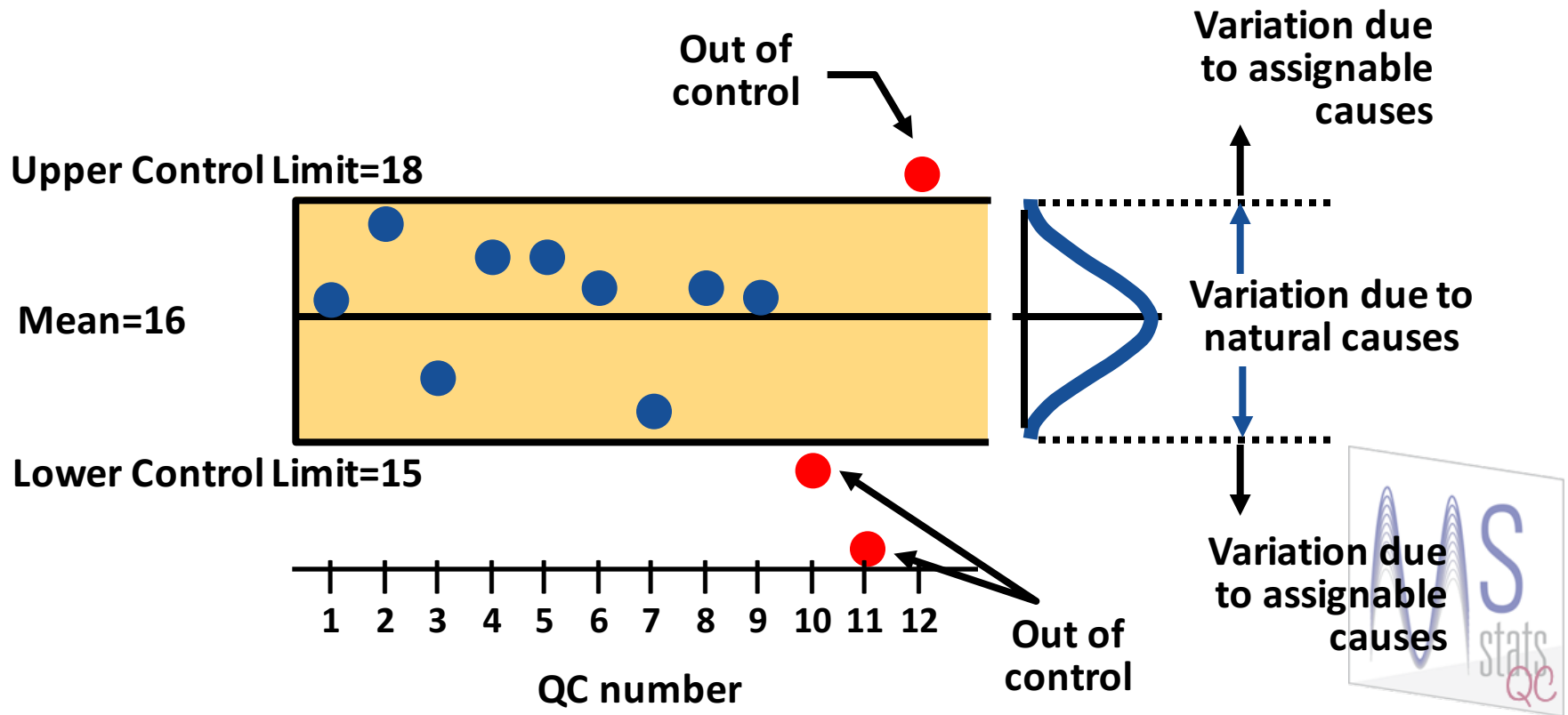
Data : CPTAC Study 9.1 Site 56A



Control Charts

Constructed from historical data, the purpose of control charts is to help distinguish between natural variations and variations due to assignable causes

Control Chart for 12 QC samples for a certain peptide

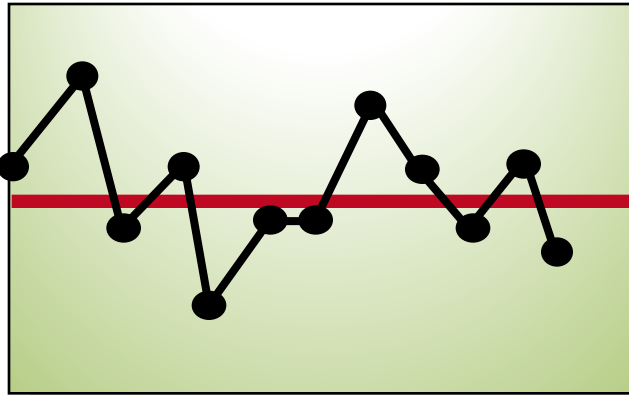


Patterns of control charts

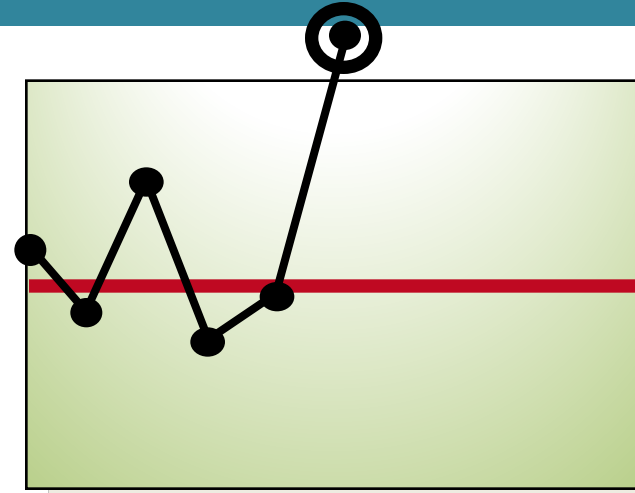
UCL

RT mean

LCL



Normal behavior. Process is "in control."

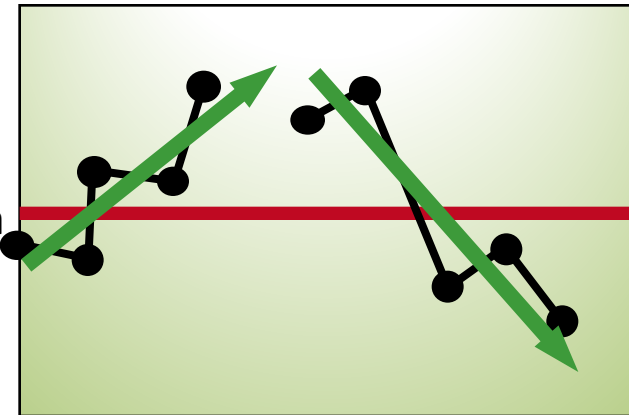


One sample out above (or below). Process is "out of control."

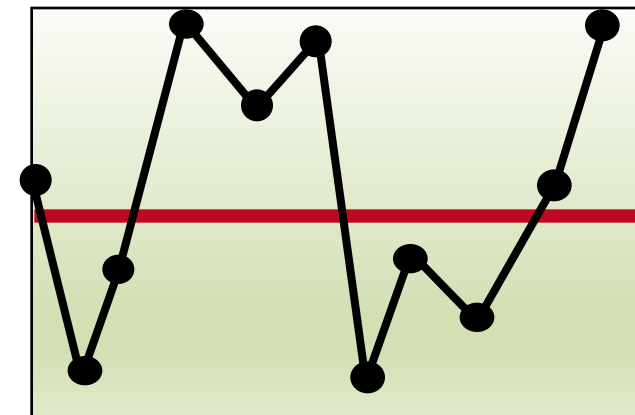
UCL

Peak area mean

LCL



Trends in either direction, 5 points. Investigate for cause of progressive change.

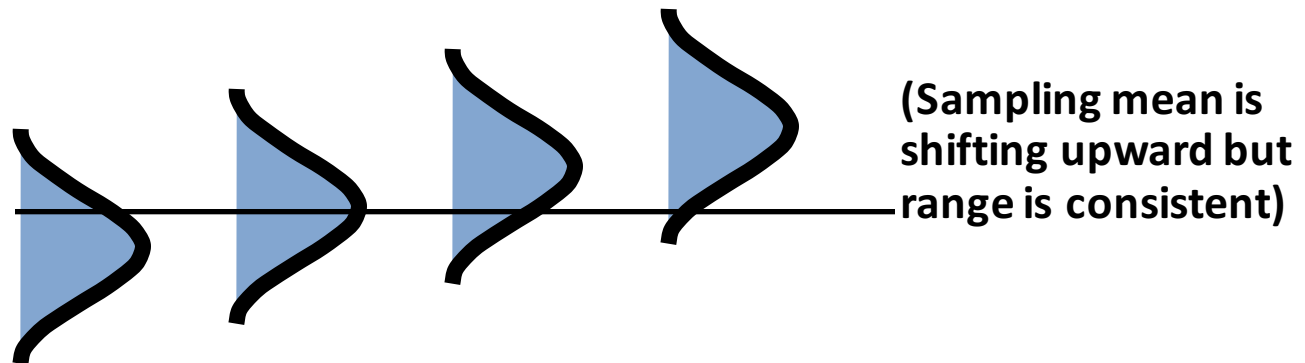


Erratic behavior. Investigate.

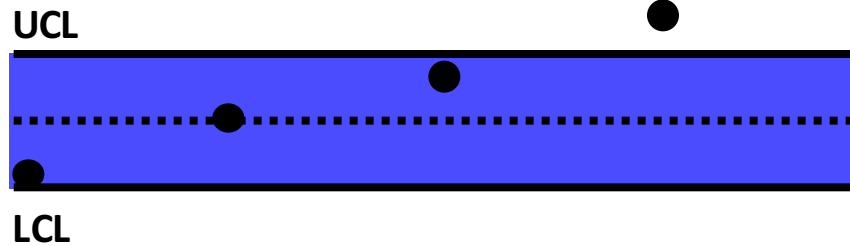
Simultaneous monitoring of LC MS/MS mean and variation

(a)

These sampling distributions result in the charts below

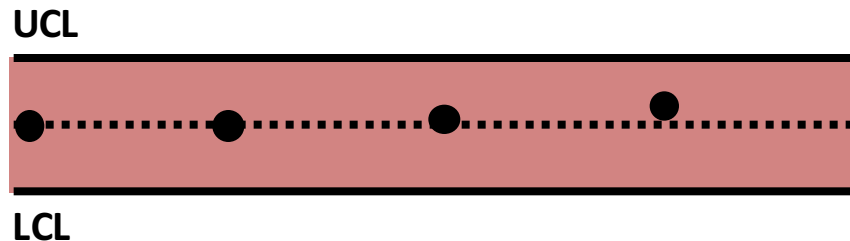


X-chart



(*X* chart detects shift in central tendency)

MR-chart



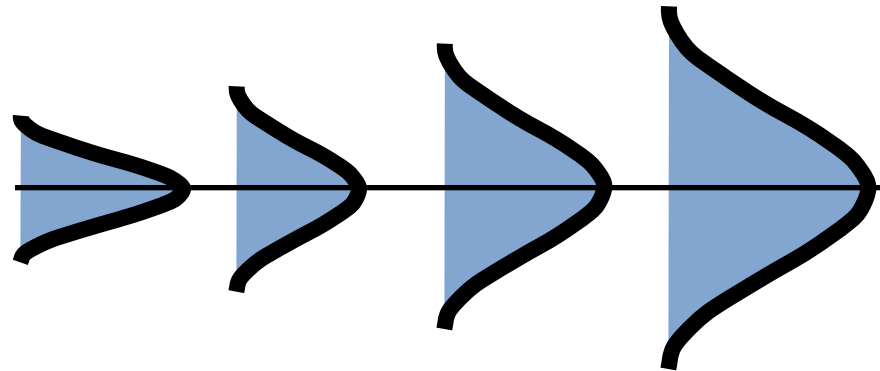
(*MR*-chart does not detect change in mean)



Simultaneous monitoring of LC MS/MS mean and variation

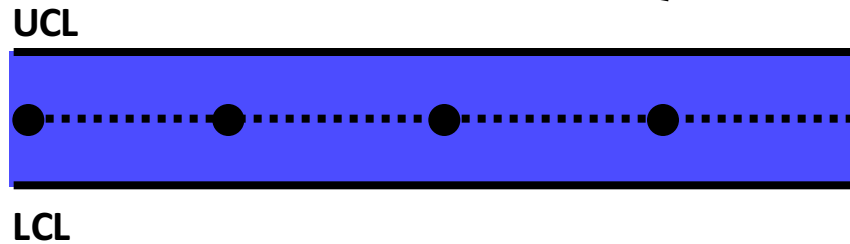
(b)

These sampling distributions result in the charts below



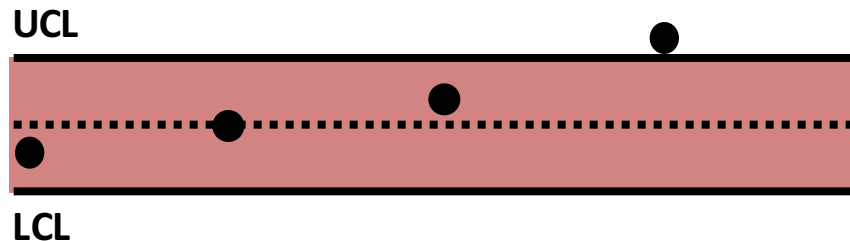
(Sampling mean is constant but dispersion is increasing)

X-chart



(*X*-chart does not detect the increase in dispersion)

MR-chart



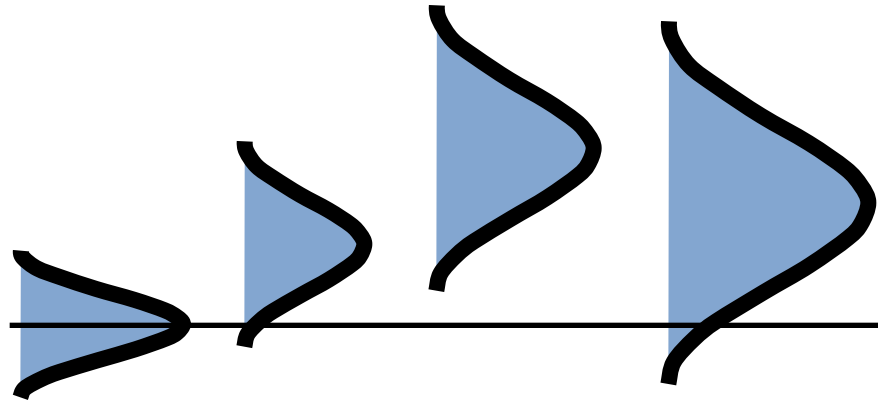
(*MR*-chart detects increase in dispersion)



Simultaneous monitoring of LC MS/MS mean and variation

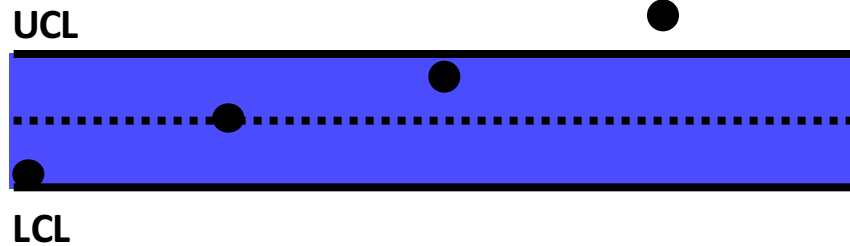
(c)

These sampling distributions result in the charts below



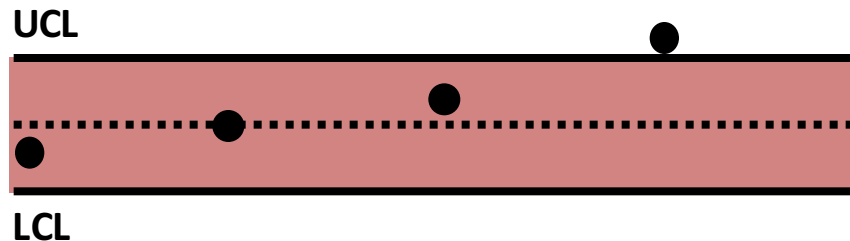
(Sampling mean is constant but dispersion is increasing)

X-chart



(*X*-chart detects shift in central tendency)

MR-chart



(*MR*-chart detects increase in dispersion)



Outline

1. Quality assurance and definition of quality
2. Basics of Statistical Process Control (SPC)
3. **MSstatsQC**
4. Case studies from CPTAC study 9.1



MSstatsQC : statistical tool for longitudinal monitoring

Open-source R-based package and web interface (msstats.org/msstatsqc) for **statistical monitoring** of system suitability and quality control (QC) samples in mass spectrometry-based proteomic experiments.

MSstatsQC

platforms **all** downloads **top 20%** posts **0** in Bioc **0.5 years**
build **ok**

DOI: [10.18129/B9.bioc.MSstatsQC](https://doi.org/10.18129/B9.bioc.MSstatsQC)  

Longitudinal system suitability monitoring and quality control for proteomic experiments

Bioconductor version: Release (3.7)

MSstatsQC is an R package which provides longitudinal system suitability monitoring and quality control tools for proteomic experiments.

Author: Eralp Dogu [aut, cre], Sara Taheri [aut], Olga Vitek [aut]

Maintainer: Eralp Dogu <eralp.dogu at gmail.com>

Citation (from within R, enter `citation("MSstatsQC")`):

Dogu E, Taheri S, Vitek O (2018). *MSstatsQC: Longitudinal system suitability monitoring and quality control for proteomic experiments*. R package version 1.2.0, <http://msstats.org/msstatsqc>.

Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("MSstatsQC")
```

MSstatsQCgui

platforms **all** downloads **available** posts **0** in Bioc **< 6 months**
build **ok**

DOI: [10.18129/B9.bioc.MSstatsQCgui](https://doi.org/10.18129/B9.bioc.MSstatsQCgui)  

A graphical user interface for MSstatsQC package

Bioconductor version: Release (3.7)

MSstatsQCgui is a Shiny app which provides longitudinal system suitability monitoring and quality control tools for proteomic experiments.

Author: Eralp Dogu [aut, cre], Sara Taheri [aut], Olga Vitek [aut]

Maintainer: Eralp Dogu <eralp.dogu at gmail.com>

Citation (from within R, enter `citation("MSstatsQCgui")`):

Dogu E, Taheri S, Vitek O (2018). *MSstatsQCgui: A graphical user interface for MSstatsQC package*. R package version 1.0.0, <http://msstats.org/msstatsqc>.

Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("MSstatsQCgui")
```

NEW!

Package for the web interface

MSstatsQC : statistical tool for longitudinal monitoring

Open-source R-based web interface (www.msstats.org/msstatsqc) for **statistical monitoring** of system suitability and quality control (QC) samples in mass spectrometry-based proteomic experiments.

MSstatsQC



QC data
gathering

Data input and data
processing

Metric
Summary

Control charts
and change
point analysis

 The image part with

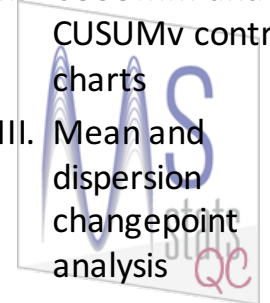
* Test peptides
for special causes of
variation

- I. MSstatsQC compatible experiments
- II. MSstatsQC input

- I. Data input
- II. Data similarity analysis
- III. Data table

- I. Box plots for each suitability metric and peptide
- II. Decision-maps
- III. Metric summaries

- I. XmR control chart
- II. CUSUMm and CUSUMv control charts
- III. Mean and dispersion change point analysis



MSstats compatible experiments and metrics

MS acquisition

- SRM
- DIA or SWATH (available with MSstatsQCgui)
- DDA or shotgun (available with MSstatsQCgui)

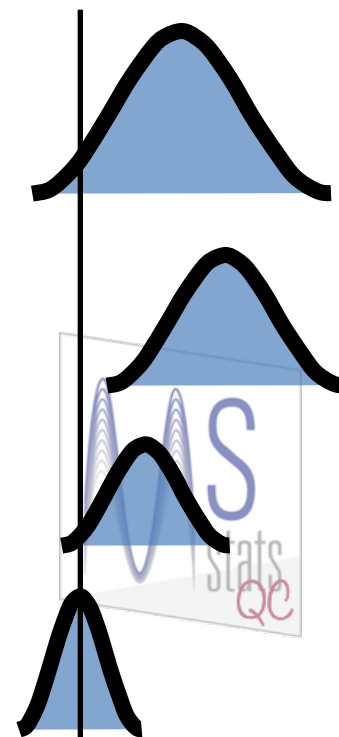
Analysis

- Decision support tools
- Control charts
- Change point analysis

Metrics

- Retention time
- Total peak area
- Full width at half maximum (FWHM)
- Peak asymmetry
- Many more...

When	Mean	Variation
Large shifts	X	MR
Small shifts	CUSUMm	CUSUMv
Time of a problem	Change point	Change point



1. Data input and data table

MSstatsQC



QC data
gathering

Data input and data
processing

Metric
Summary

Control charts
and change
point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Home Data Import and Selection Create Decision Rules Metric Summary Control Charts Help

Data Import Options

Upload your data (Comma-separated (*.csv) QC file format)

To see acceptable example data, look at [Help](#) tab

Upload file

Browse... Study9.1Site

Upload complete

If you want to run **MSstatsQC** with example data file, click this button

Run with example data

Show 25 entries

Search:

Acquired time	Peptide	Annotations	Retention time	Full width at half maximum
9/19/11 13:14	VLVLDTDYK		24.62	0.29
9/19/11 14:45	VLVLDTDYK		24.70	0.31
9/19/11 16:15	VLVLDTDYK		24.53	0.31
9/19/11 17:46	VLVLDTDYK		24.59	0.31
9/19/11 19:16	VLVLDTDYK		24.62	0.30
9/19/11 20:47	VLVLDTDYK		24.76	0.30

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Home Data Import and selection Create decision rules Metric summary Control charts Help

Data Import Options

Select metrics for all further analyses:

☒ Retention time ☒ Total peak area ☐ Full width at half maximum ☐ Peak asymmetry

If you want to select mean and standard deviation yourself select them here. Otherwise choose the guide set button.

☐ Mean and standard deviation estimated by the user
☒ Mean and standard deviation estimated from guide set

Select a guide set to estimate control limits

Lower bound of guide set

1

Upper bound of guide set

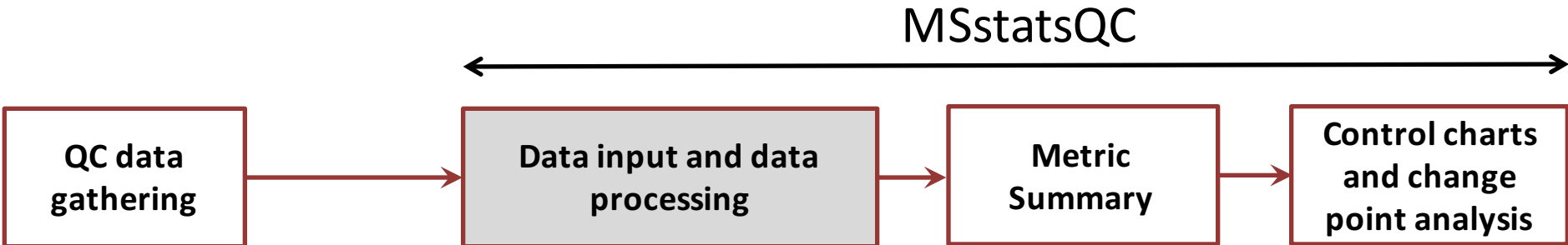
20

Select a precursor or select all

Choose peptide

TAAYVNAIEK

1. Data input and data table



MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Home Data Import and selection **Create decision rules** Metric summary Control charts ▾ Help

Create your decision rule:

RED FLAG

System performance is UNACCEPTABLE when:

1. greater than the selected % of peptides are **out of control** and
2. greater than the selected # of metrics are **out of control**.

% out of control peptides:

out of control metrics:

25

1

YELLOW FLAG

System performance is POOR when:

1. greater than the selected % of peptides are **out of control** and
2. greater than the selected # of metrics are **out of control**.

Warning: The limits should be less than or equal to the RED FLAG limits

% out of control peptides:

of out of control metrics:

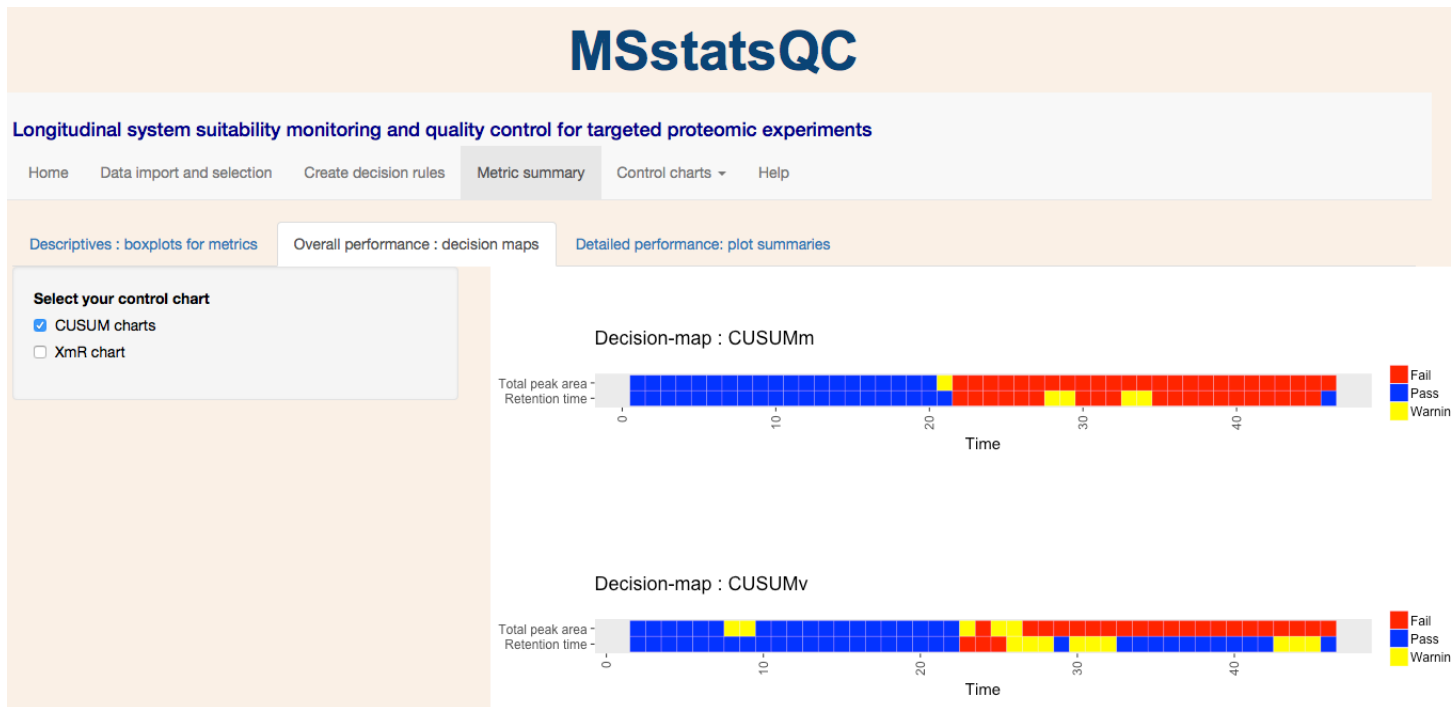
10

1



2. Metric Summary

MSstatsQC



2. Metric Summary

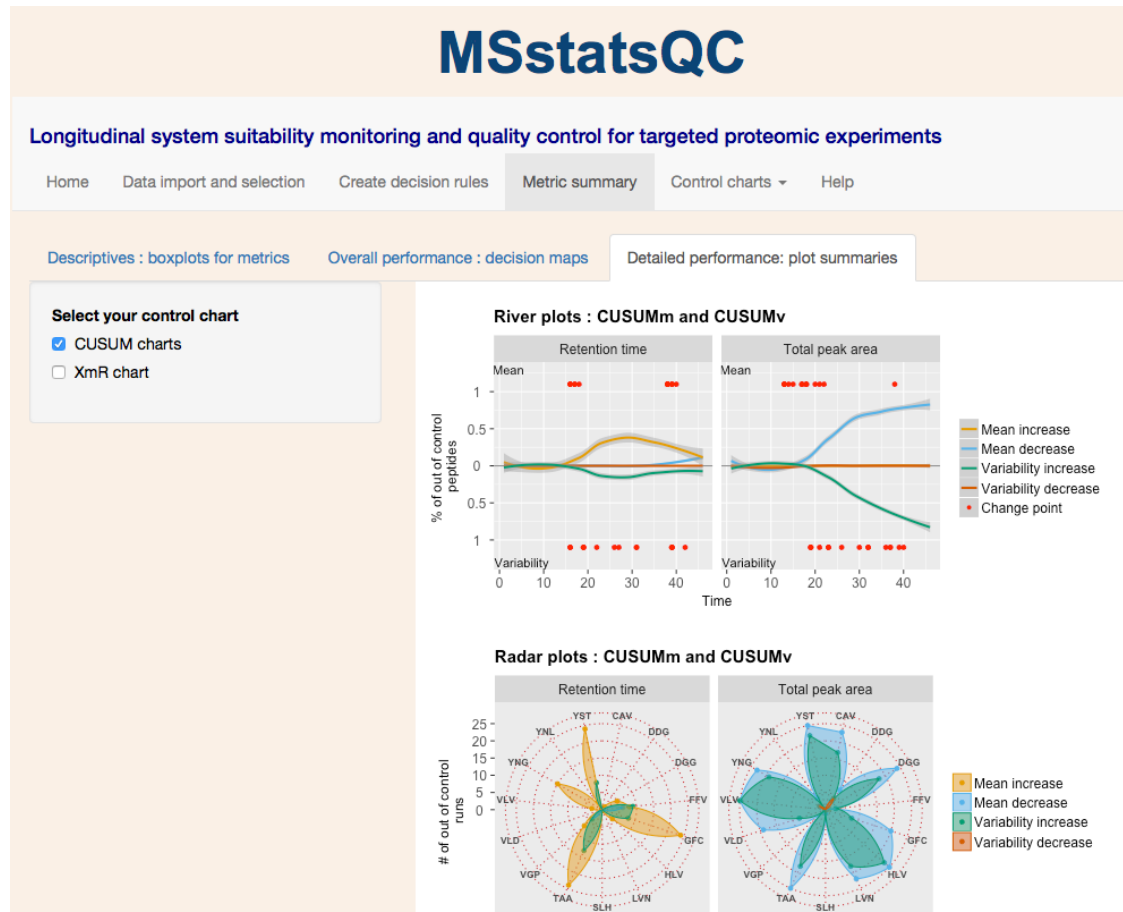
MSstatsQC

QC data
gathering

Data input and data
processing

Metric
Summary

Control charts
and change
point analysis



3. Control charts-Individual (X) and Moving Range (mR)

MSstatsQC



QC data
gathering

Data input and data
processing

Metric
Summary

Control charts
and change
point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

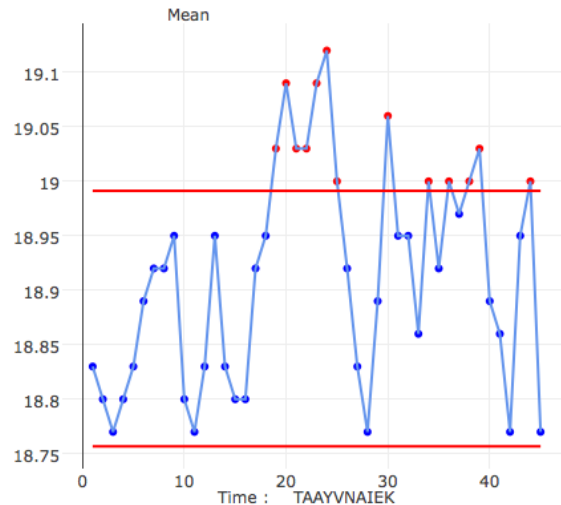
Home Data import and selection Create decision rules Metric summary Control charts Help

Retention time Total peak area

XmR control charts

CUSUMm and CUSUMv control charts

Change point analysis for mean and variability



1. XmR chart is useful when large shifts and isolated outliers exist in the dataset.
2. Analyst are encouraged to go back to their records and investigate the causes of out-of-control observations and try to eliminate it.

3. Control charts-Cumulative Sum (CUSUM)

MSstatsQC



QC data
gathering

Data input and data
processing

Metric
Summary

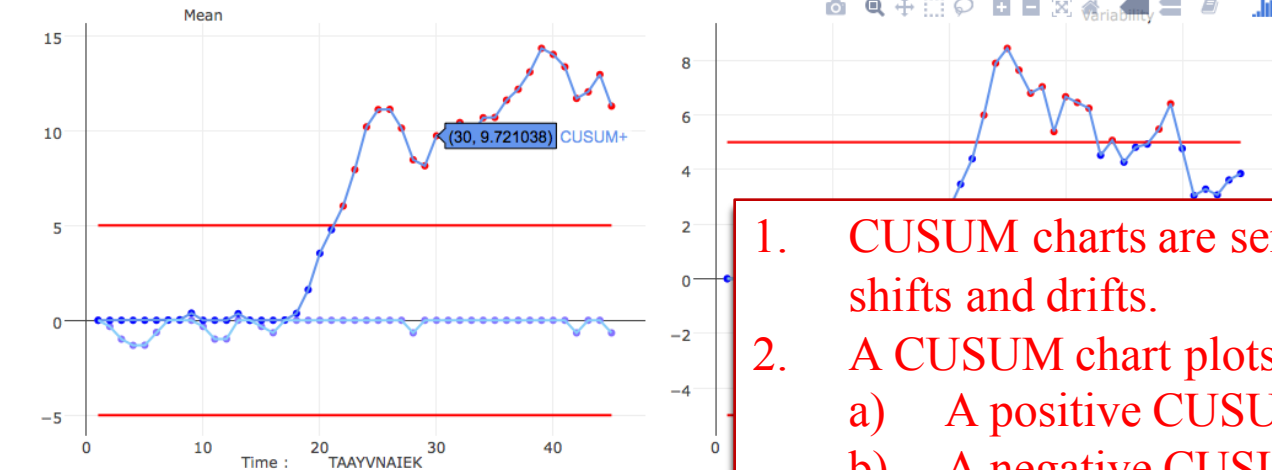
Control charts
and change
point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Home Data import and selection Create decision rules Metric summary Control charts Help

Retention time Total peak area Full width at half maximum Peak Assymetry



1. CUSUM charts are sensitive to small, sustained shifts and drifts.
2. A CUSUM chart plots two statistics:
 - a) A positive CUSUM for increases and
 - b) A negative CUSUM for decreases.

4. Change point analysis

MSstatsQC



QC data
gathering

Data input and data
processing

Metric
Summary

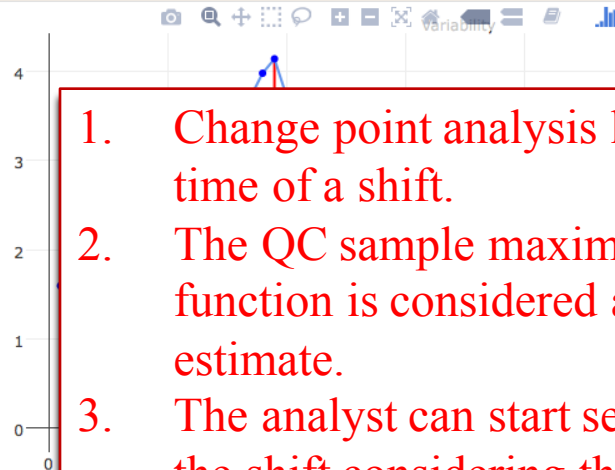
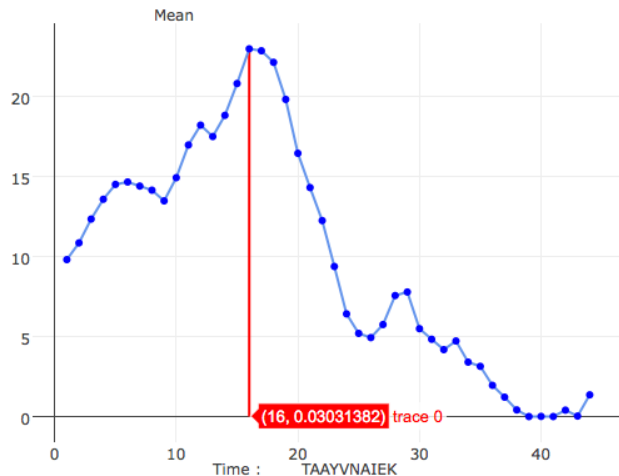
Control charts
and change
point analysis

MSstatsQC

Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Home Data import and selection Create decision rules Metric summary Control charts Help

Retention time Total peak area Full width at half maximum Peak Assymetry



1. Change point analysis help identify the exact time of a shift.
2. The QC sample maximizes the change point function is considered as the change point estimate.
3. The analyst can start searching for the causes of the shift considering this information.

Outline

1. Quality assurance and definition of quality
2. Basics of Statistical Process Control (SPC)
3. MSstatsQC
4. **Case studies from CPTAC study 9.1**



SRM Data : CPTAC Multisite Study 9.1

Mol Cell Proteomics. 2013 Sep;12(9):2623-39.

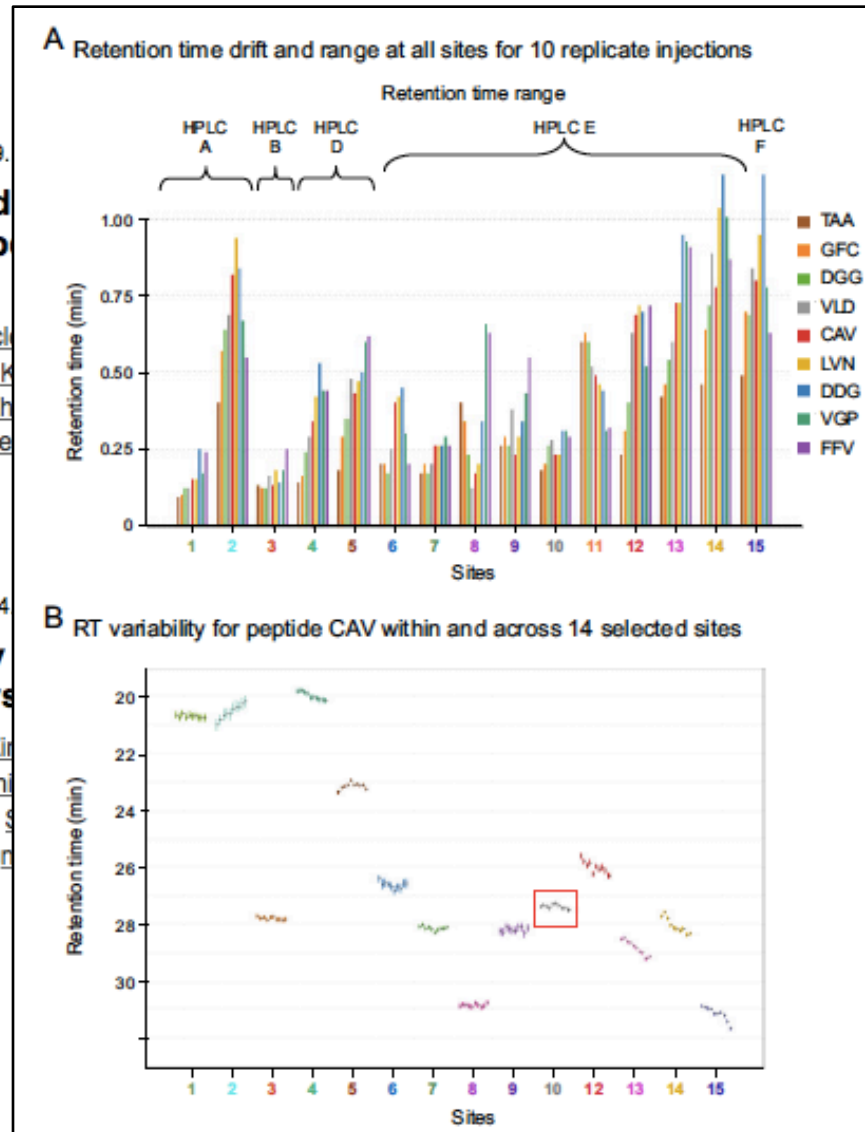
Design, implementation and assessment of instrument performance for the quantitative monitoring-MS (LC-MRM-MS).

Abbatiello SE¹, Mani DR, Schilling B, Mack JM, Hedrick V, Inerowicz HD, Jackson A, K A, Waldemarson S, Whitwell CA, You J, Zh Neubert TA, Paulovich A, Regnier F, Skates

Mol Cell Proteomics. 2015 Sep;14(9):2357-74

Large-Scale Interlaboratory Quantitative Peptide Assays

Abbatiello SE¹, Schilling B², Mani DR¹, Zin JM², Inerowicz HD⁷, Jackson A⁸, Keshishi N¹³, Shaddox K³, Skates SJ¹⁴, Kuhn E¹, S MJ⁵, Neubert TA¹⁰, Paulovich AG¹⁵, Regnier



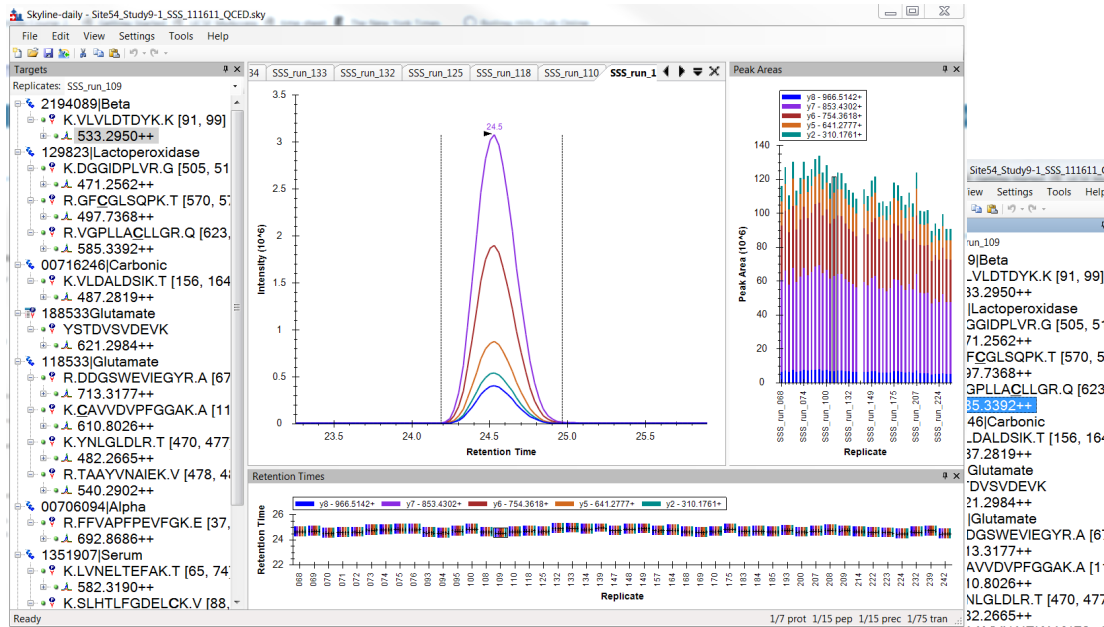
Method for the quantitative monitoring-MS (LC-

Allen S, Dodder NG, Ghosh M, Held
Box K, Smith D, Tomazela D, Wahlander
Gibson BW, Liebler D, MacCoss M,

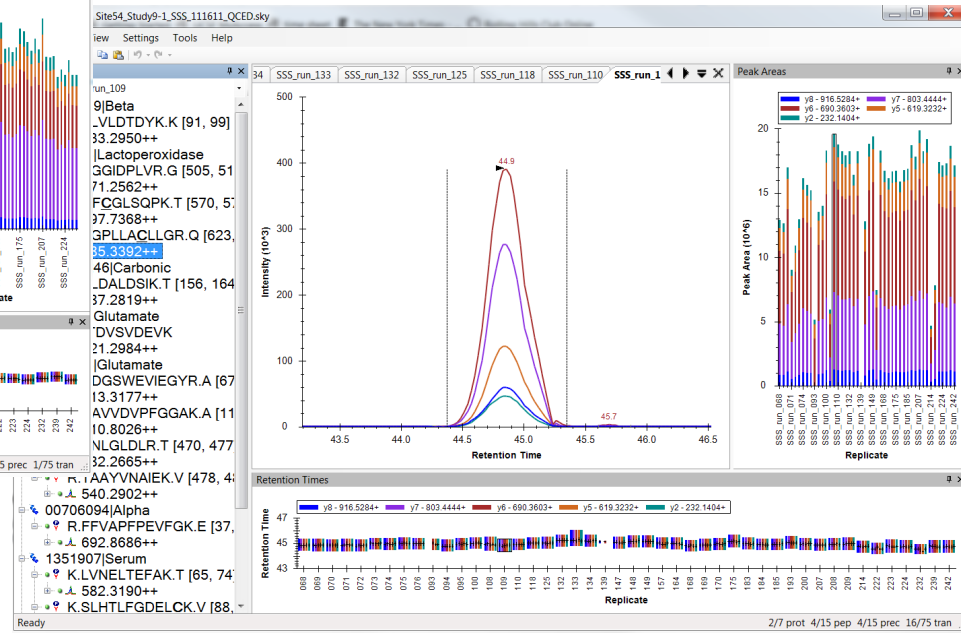
Highly Multiplexed,

usack MP², Gosh M⁶, Hedrick V⁷, Held
Kudnick P¹², Sadowski P¹⁰, Sedransk
Gibson BW², Liebler DC³, MacCoss

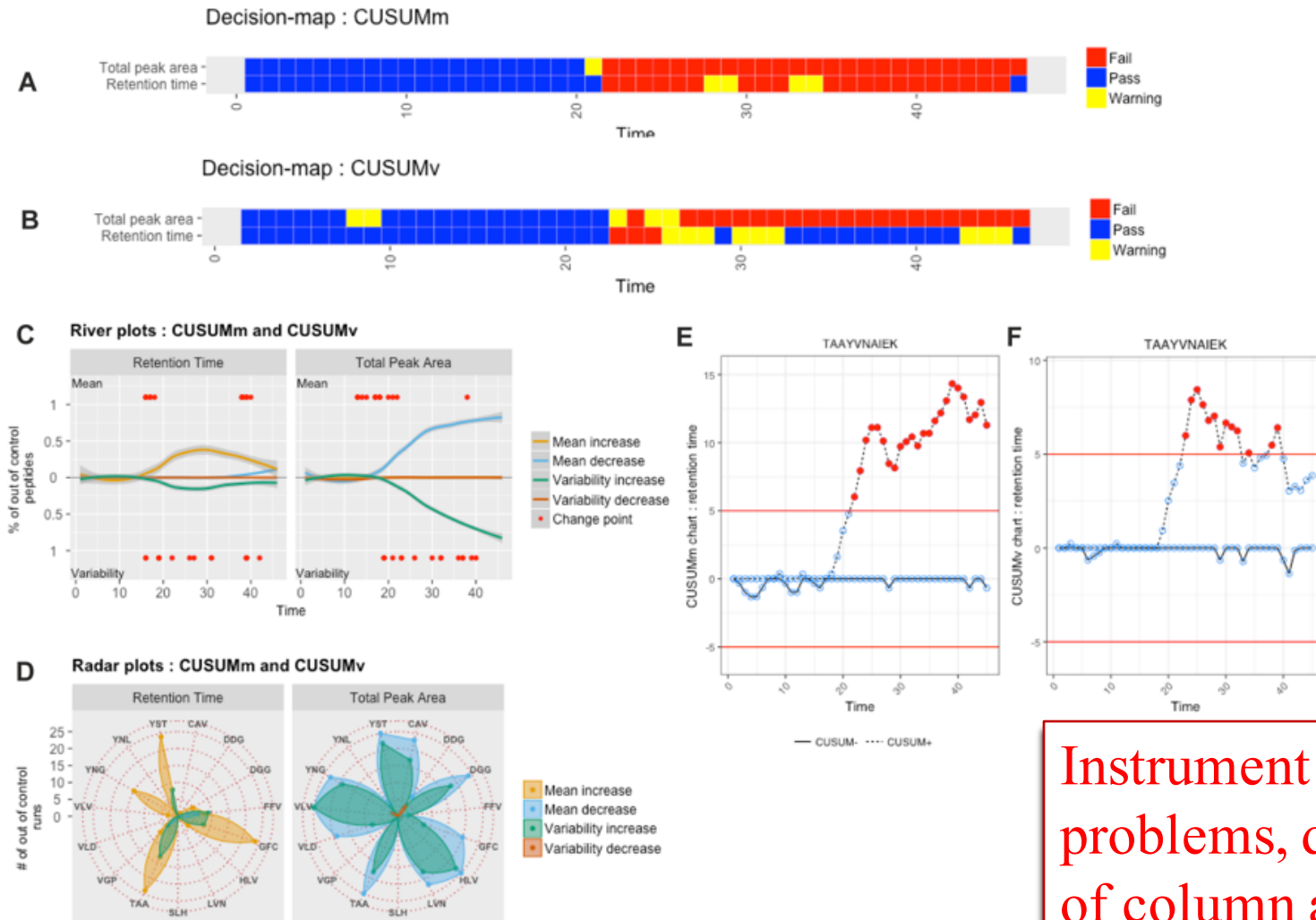




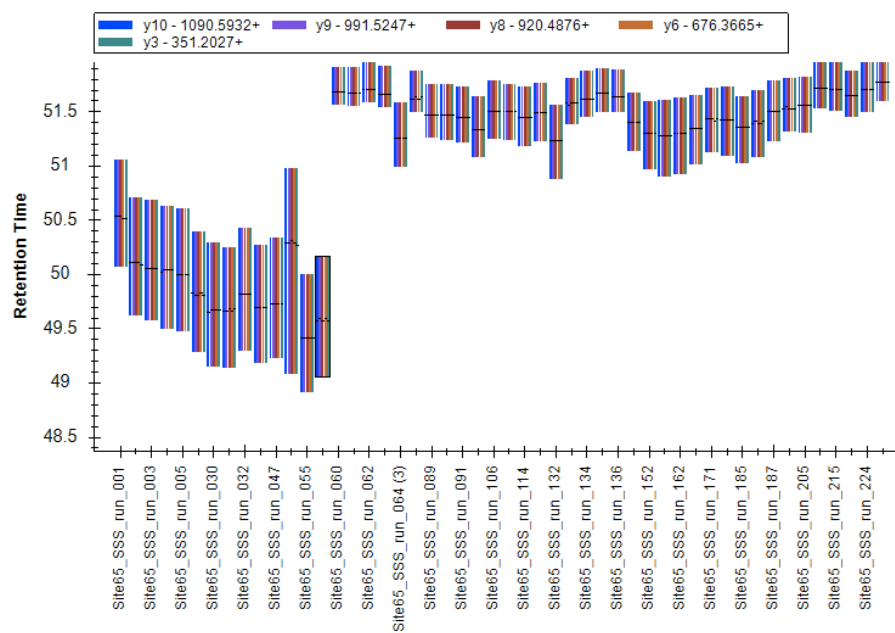
Very nice SST for some peptides



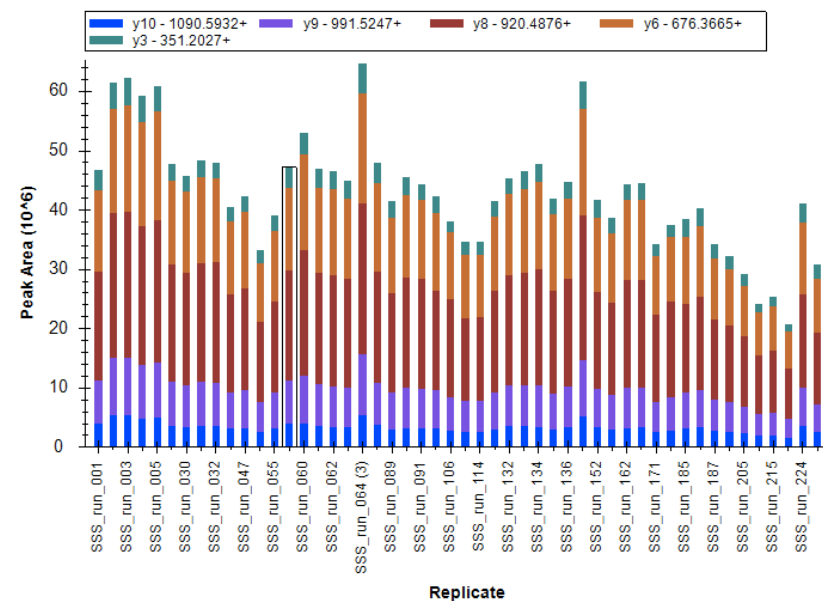
Changes in retention time for some peptides

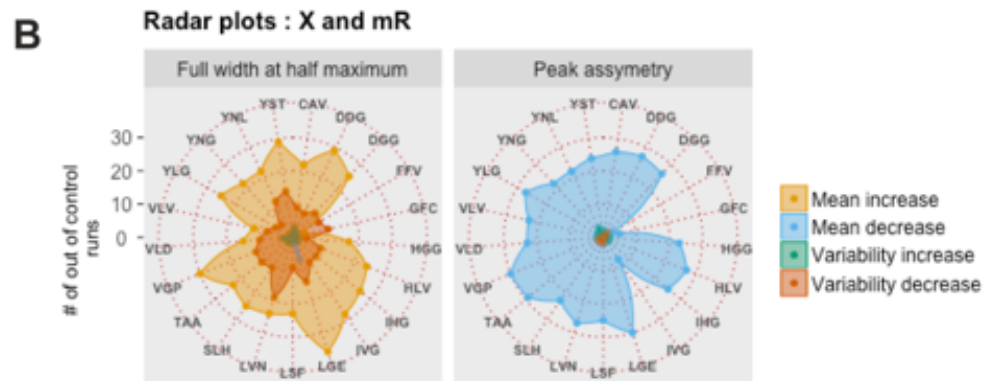
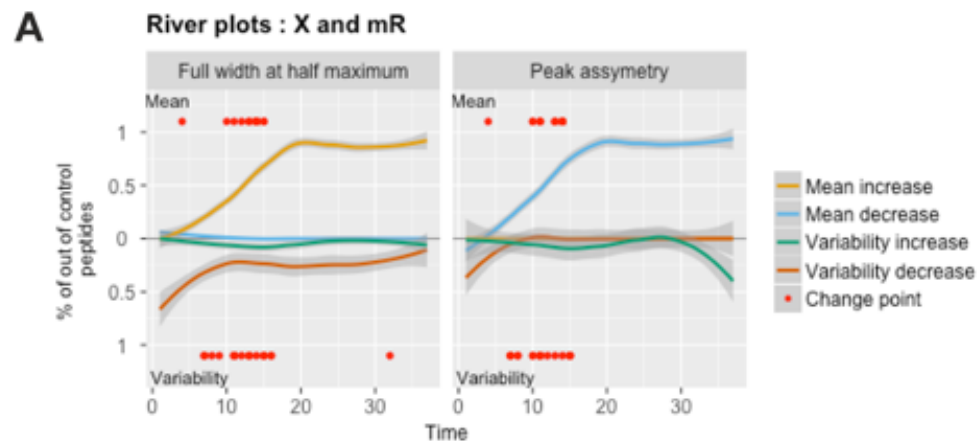


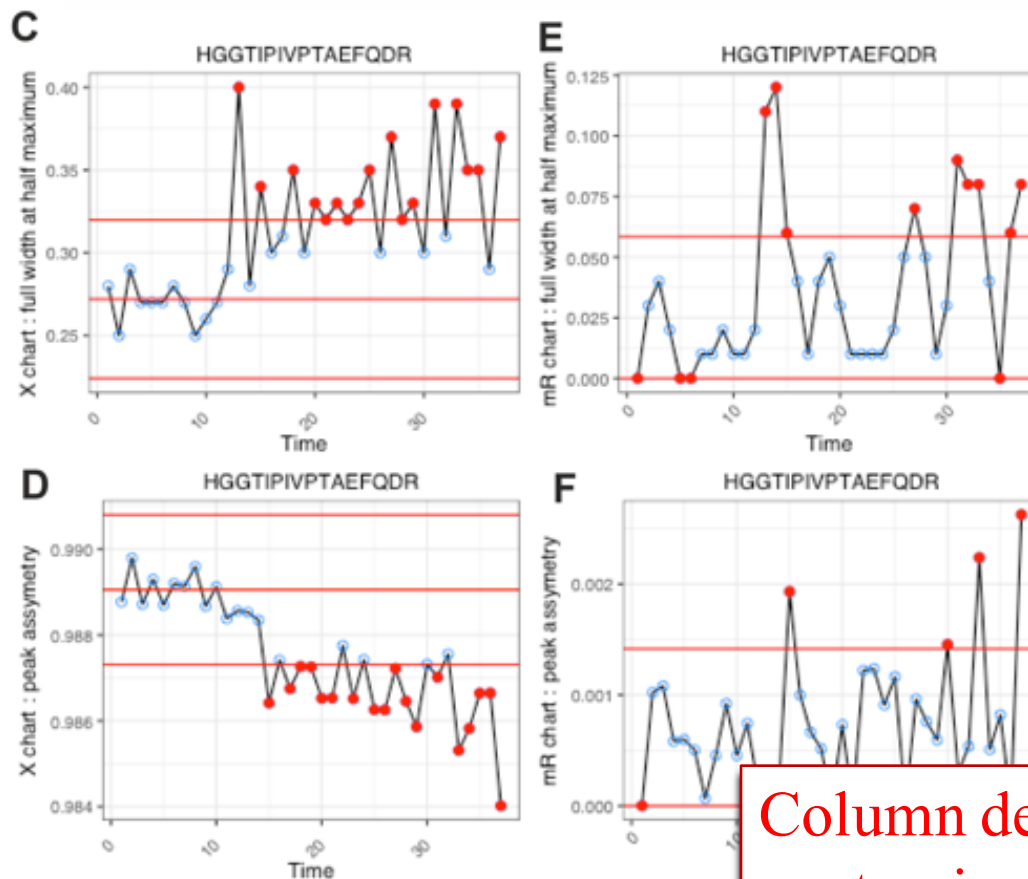
Instrument calibration problems, deterioration of column and emitter and wear in parts



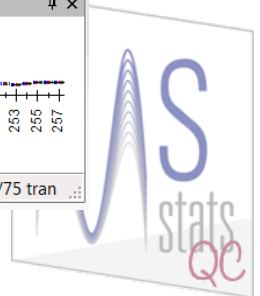
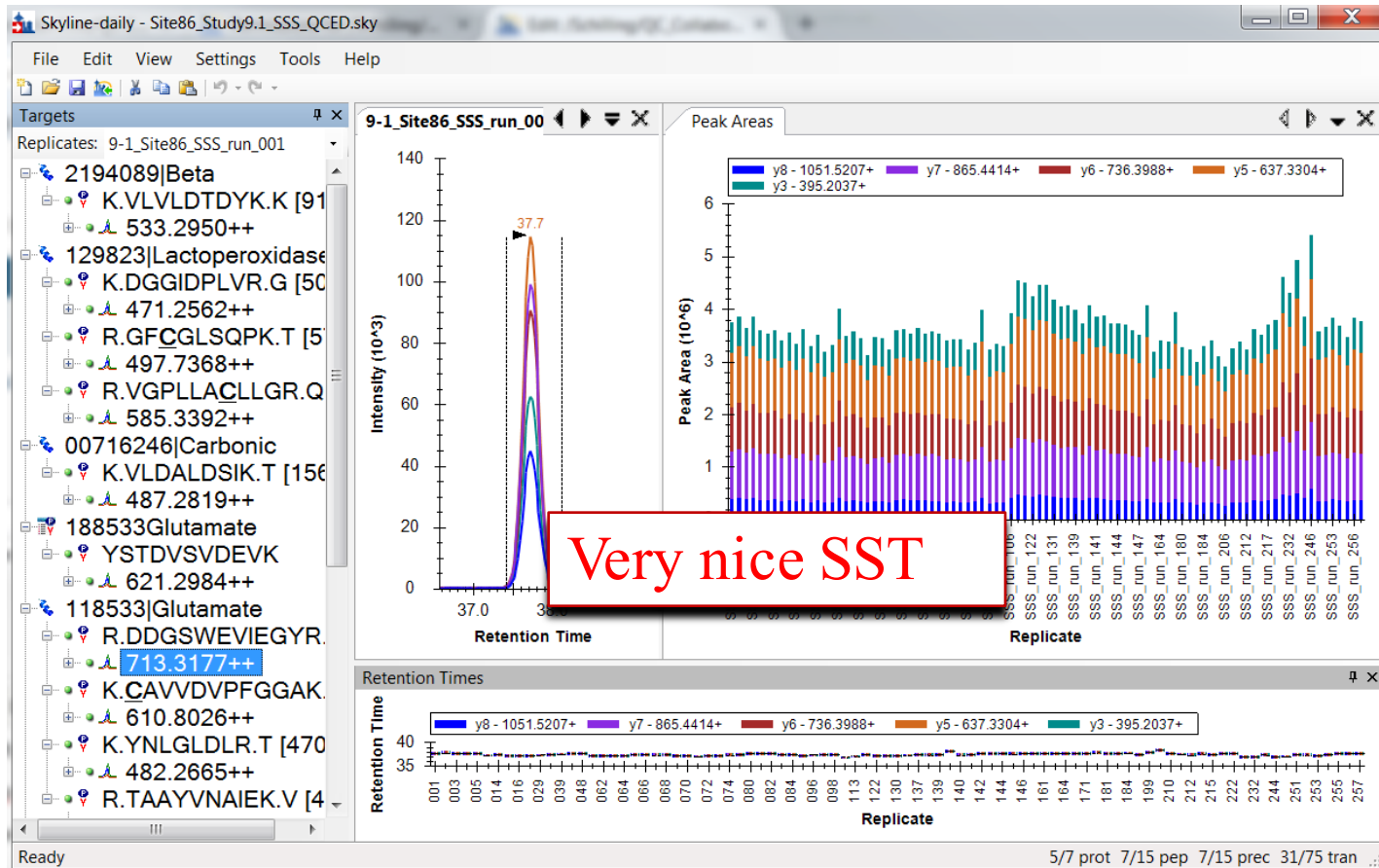
Peak Areas

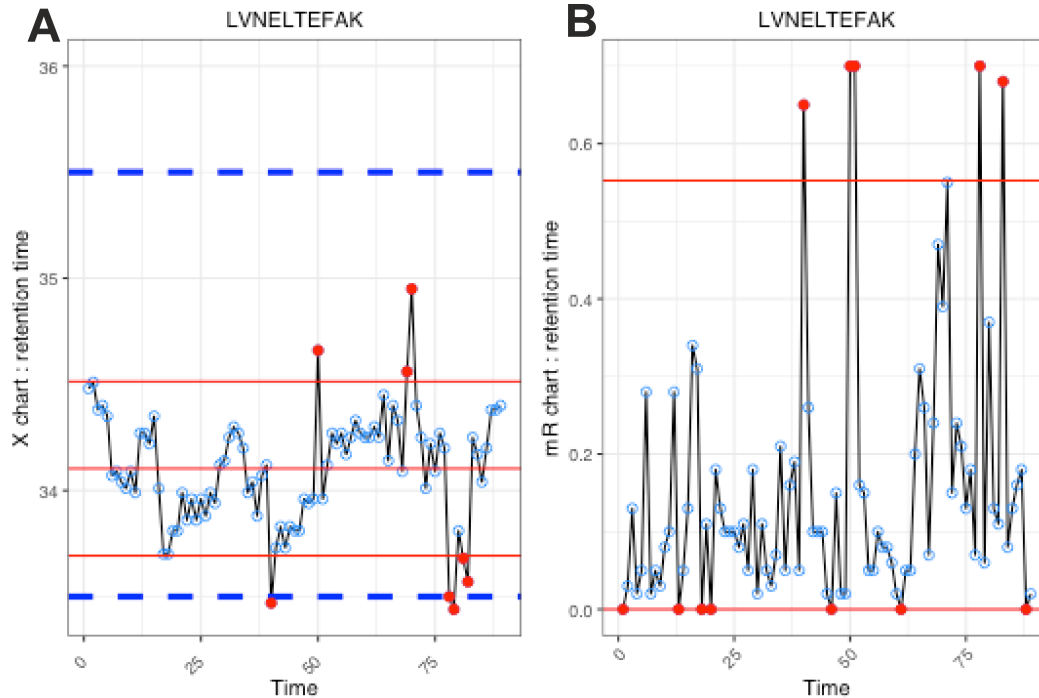




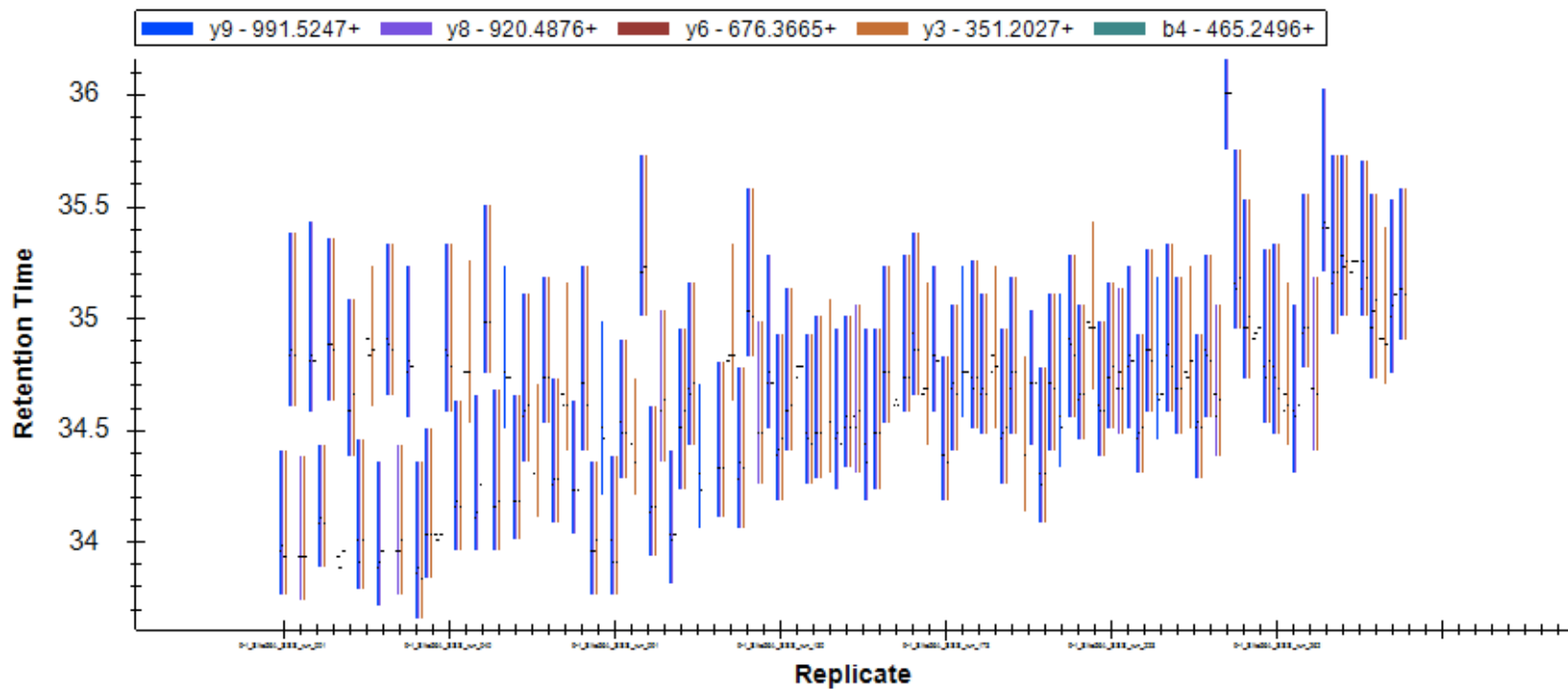


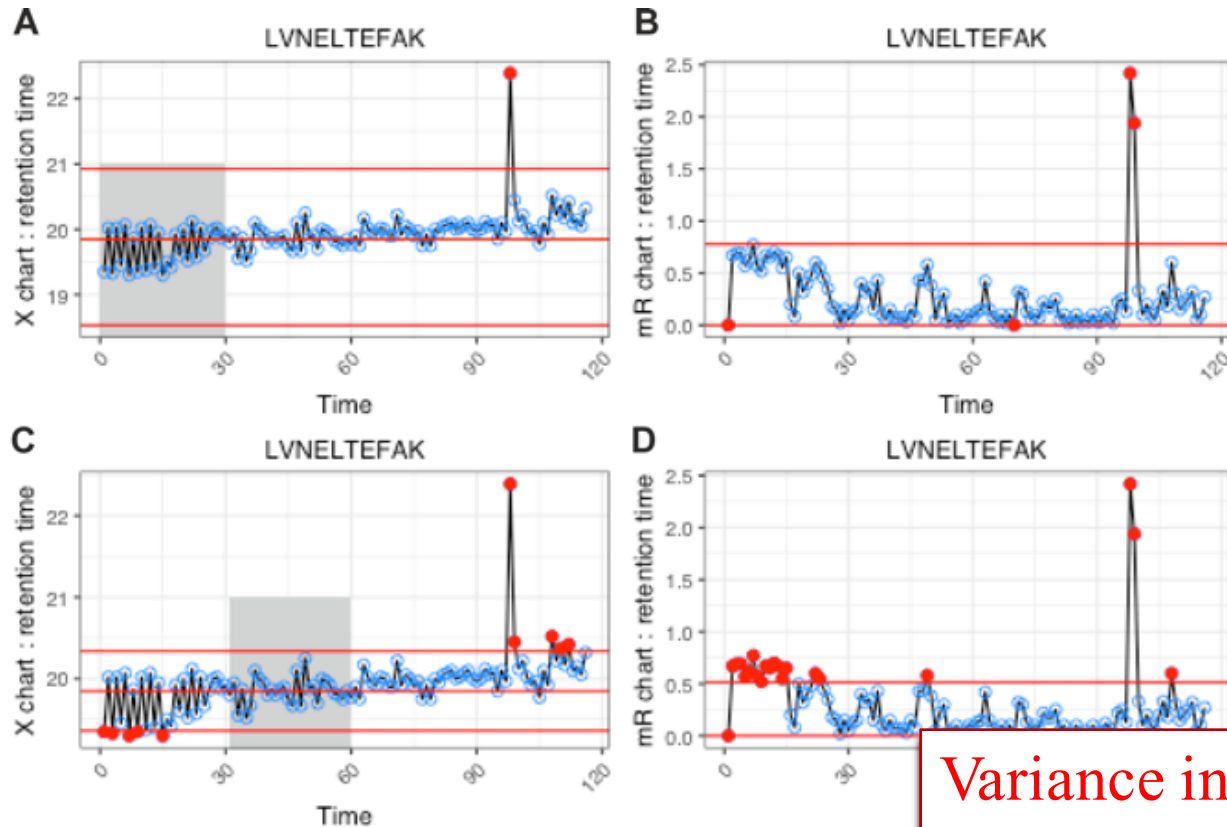
Column deterioration,
contamination, and fitting
fatigue





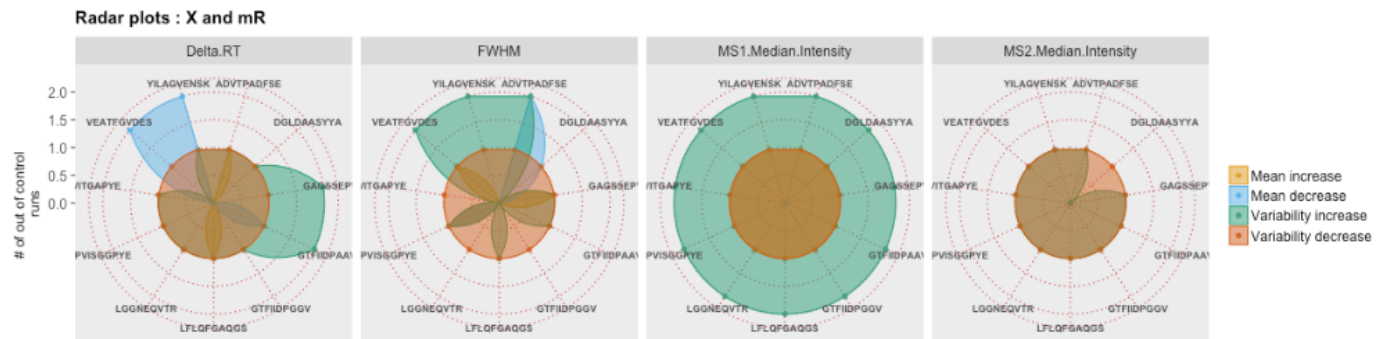
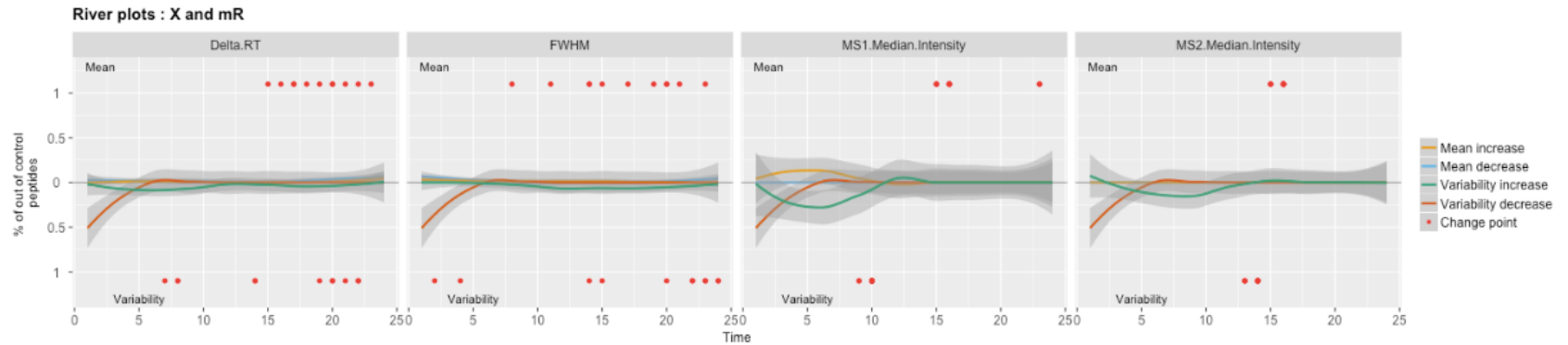
Environmental factors
such as temperature or
pressure changes





Variance inflation
could occur during the
equilibration phase of
a new LC column

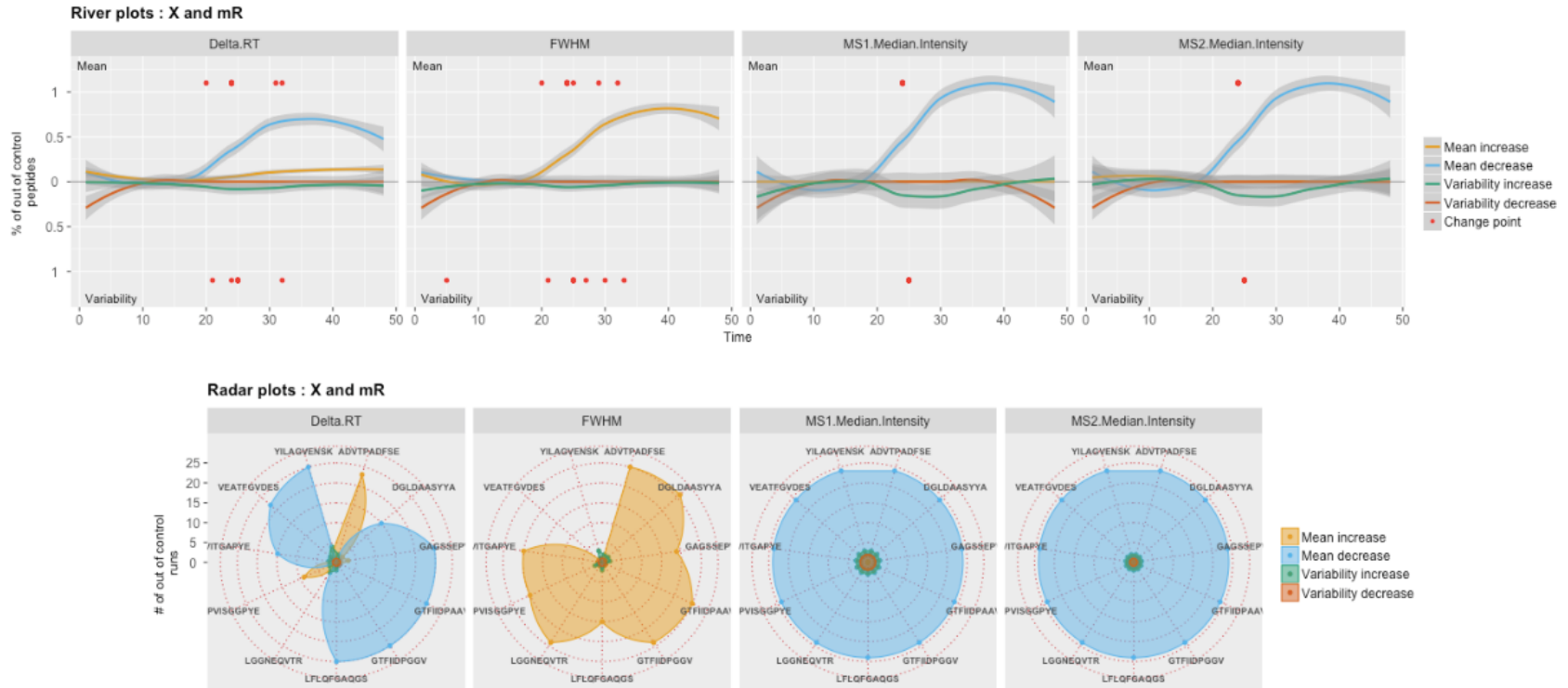
DDA Data: iRT peptides



Most of the peptides are stable and system is acceptable



DIA Data: iRT peptides



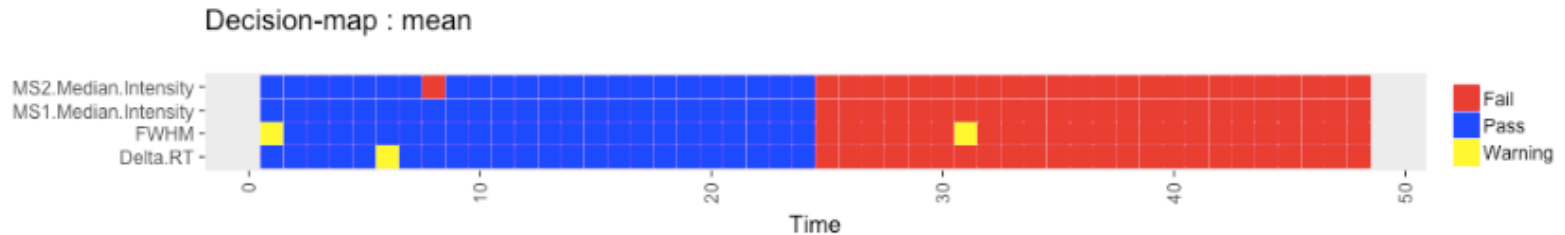
Most of the peptides are unstable and system is unacceptable due to changes in all metrics monitored.

Mean level of FWHM increases over time while mean levels of retention time, MS1 and MS2 intensities decreases.

Change point is expected to be around 20th run.



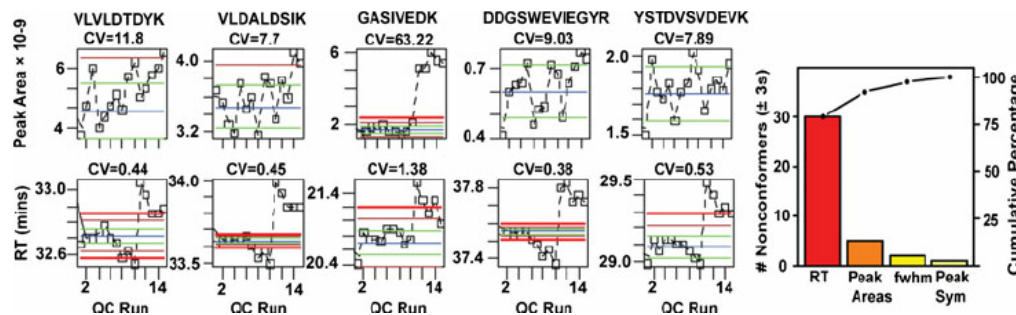
DIA Data: iRT peptides



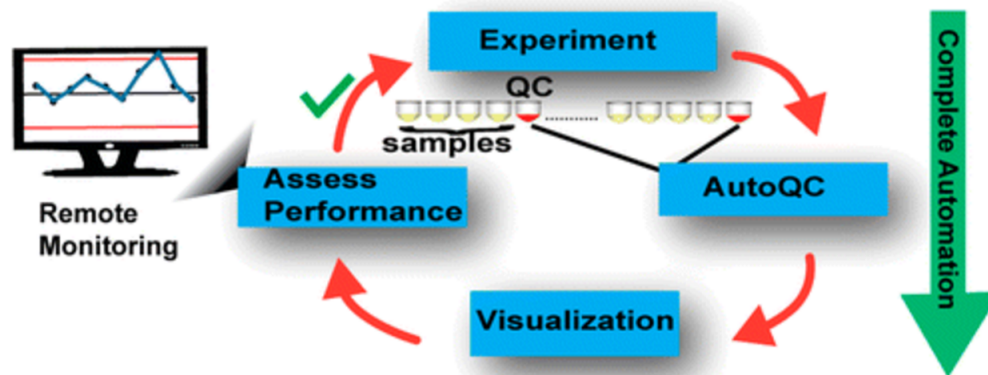
System is not able to pass predefined performance criteria (decision-map)



SPC applied to mass spectrometry proteomics



SProCop



Panorama
AutoQC

Bereman et. al. (2014) *J. Am. Soc. Mass Spectrom*
Bereman et. al. (2016) *J. Proteome Res.*

SPC applied to mass spectrometry proteomics

QC Summary

47 sample files

7 precursors

AutoQC ○

- 2013/08/27 14:45:49 - 3/56 (Levey-Jennings), 1/56 (Moving Range), 1/56 (CUSUMv) outliers
- 2013/08/27 03:19:45 - 5/56 (Levey-Jennings), 8/56 (Moving Range), 2/56 (CUSUMv) outliers
- 2013/08/26 04:27:53 - 6/56 (Levey-Jennings), 5/56 (Moving Range), 1/56 (CUSUMv) outliers

QC Plots

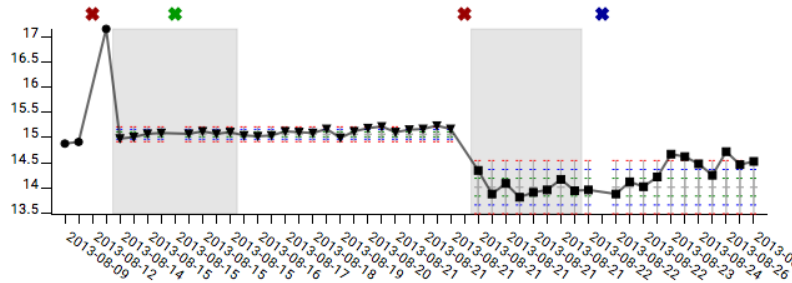
Metric: Retention Time Date Range: All dates

Plot Size: ☒ Small ☐ Large | QC Plot Type: ☒ Levey-Jennings ☐ Moving Range ☒ CUSUMm ☐ CUSUMv

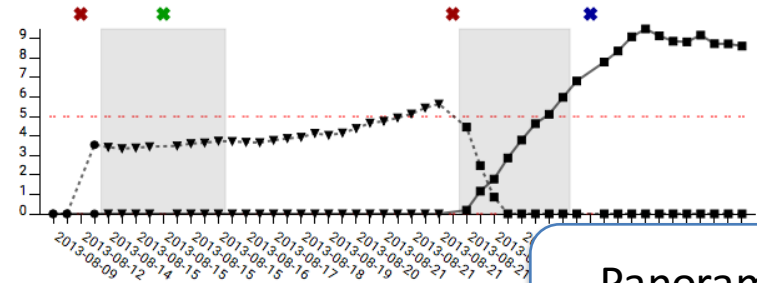
Y-Axis Scale: Linear | ☐ Group X-Axis Values by Date ☐ Show All Series in a Single Plot [VIEW LEGEND](#)

ATEEQLK


Levey-Jennings



CUSUMm



Panorama
AutoQC



MSstats

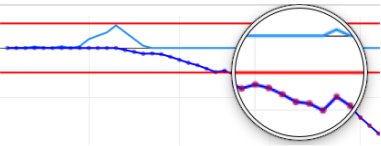
Statistical Tool For Quantitative Mass Spectrometry-Based Proteomics

HOME INSTALLATION WORKFLOWS DATASETS **MSSTATSQC** NEWS CONTACT

MSSTATSQC

Longitudinal system suitability monitoring tools for quantitative mass spectrometry based proteomic experiments

Statistical process control (SPC) is a general and well-established method of quality control (QC) which can be used to monitor and improve the quality of a process such as LC MS/MS. 'MSstatsQC' is an open-source R-based web application for statistical analysis and monitoring of quality control and system suitability testing (SST) samples produced by spectrometry-based proteomic experiments. Our framework termed 'MSstatsQC' is available through <http://www.msstats.org/msstatsqc>. It uses SPC tools to track ID free system suitability metrics including total peak area, retention time, full width at half maximum (FWHM) and peak asymmetry for selection reaction monitoring (SRM) based proteomic experiments. We introduce simultaneous and time weighted monitoring tools and change point analysis to monitor mean and dispersion of system suitability metrics such as retention time. Proposed longitudinal monitoring approach significantly improves the ability of



- News about MsstatsQC
- Example datasets
- Related publications

srtaheri / msstats-qc

Watch 3 Star 0 Fork 0

Code Issues 0 Pull requests 0 Pulse Graphs

No description or website provided.

128 commits 1 branch 1 release 1 contributor

Branch: master New pull request Find file Clone or download

srtaheri order of summary and radar plots changed ... Latest commit c5eebcc 2 days ago		
Datasets	fake data included	19 days ago
UserManual	manual updated	a month ago
rsconnect/shinyapps.io/lcms	shiny republished	a month ago
www	unrelated movie is deleted from www folder	a month ago
.gitignore	session info removed	2 months ago
QCMatrix.R	codes of summary and radar plots changed	2 days ago

- MSstatsQC daily
- Example datasets
- R shiny codes and functions

Acknowledgements

Northeastern University

- Olga Vitek
- Sara Taheri
- Meena Choi

University of Washington

Brendan MacLean

North Carolina State University

Michael Bereman

Buck Institute

Birgit Schilling

Thermo Fisher

Sue Abbatiello



Related work

MCP MOLECULAR & CELLULAR
PROTEOMICS

QUICK SEARCH Author: Keyword: Year: Vol: Page: [Advanced Search]

Home | Current issue | Papers in Press | Archive | Reviews | Editorials | Special Issues

 Check for updates 

MSstatsQC: Longitudinal system suitability monitoring and quality control for targeted proteomic experiments

Eralp Dogu¹, Sara Mohammad-Taheri², Susan E Abbatiello³, Michael S Bereman⁴, Brendan MacLean⁵, Birgit Schilling⁶ and Olga Vitek^{2,*}

 Author Affiliations

 * Corresponding author; email: o.vitek@neu.edu

Abstract

Selected Reaction Monitoring (SRM) is a powerful tool for targeted detection and quantification of peptides in complex matrices. An important objective of SRM is to obtain peptide quantifications that are (1) suitable for the purpose of the investigation, and (2) reproducible across laboratories and runs. The first objective is achieved by system suitability tests (SST), which verify that mass spectrometric instrumentation performs as specified. The second objective is achieved by quality control (QC), which provides in-process quality assurance of the sample profile. A common aspect of SST and QC is the longitudinal nature of the data. Although SST and QC have received a lot of attention in the proteomic community, the currently used statistical methods are fairly limited. This manuscript improves upon the statistical methodology for SST and QC that is currently used in proteomics. It adapts the modern methods of longitudinal statistical process control, such as simultaneous and time weighted control charts and change point analysis, to SST and QC of SRM experiments, discusses their advantages, and provides

This Article

First Published on May 8, 2017,
doi: 10.1074/mcp.M116.064774
mcp.M116.064774.

» Abstract Free
Full Text (PDF) Free
Supplemental Data

☐ Classifications

Research

☐ Services

- Email this article to a friend
- Alert me when this article is cited
- Alert me if a correction is posted
- Alert me when eletters are published
- Similar articles in this journal
- Similar articles in PubMed
- Download to citation manager
- Request Permissions

☐ Responses

Submit a Letter to the Editor
No responses published

☐ Google Scholar

Articles by Dogu, E.
Articles by Vitek, O.

☐ PubMed

☐ Social Bookmarking

Current Issue

May 2017, 16 (5)



Alert me to new issues of
Molecular & Cellular Proteomics

Authors

Frequently Asked Questions (FAQs)

Submit

Subscribers

About the Journal

Editorial Board

Copyright Permissions

RSS and Alerts

Advertise

