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

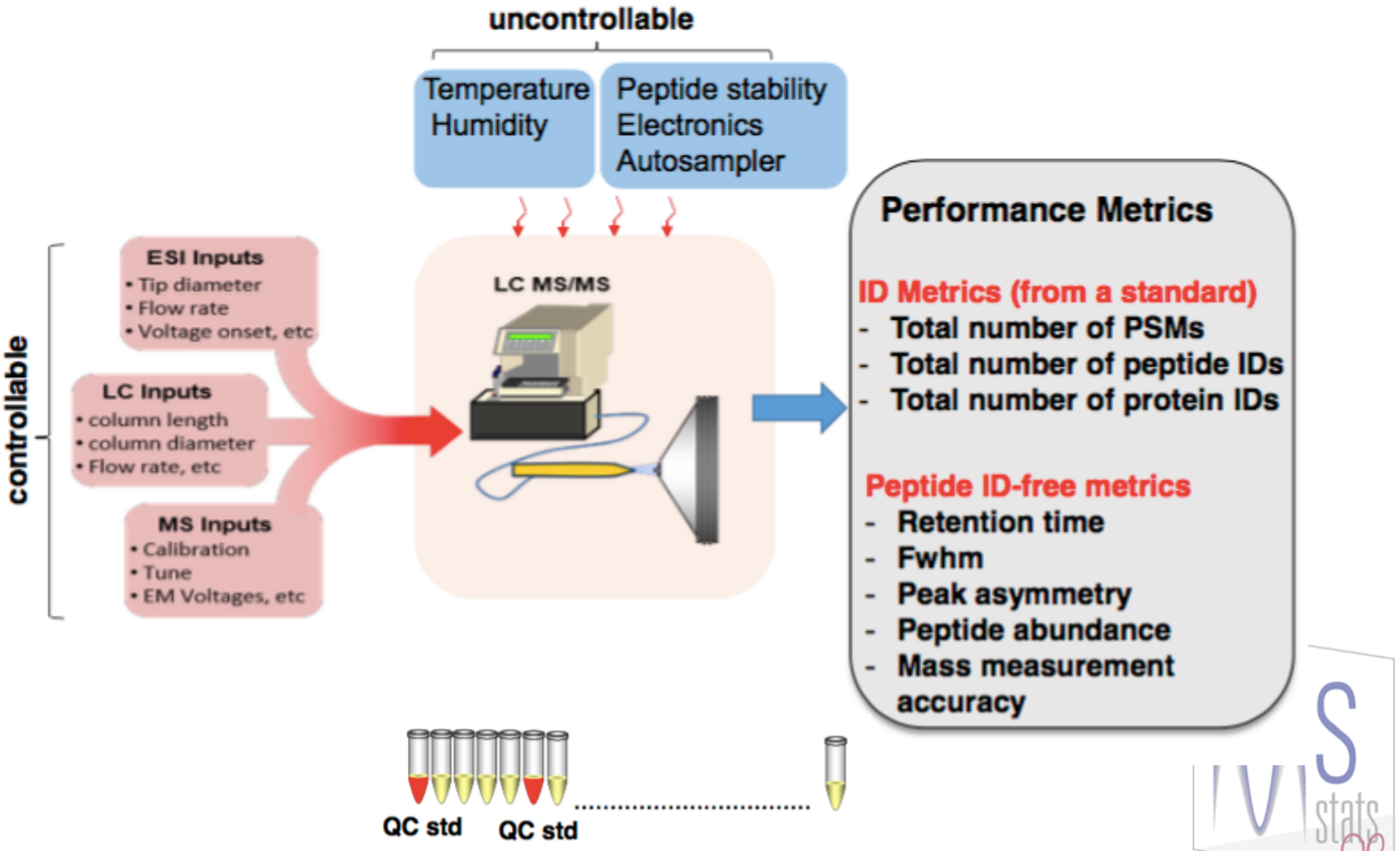
Dr. Eralp DOGU
Mugla University, TR
Prof. Olga Vitek Group
Northeastern University

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!



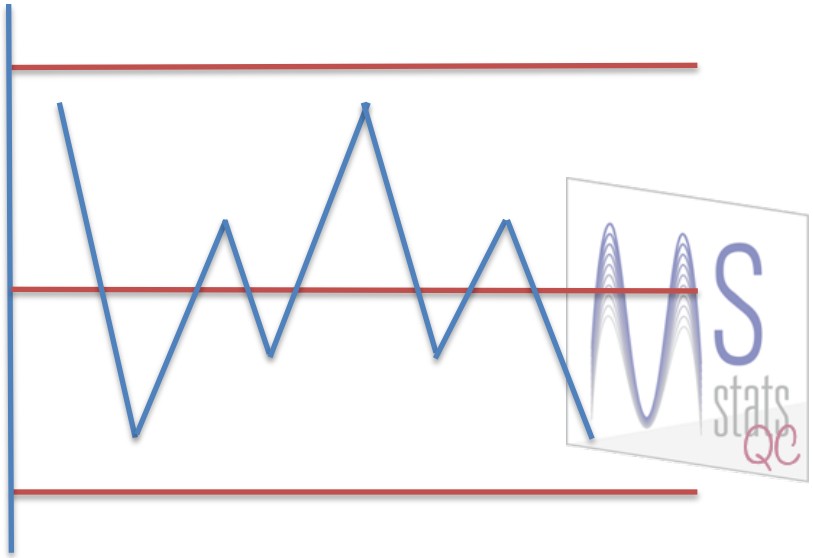
Outline

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

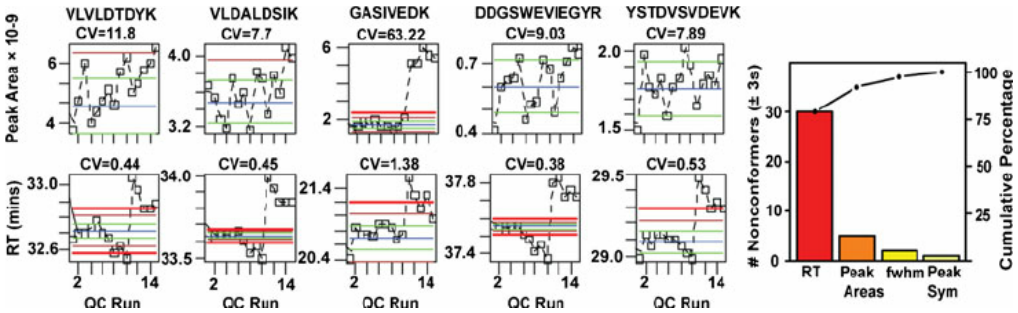
Statistical Process Control (SPC)

- 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

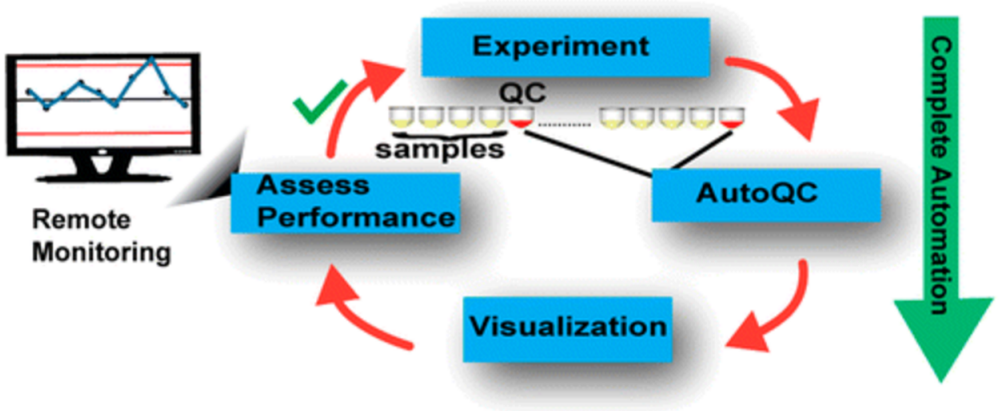
- A key SPC tool is the control chart, which is the focus of this presentation
  - Combines time-series analysis with graphical representation of data



SPC applied to mass spectrometry proteomics



SProCop



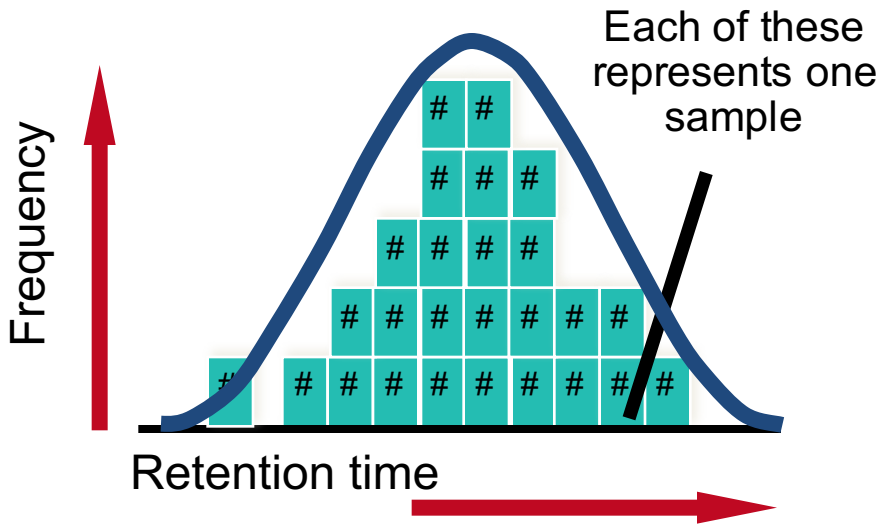
Panorama AutoQC

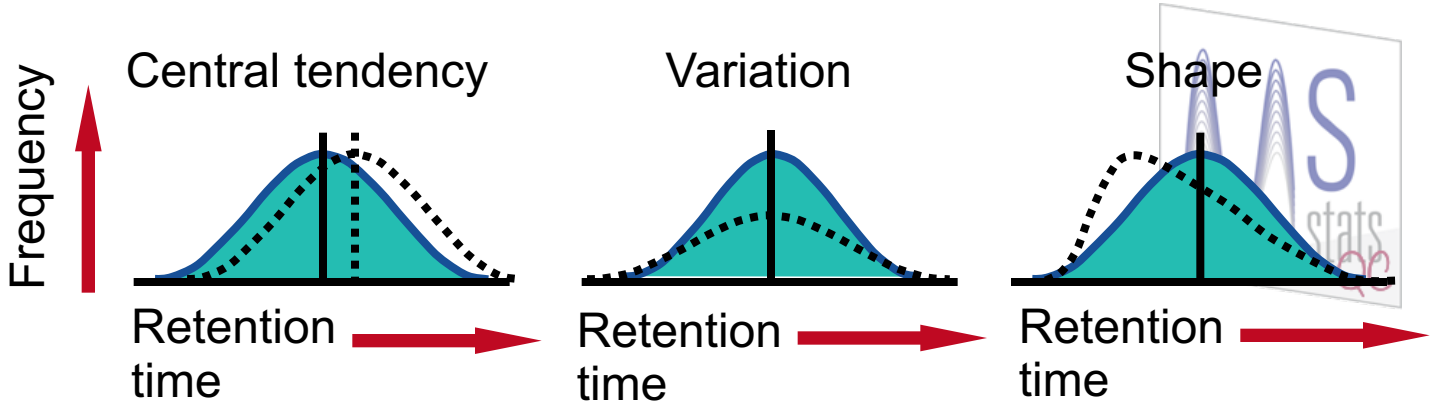
Bereman et. al. (2014) *J. Am. Soc. Mass Spectrom*
Bereman et. al. (2016) *J. Proteome Res.*

Sampling to set up a program

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

a) QC or system suitability samples, vary from each other in terms of retention time
b) After enough samples (guide set) are taken from a stable process, they form a pattern called a distribution
c) 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

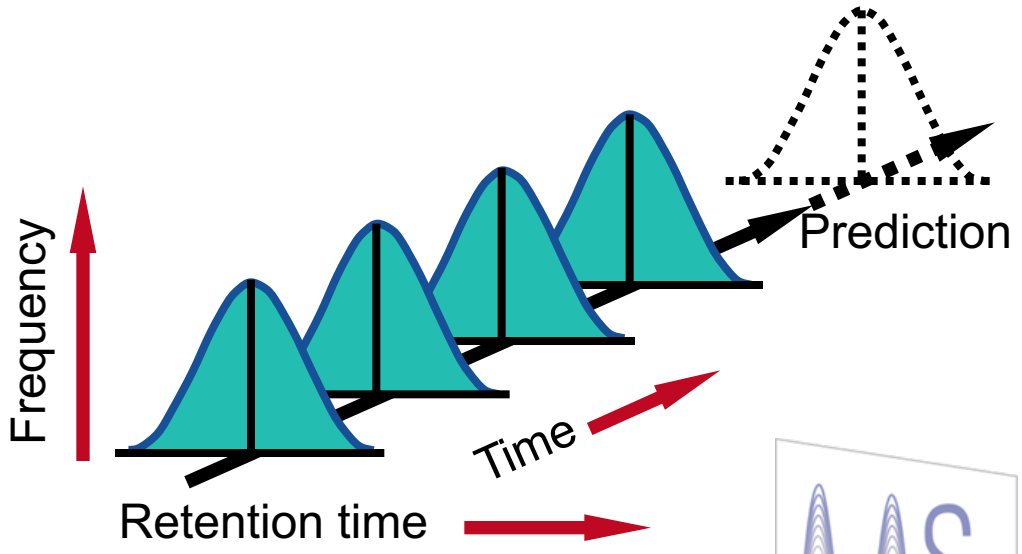




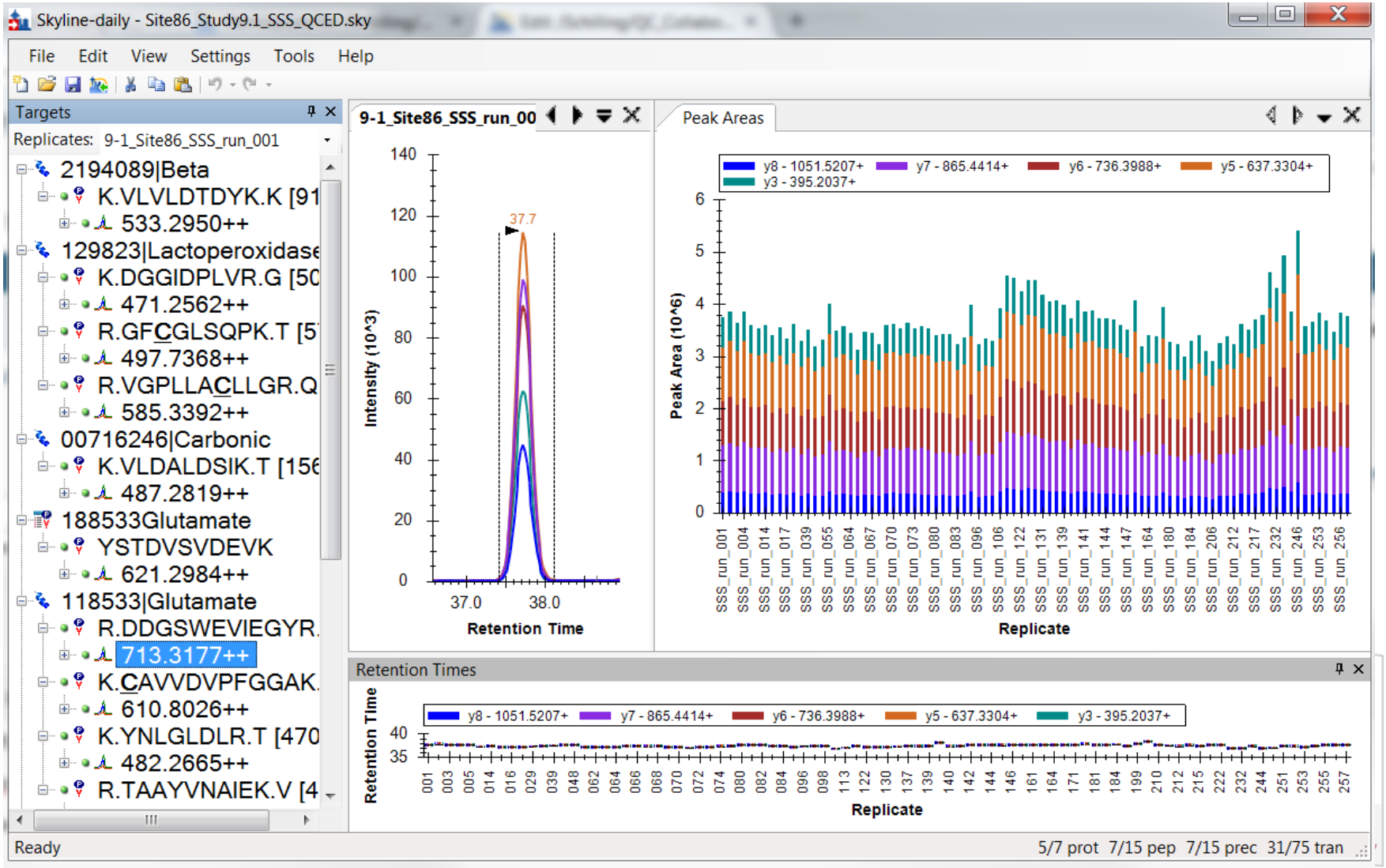
Sampling to set up a suitability program

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



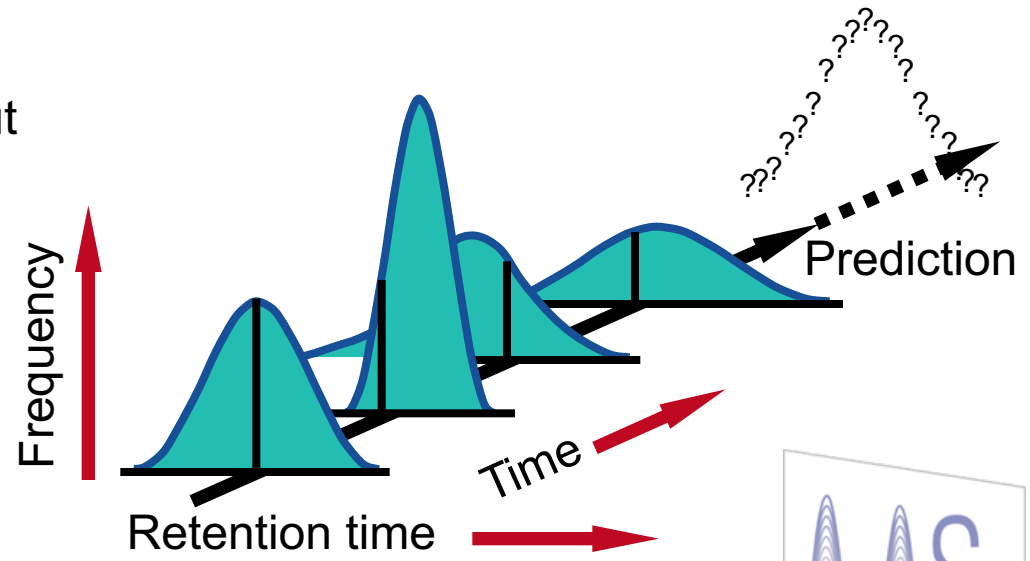
Data : CPTAC Study 9.1 Site 86



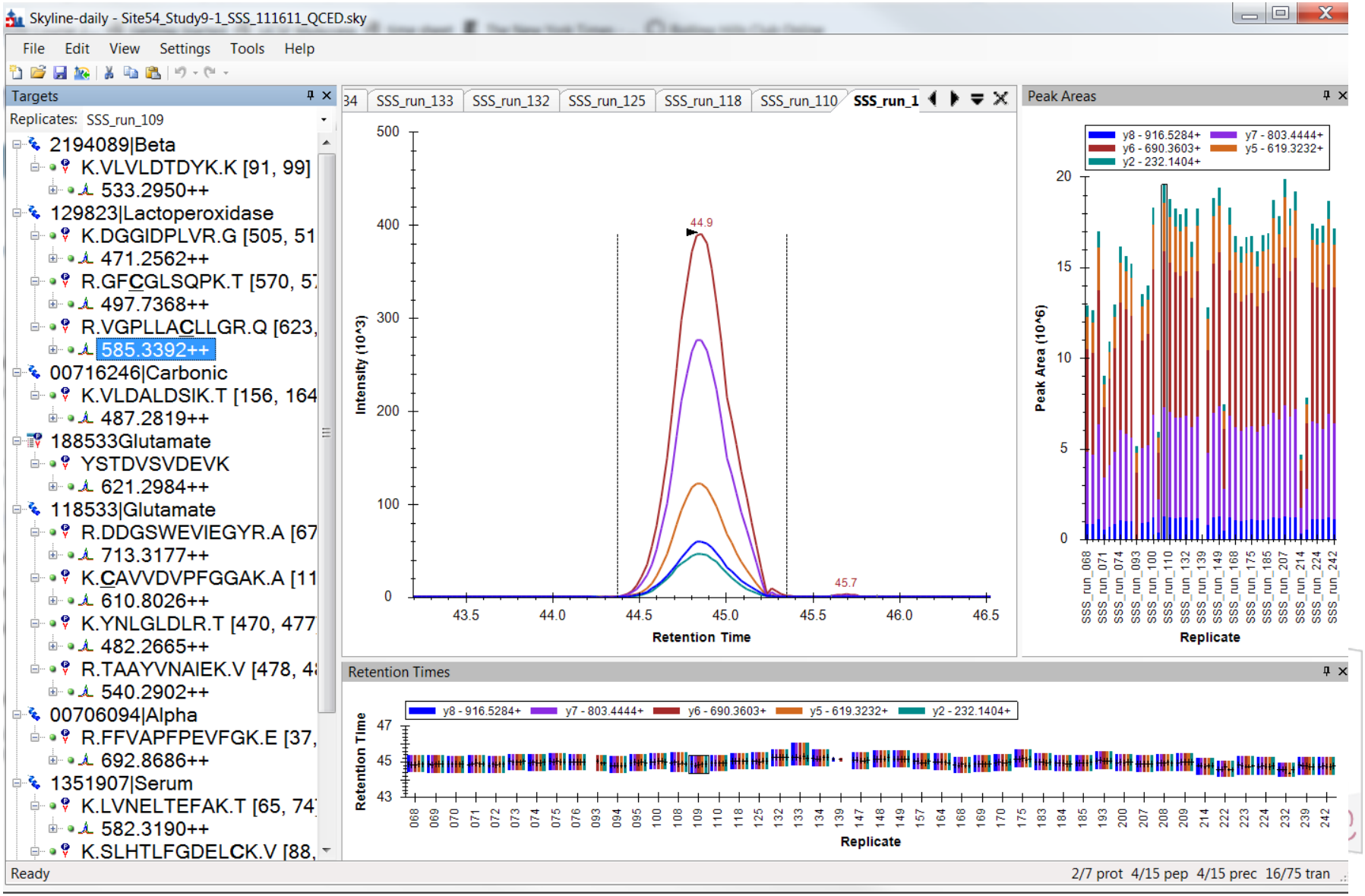
Sampling to set up a suitability program

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 predicable



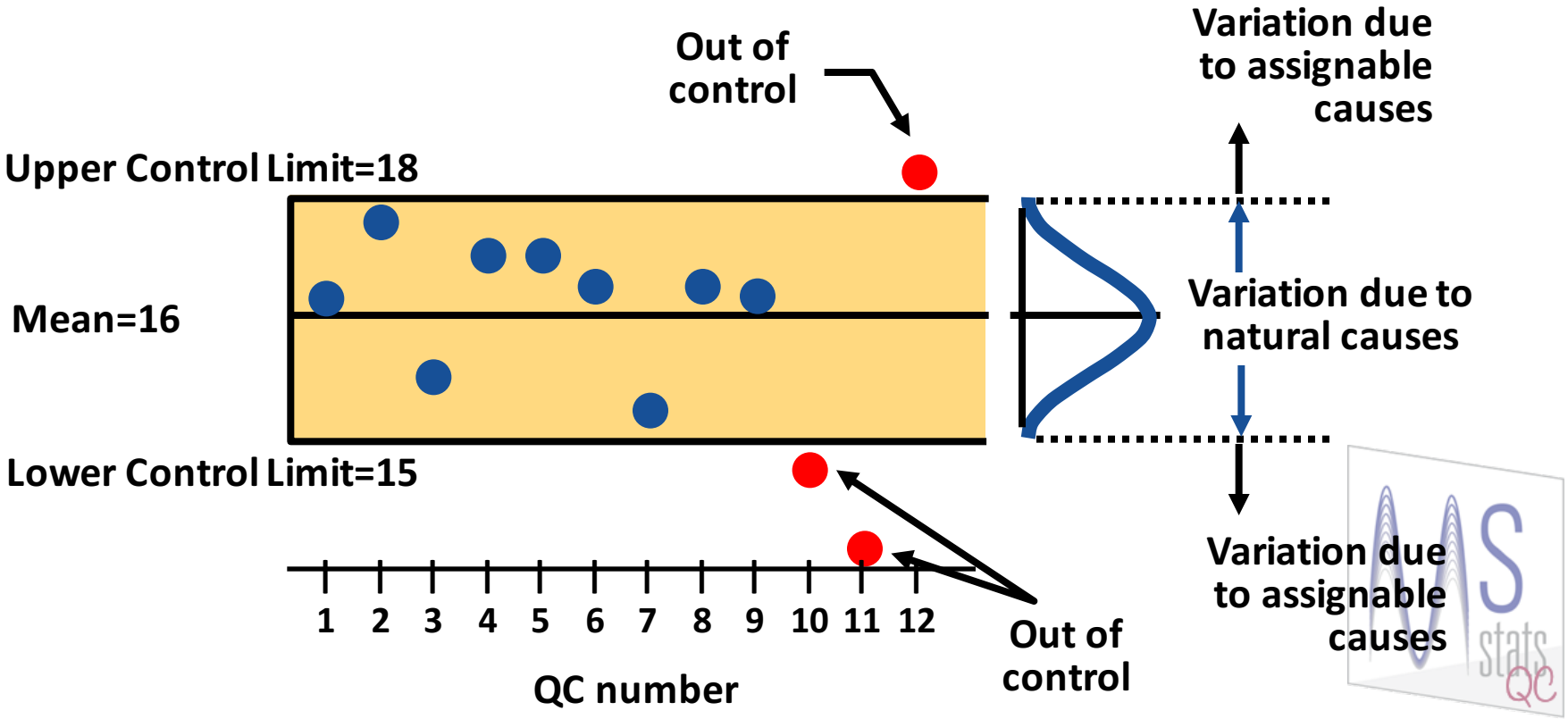
Data : CPTAC Study 9.1 Site 56A



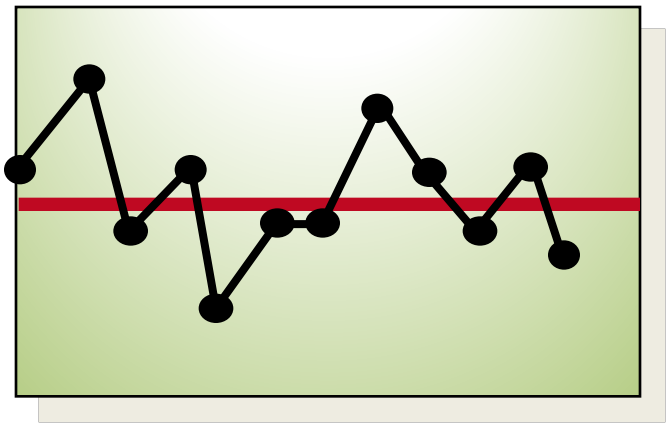
Control Charts

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

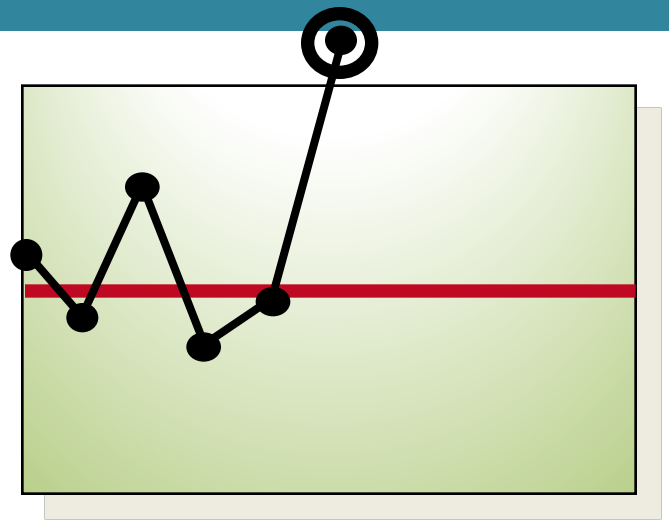
Control Chart for 12 QC samples for a certain peptide



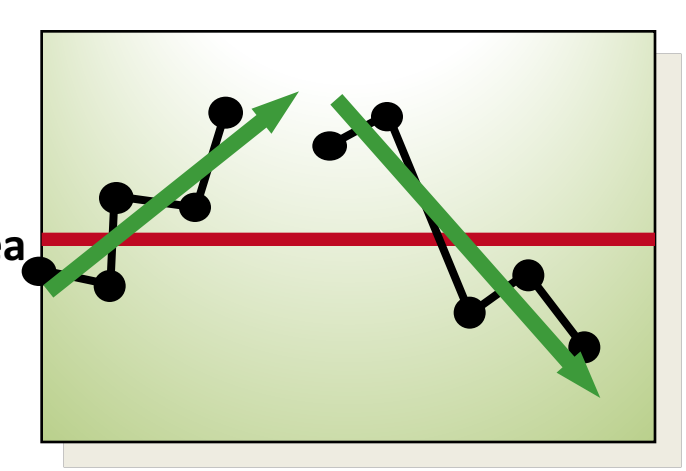
Patterns of control charts



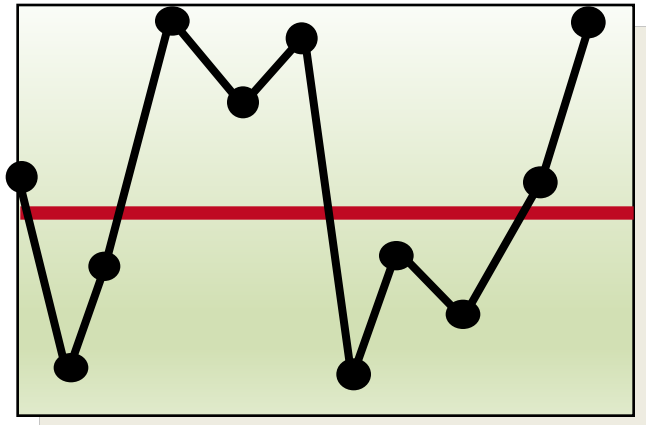
Normal behavior. Process is "in control."



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



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

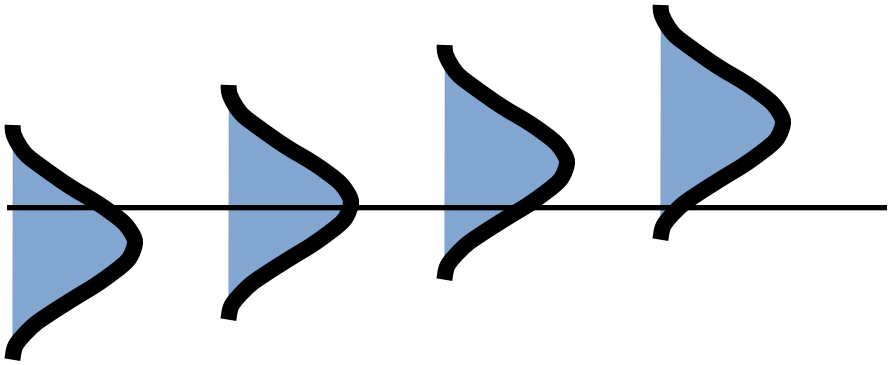


Erratic behavior. Investigate.

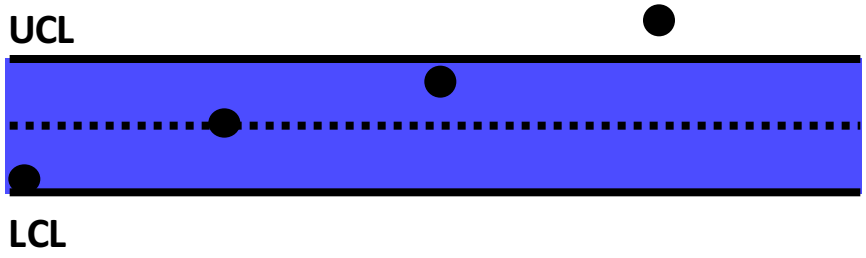
Simultaneous monitoring of LC MS/MS mean and variation

(a)

These sampling distributions result in the charts below

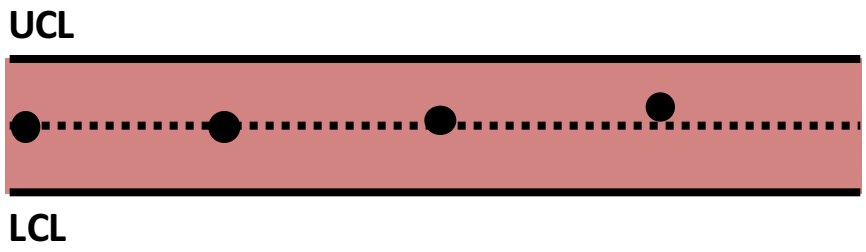


(Sampling mean is shifting upward but range is consistent)



***X*-chart**

(*X* chart detects shift in central tendency)



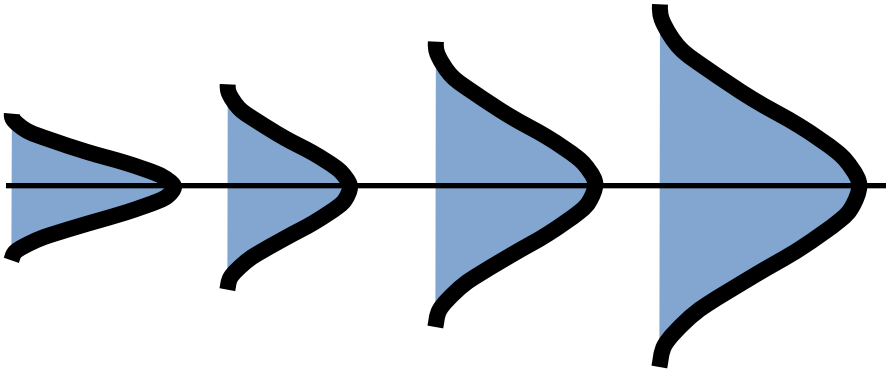
MR-chart

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

Simultaneous monitoring of LC MS/MS mean and variation

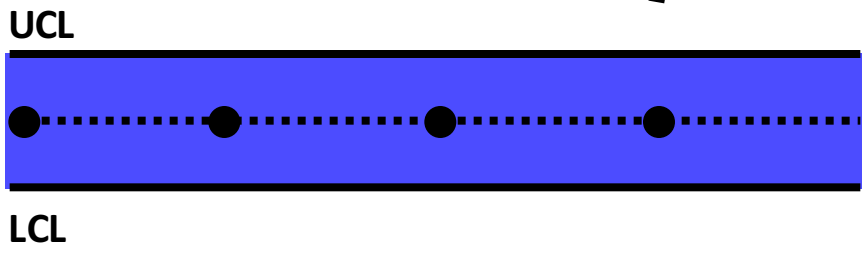
(b)

These sampling distributions result in the charts below



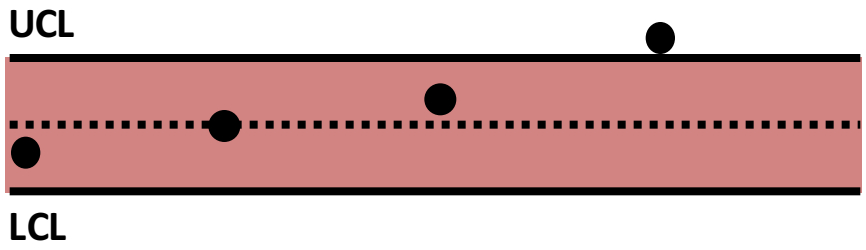
(Sampling mean is constant but dispersion is increasing)

***X*-chart**



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

***MR*-chart**

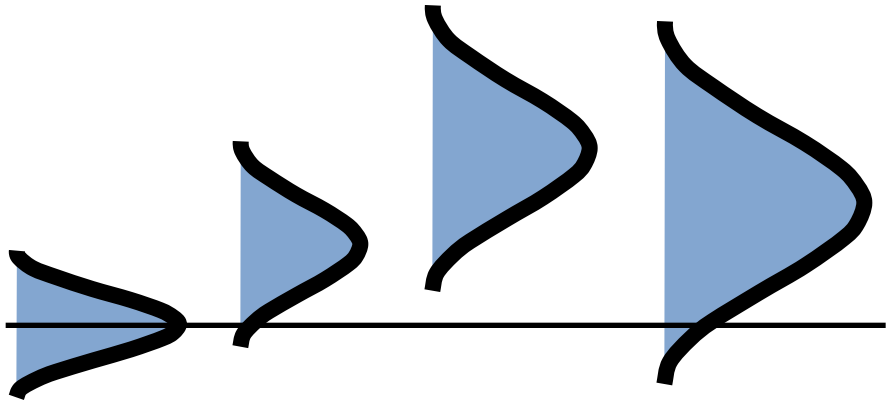


(*MR*-chart detects increase in dispersion)

Simultaneous monitoring of LC MS/MS mean and variation

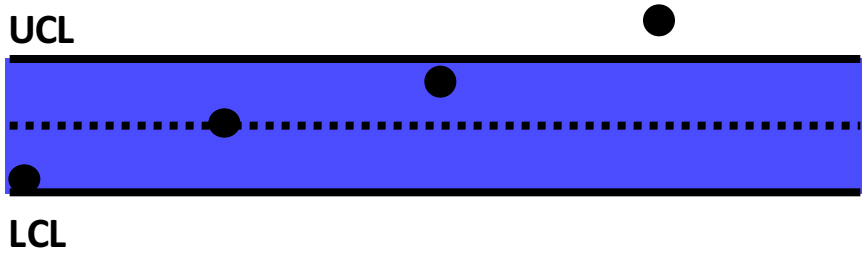
(c)

These sampling distributions result in the charts below



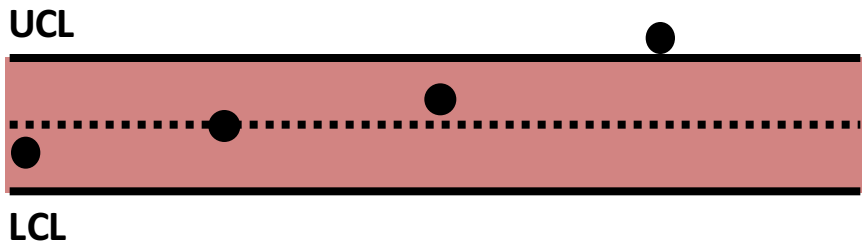
(Sampling mean is constant but dispersion is increasing)

***X*-chart**



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

***MR*-chart**



(M*R*-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

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

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:

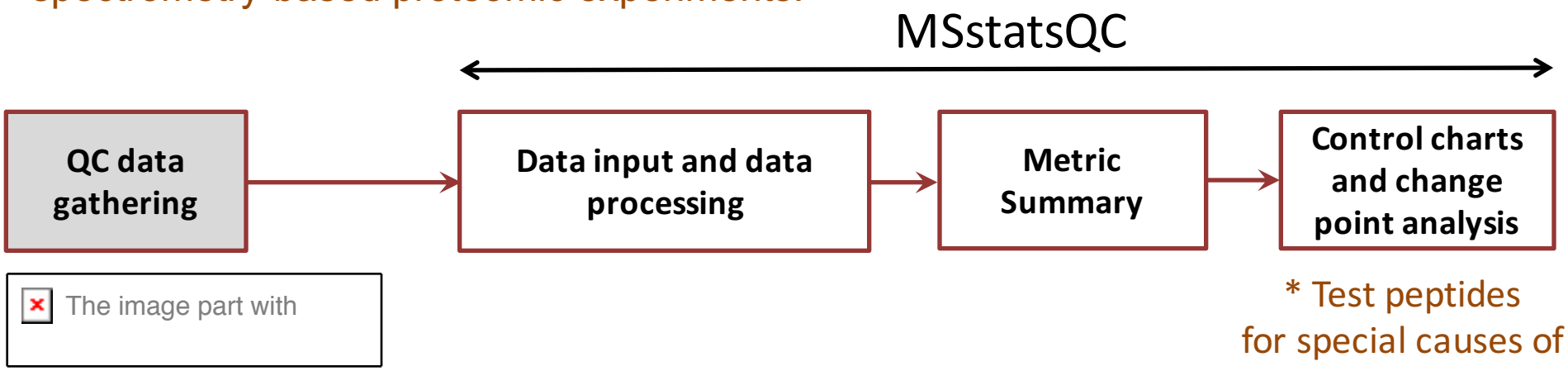
```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("MSstatsQCgui")
```

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.



MSstats compatible experiments and metrics

MS acquisition

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

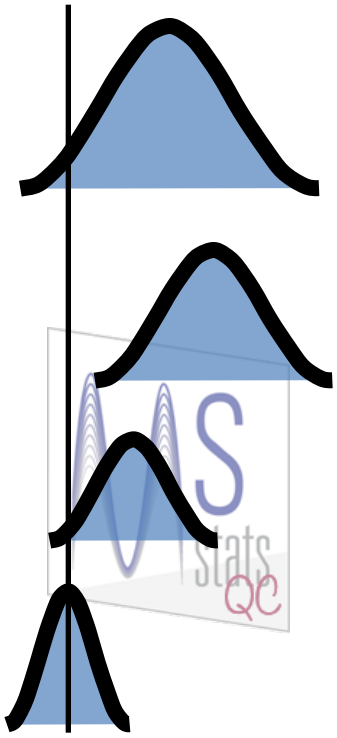
Analysis

- **Decision support tools**
- **Control charts**
- **Change point analysis**

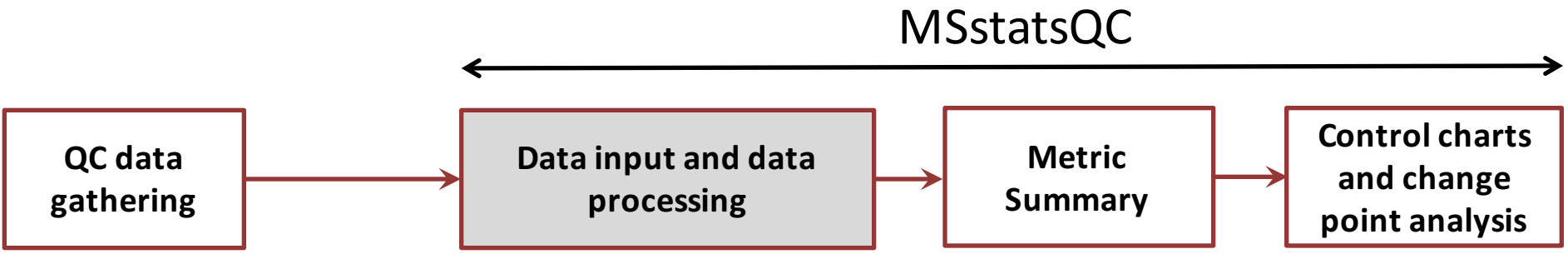
Metrics

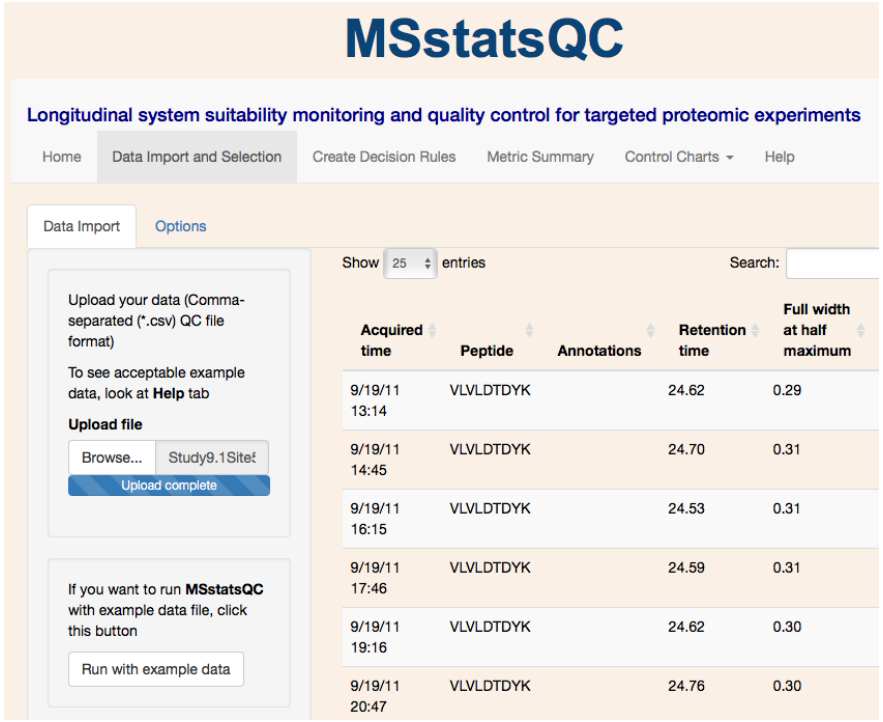
- **Retention time**
- **Total peak area**
- **Full width at half maximum (FWHM)**
- **Peak asymmetry**
- **Many more…**

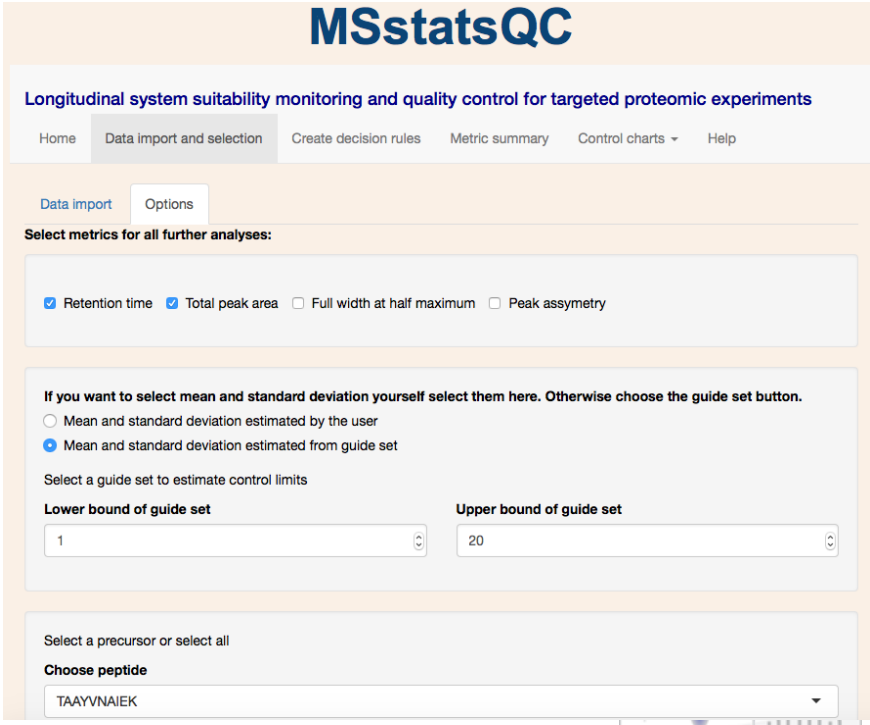
When	Mean	Variation
Large shifts	X	MR
Small shifts	CUSUMm	CUSUMv
Time of a problem	Change point	Change point



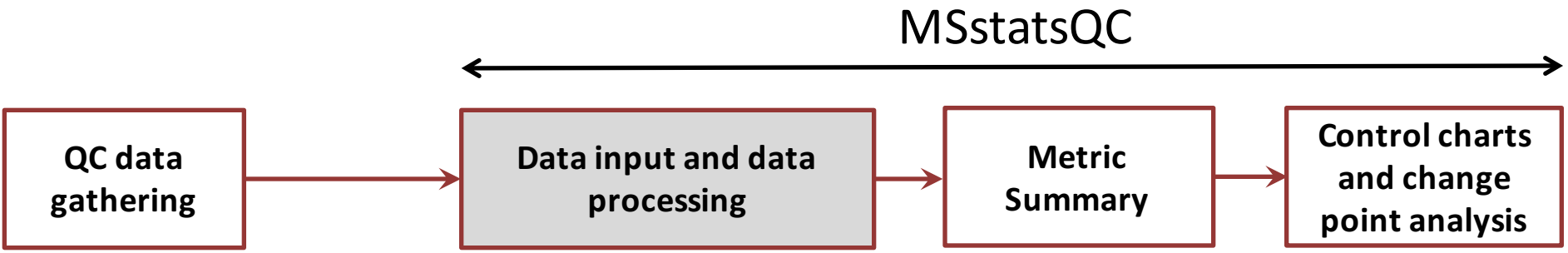
1. Data input and data table

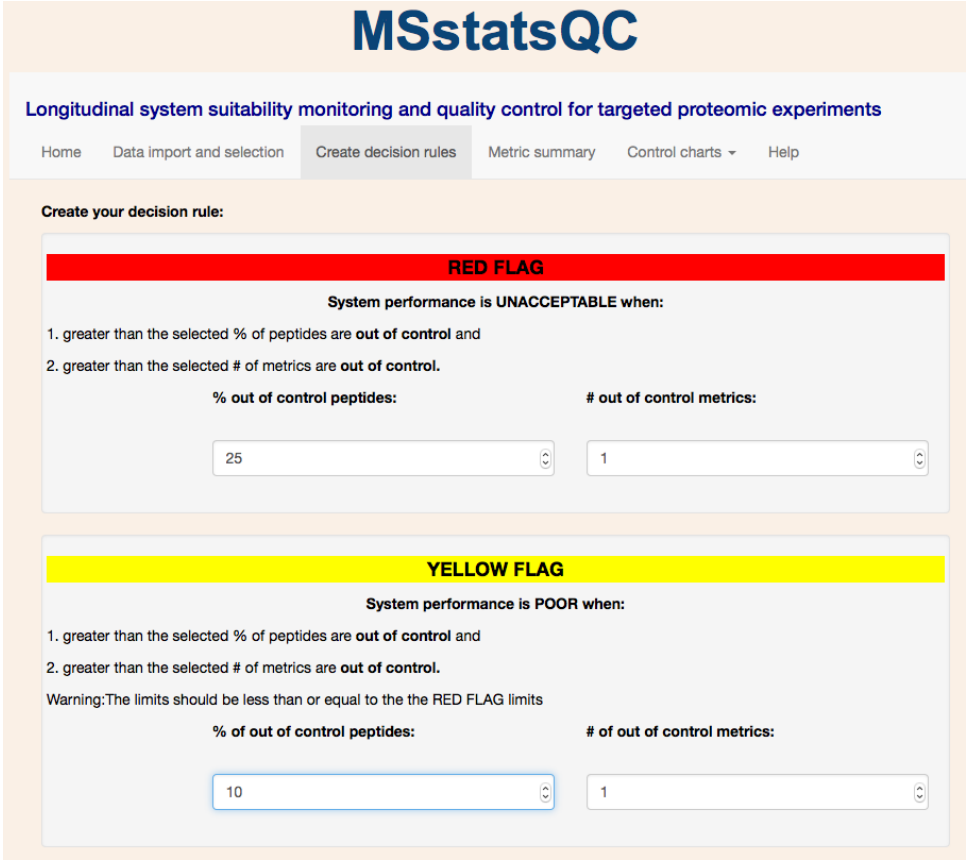






1. Data input and data table

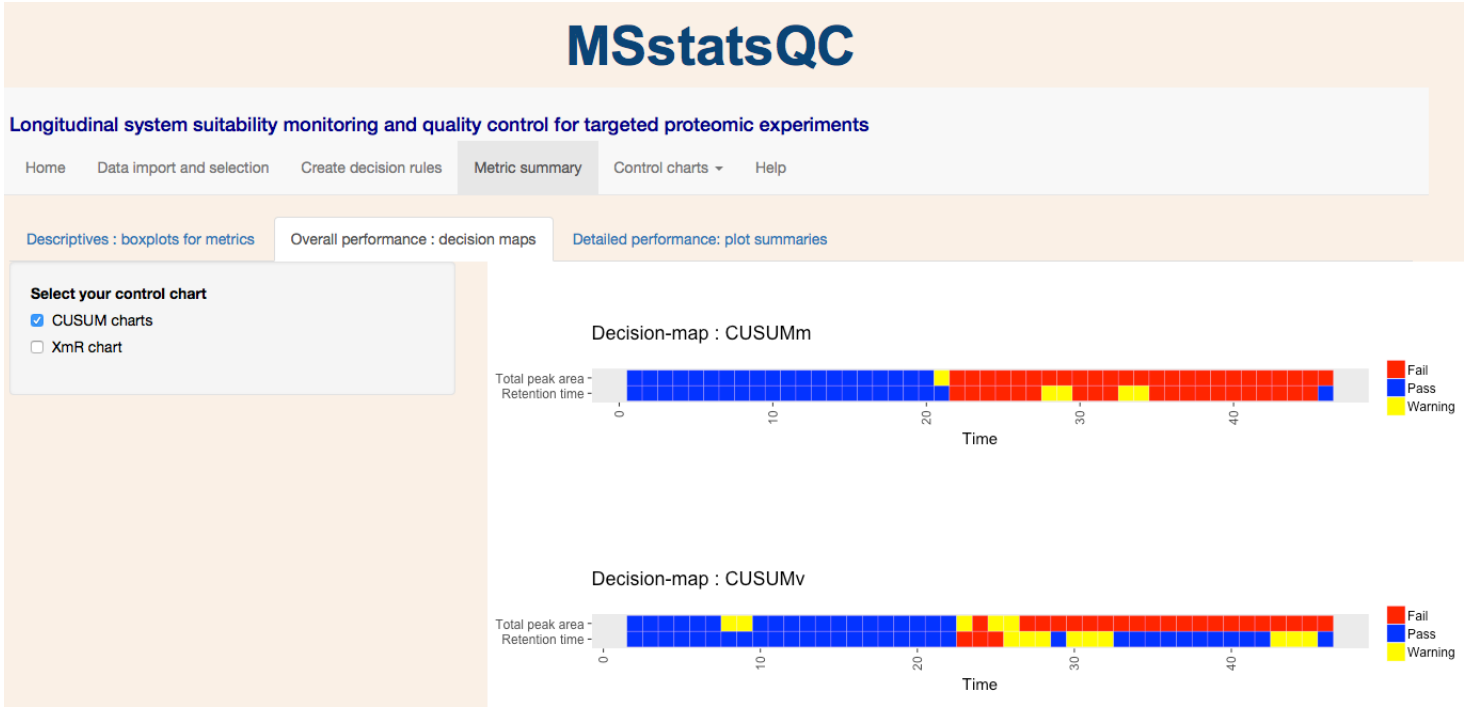




2. Metric Summary

MSstatsQC

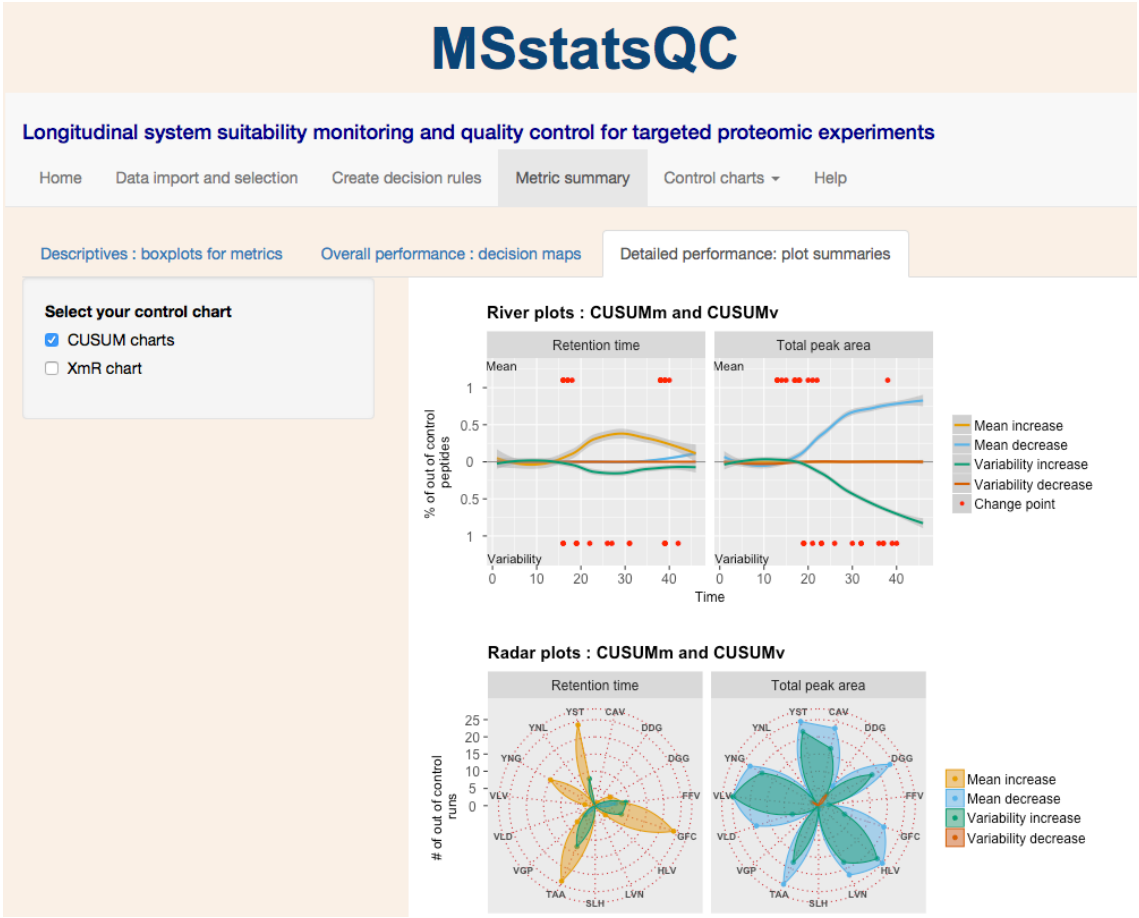
QC data gathering → Data input and data processing → **Metric Summary** → Control charts and change point analysis



2. Metric Summary

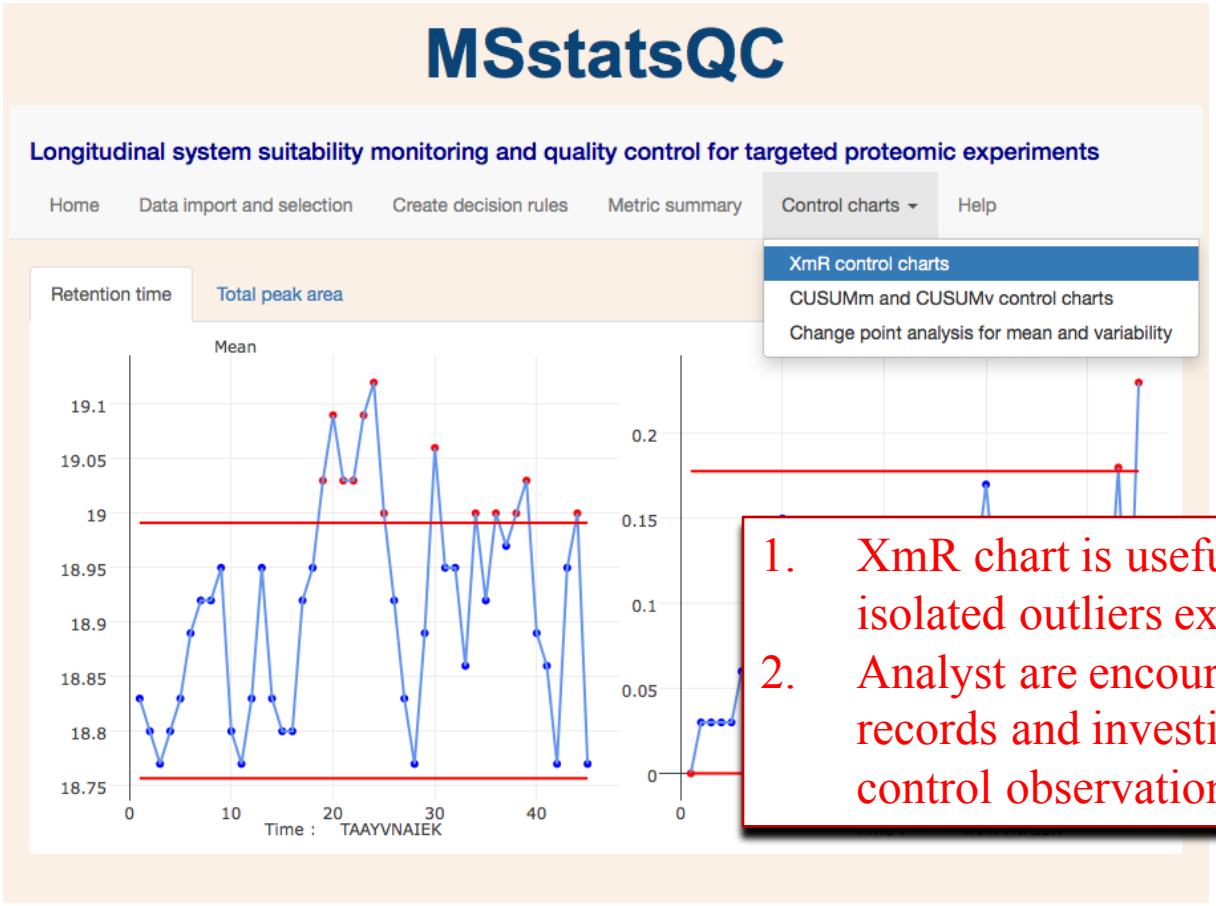
MSstatsQC

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)

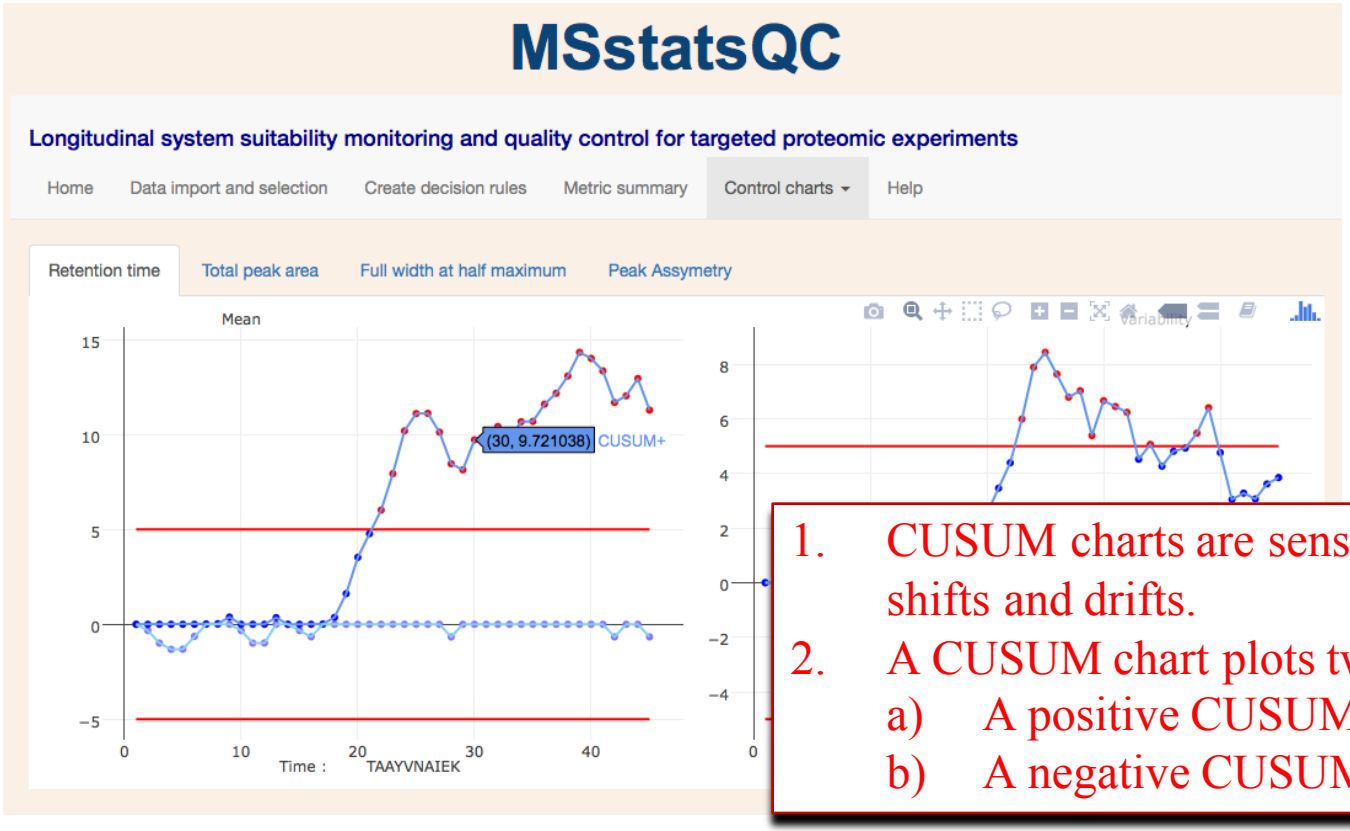
MSstatsQC



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.

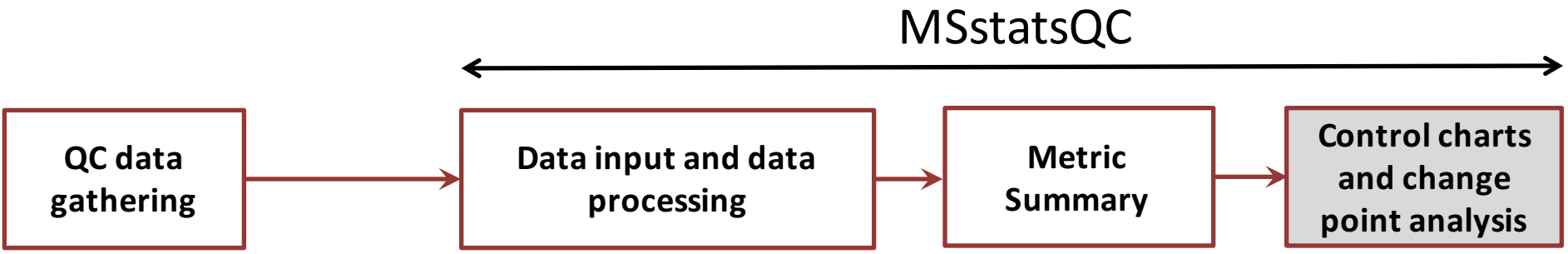
3. Control charts-Cumulative Sum (CUSUM)

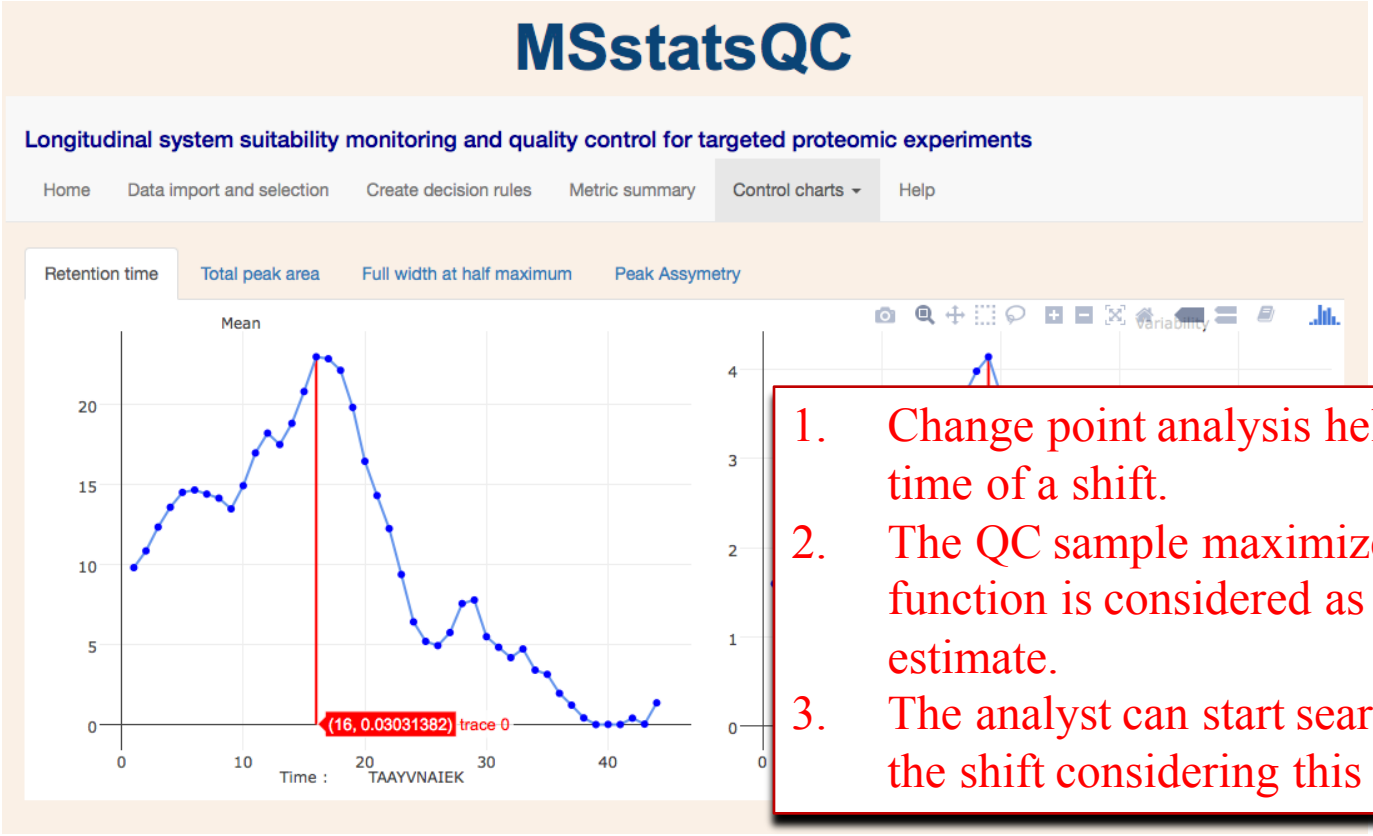
MSstatsQC



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





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 p… MRM-MS).

Abbatiello SE[1], Mani DR, Schilling B, Macl… JM, Hedrick V, Inerowicz HD, Jackson A, K… A, Waldemarson S, Whitwell CA, You J, Zh… Neubert TA, Paulovich A, Regnier F, Skate…

…l for the quantitative …tion monitoring-MS (LC-

…Allen S, Dodder NG, Ghosh M, Held …ox K, Smith D, Tomazela D, Wahlander …, Gibson BW, Liebler D, Maccoss M,

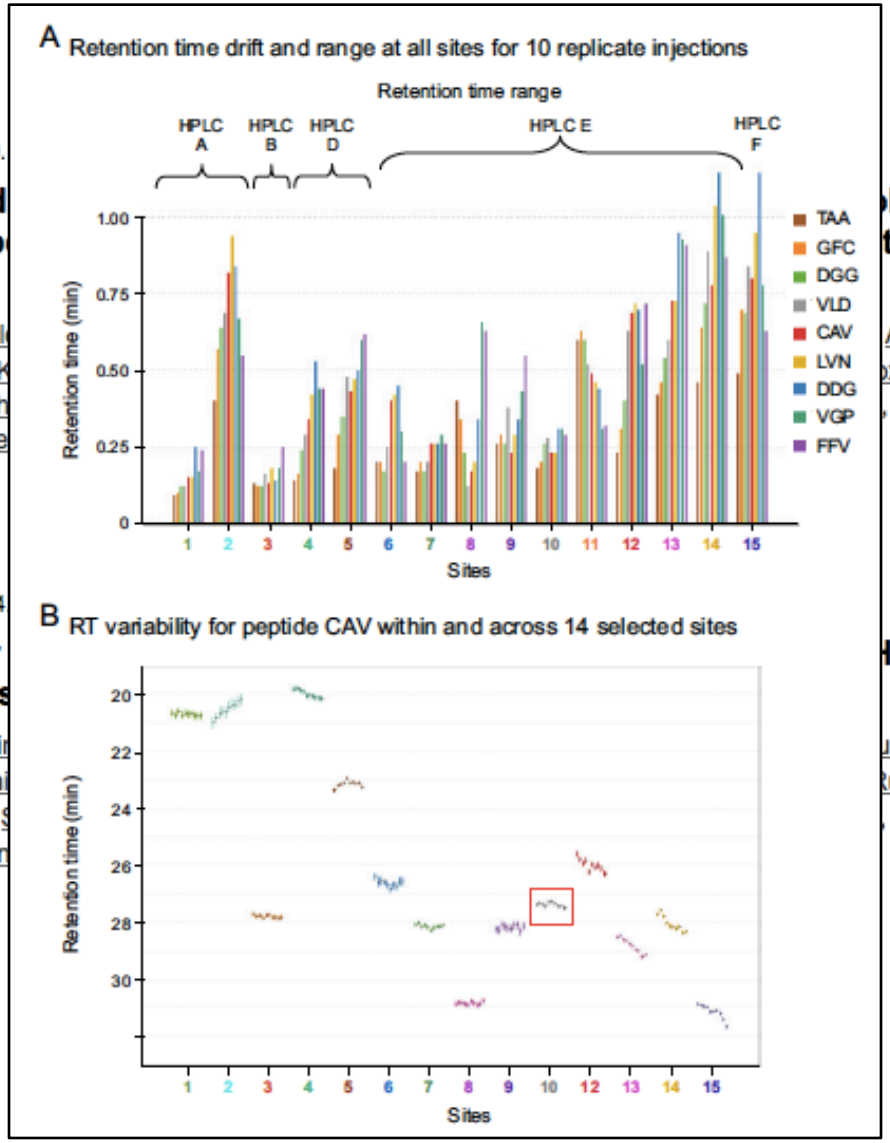
Mol Cell Proteomics. 2015 Sep;14(9):2357-74

Large-Scale Interlaboratory … Quantitative Peptide Assays

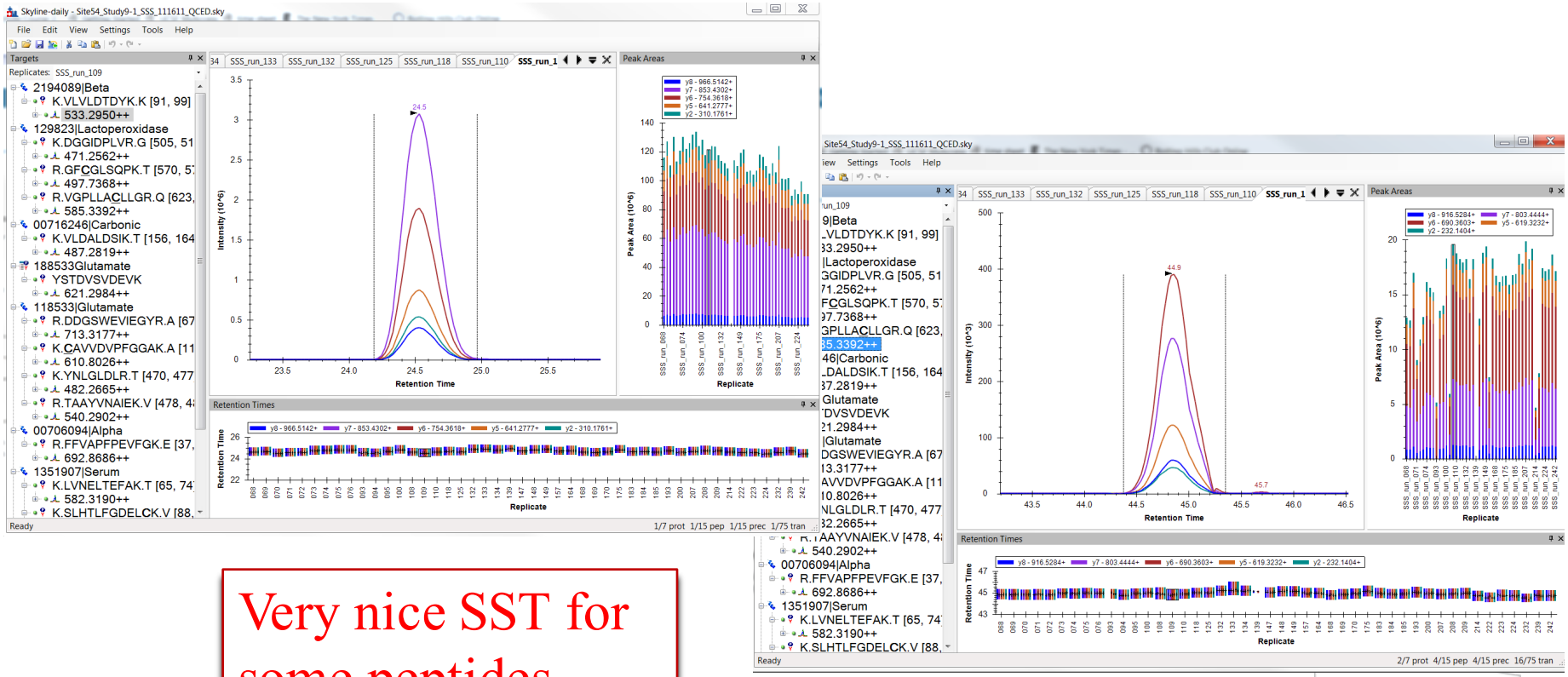
Abbatiello SE[1], Schilling B[2], Mani DR[1], Zi… JM[2], Inerowicz HD[7], Jackson A[8], Keshishi… N[13], Shaddox K[3], Skates SJ[14], Kuhn E[1], … MJ[5], Neubert TA[10], Paulovich AG[15], Regn…

…ighly Multiplexed,

…usack MP[2], Gosh M[6], Hedrick V[7], Held …udnick P[12], Sadowski P[10], Sedransk …, Gibson BW[2], Liebler DC[3], MacCoss



CPTAC Study 9.1 **Site 54**

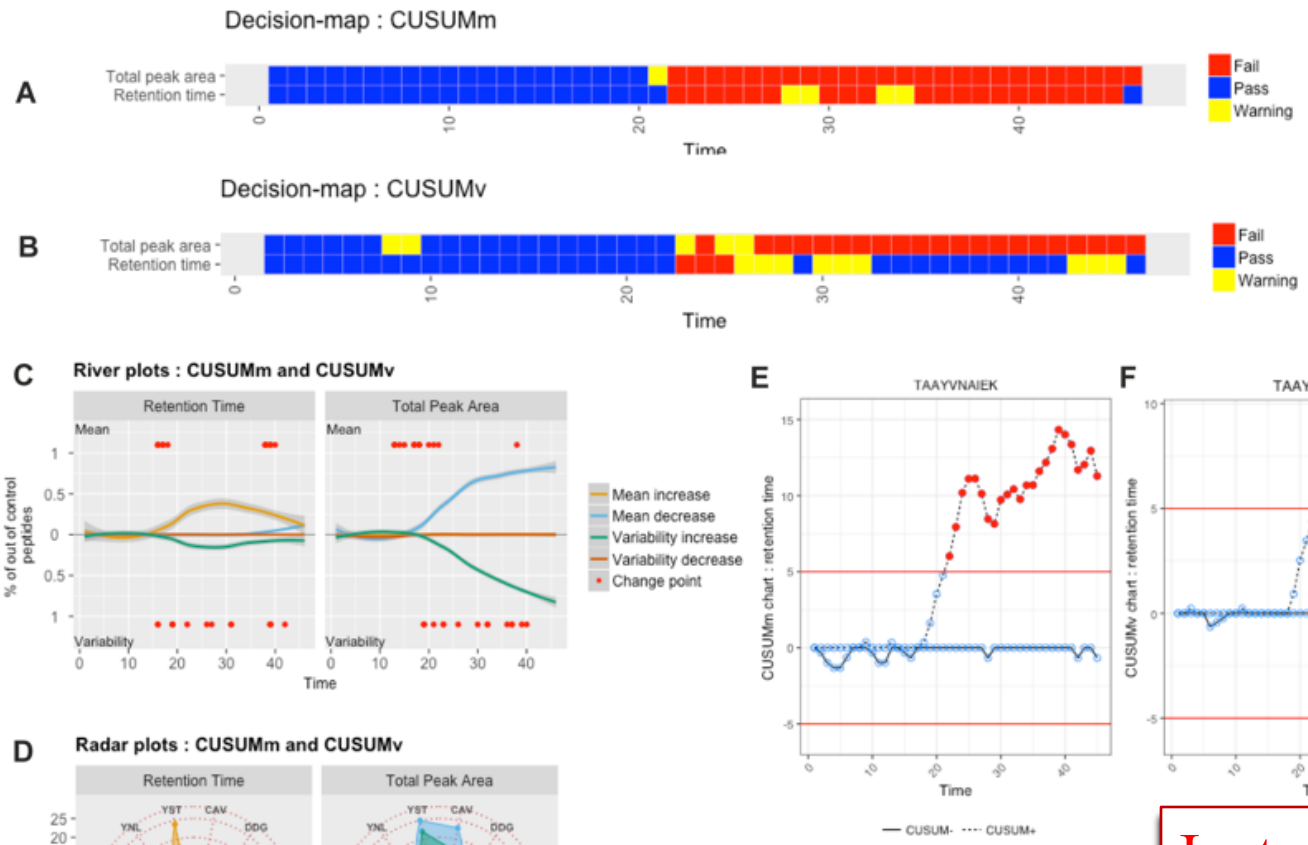


Very nice SST for some peptides

Changes in retention time for some peptides

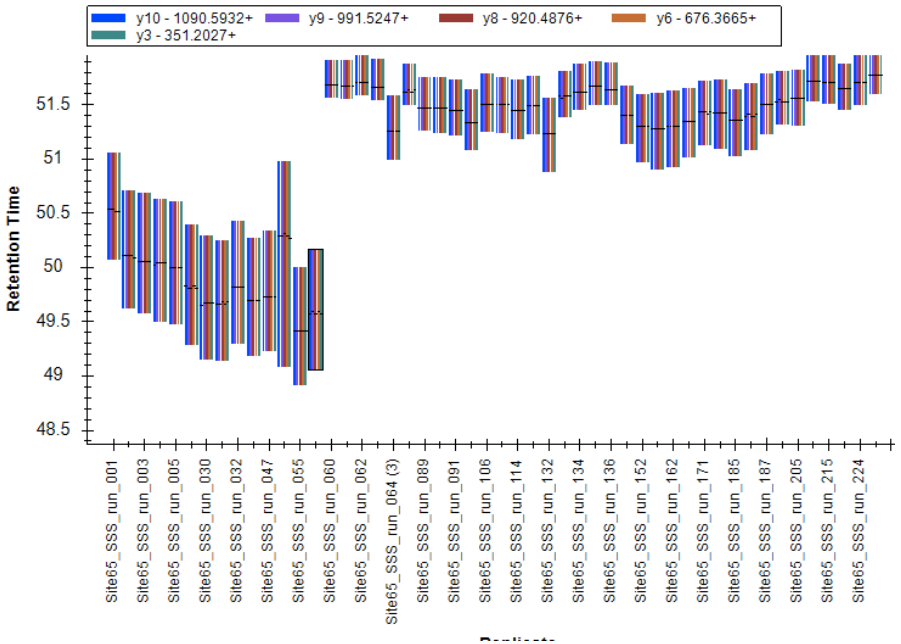
CPTAC Study 9.1

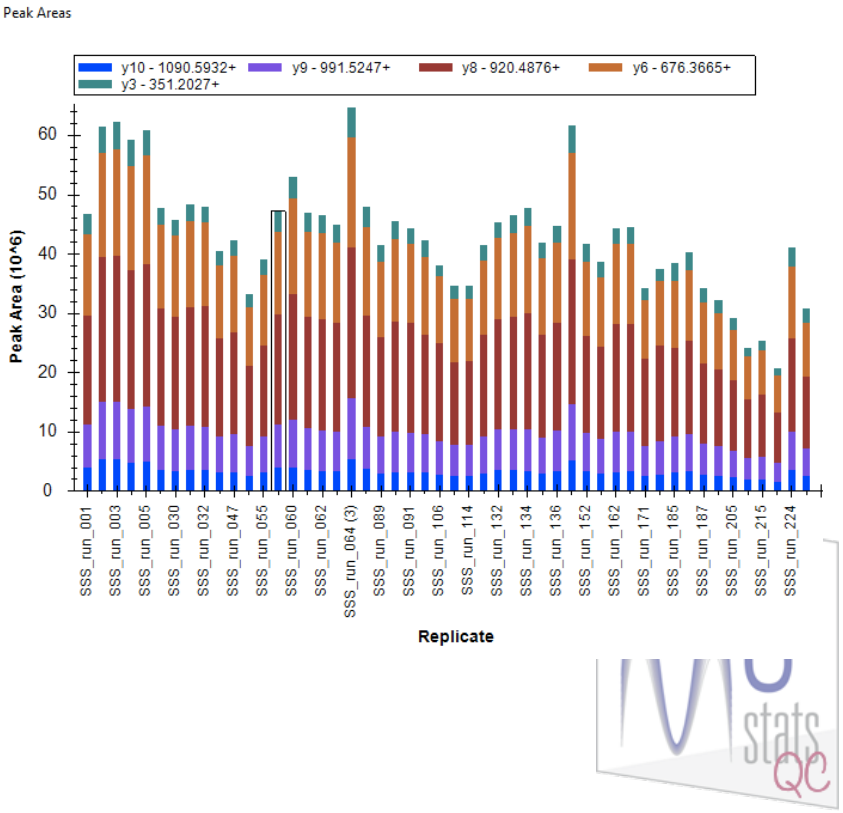
Site 54



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

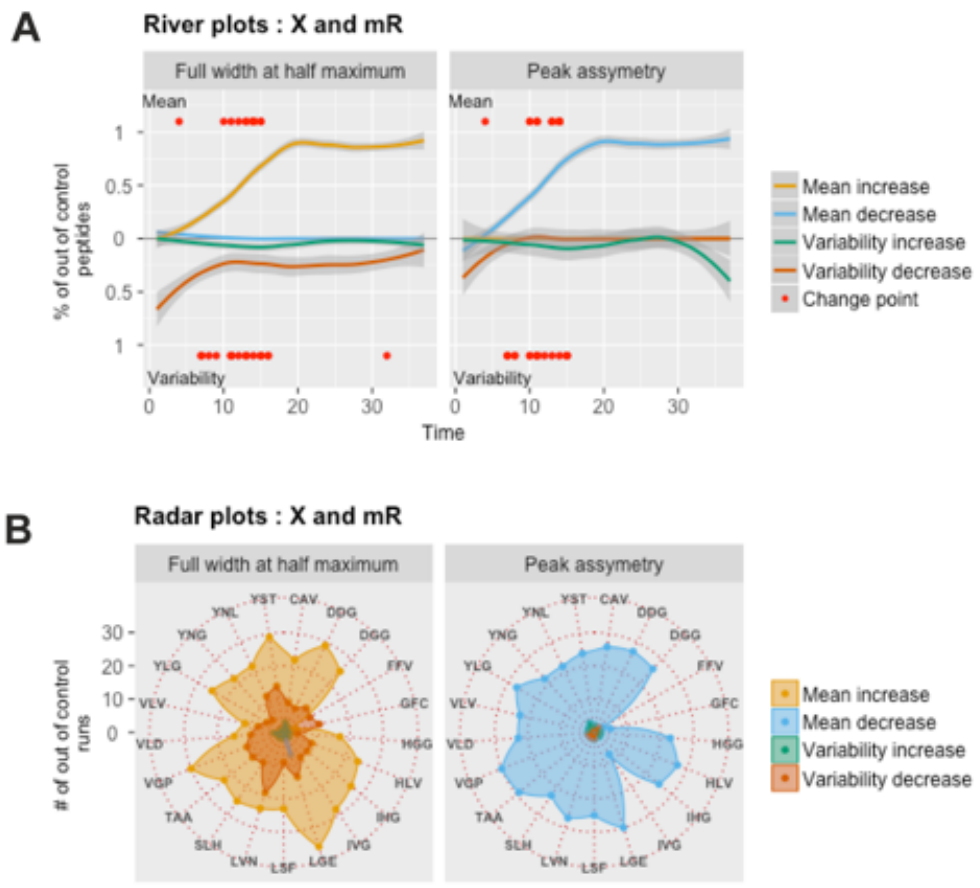
CPTAC Study 9.1 Site 65



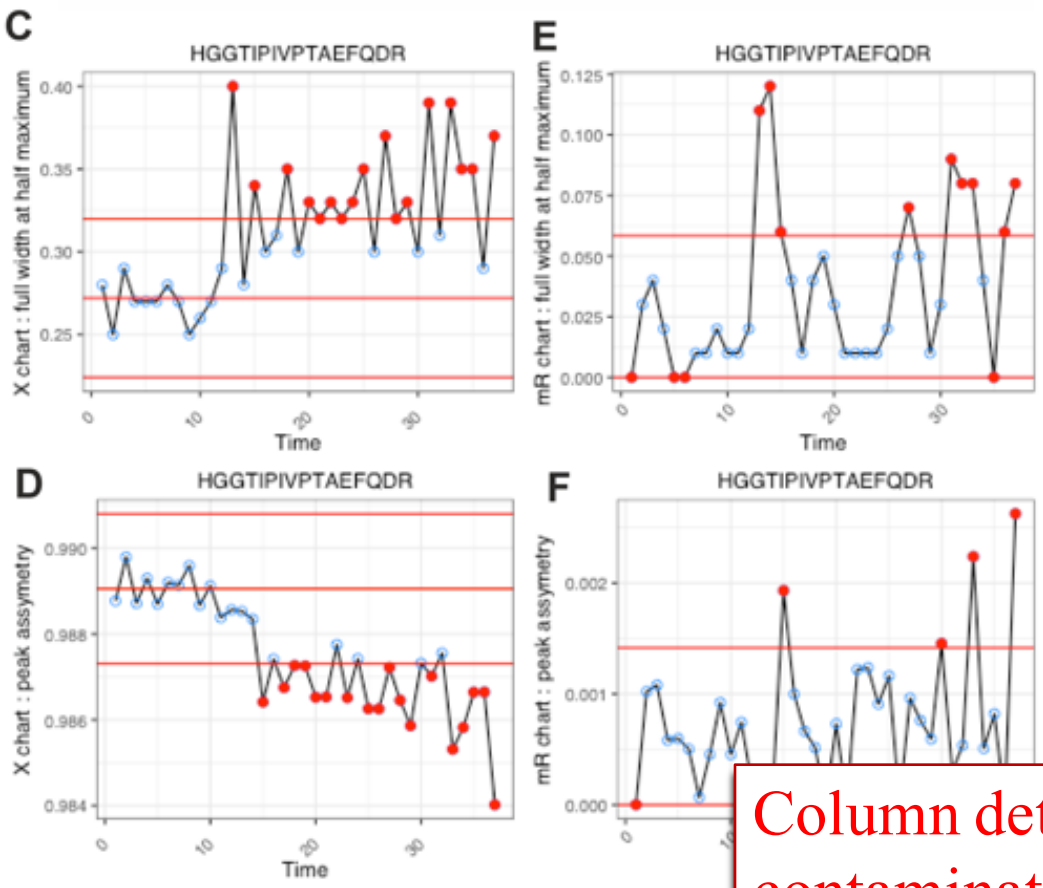


CPTAC Study 9.1

Site 65

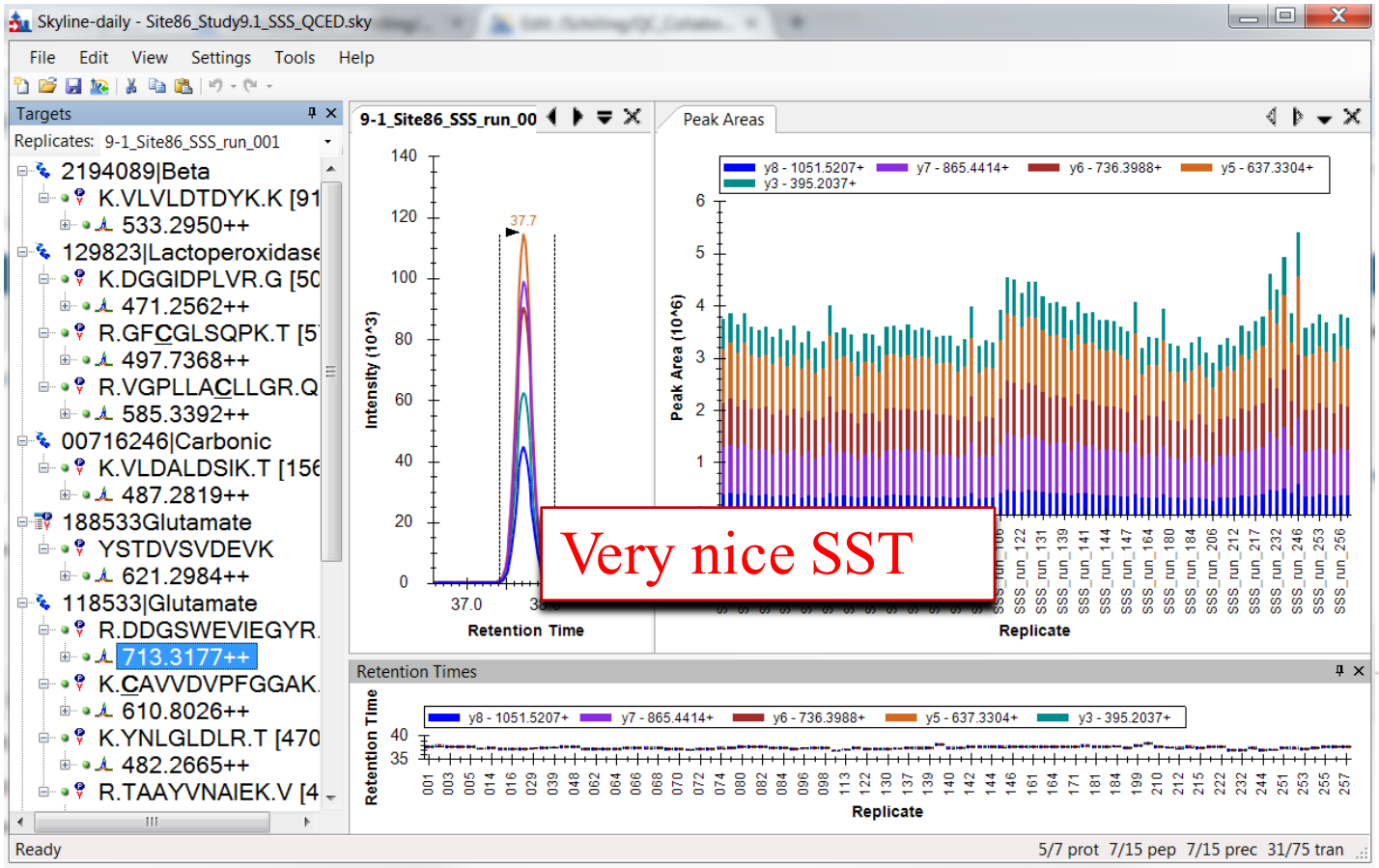


CPTAC Study 9.1 **Site 65**

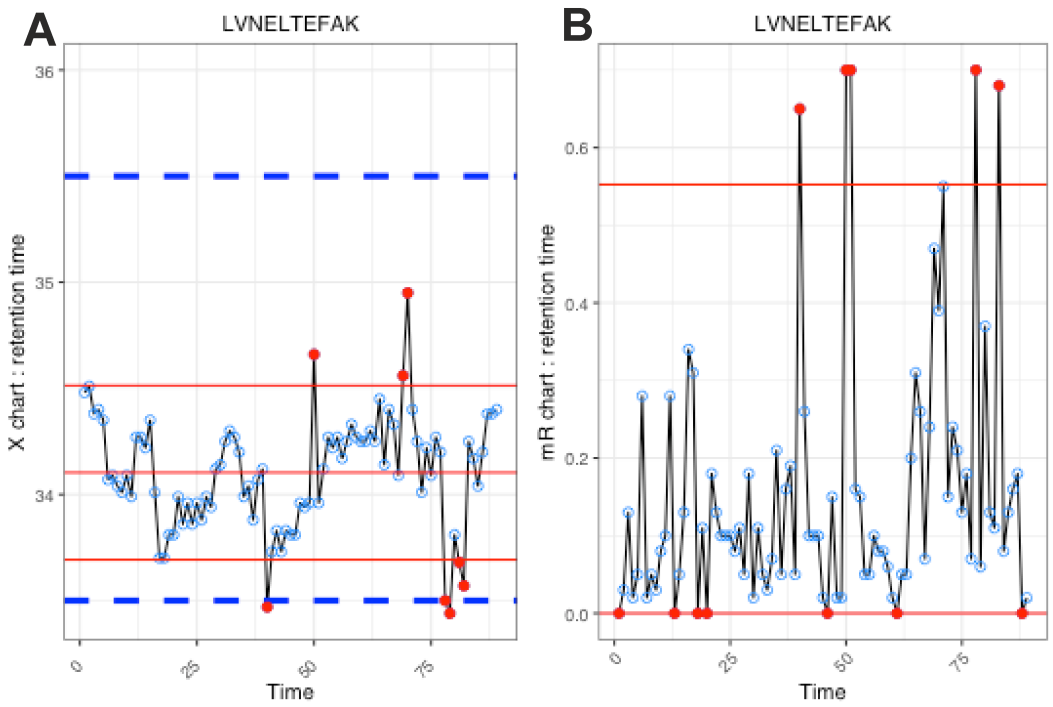


Column deterioration, contamination, and fitting fatigue

CPTAC Study 9.1 Site 86

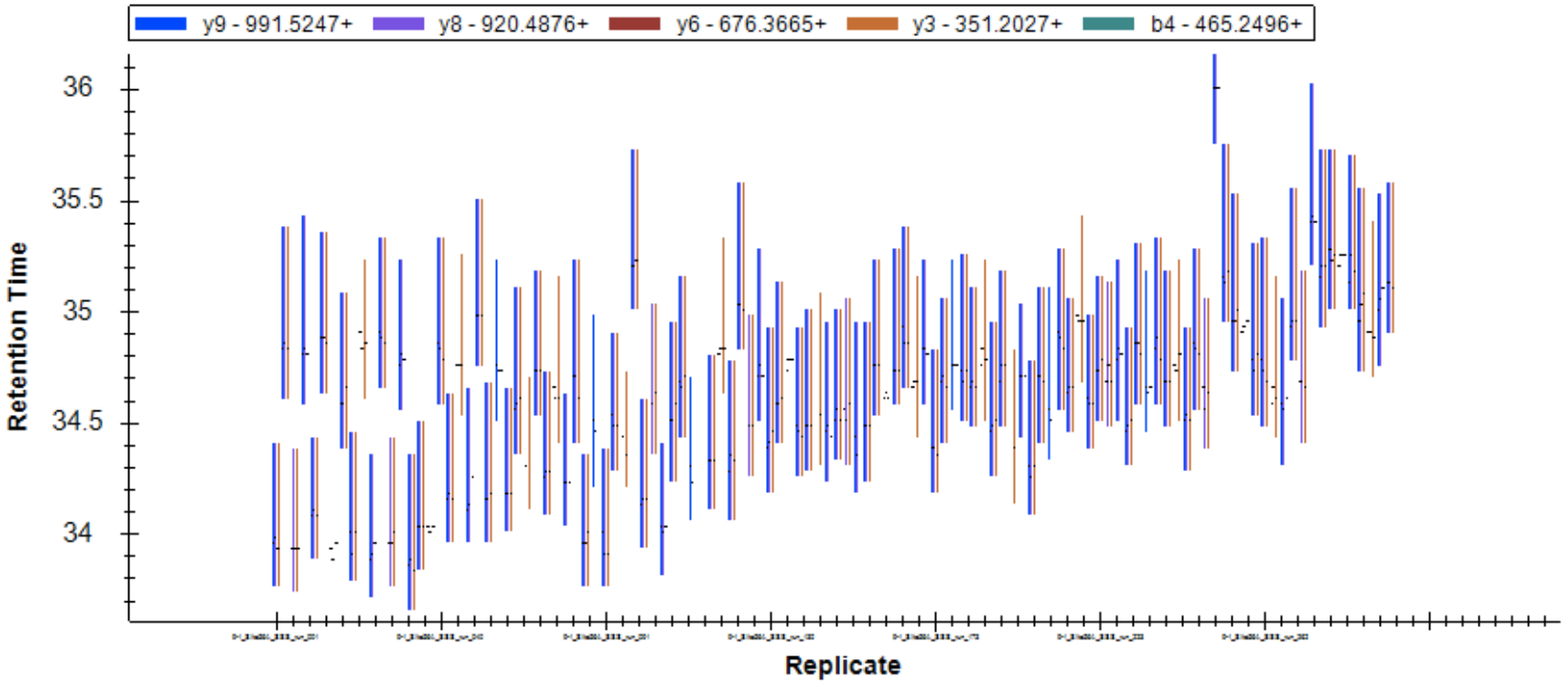


CPTAC Study 9.1 **Site 86**

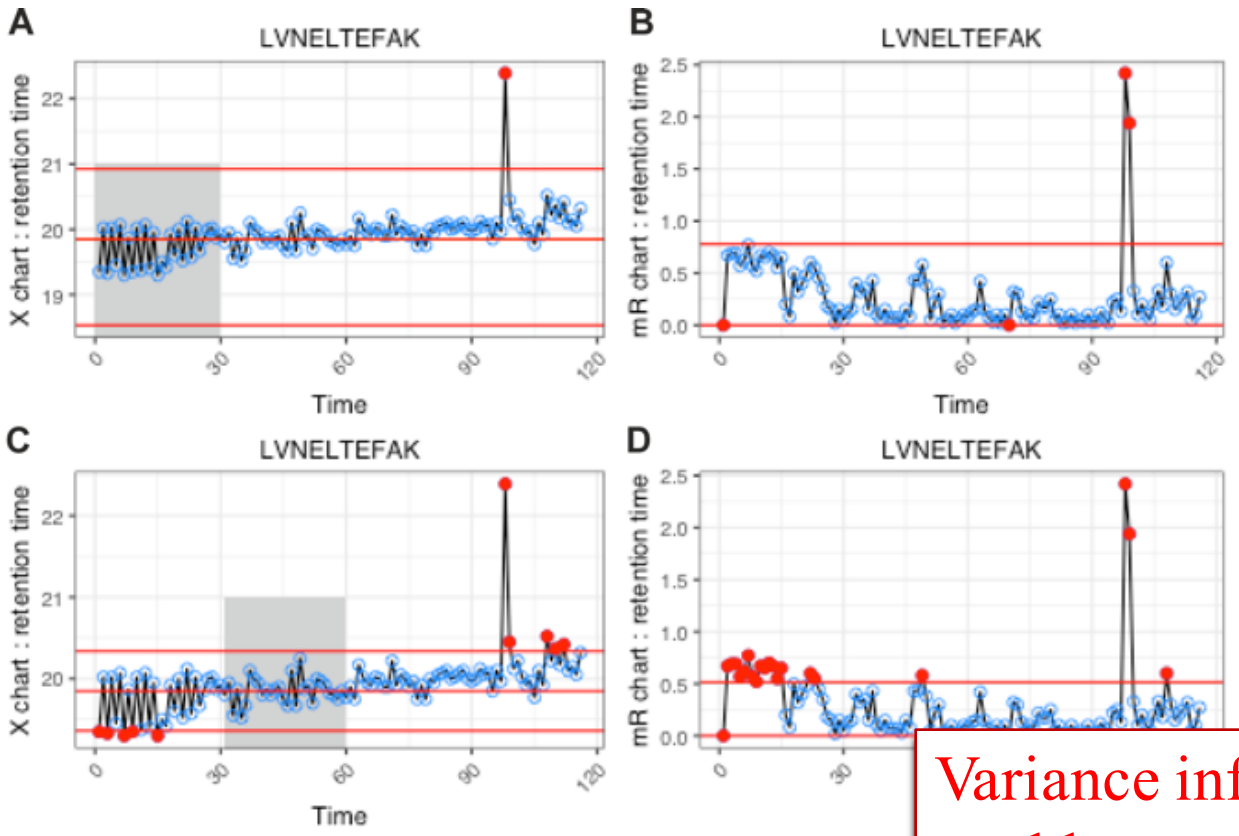


Environmental factors such as temperature or pressure changes

CPTAC Study 9.1 Site 56A

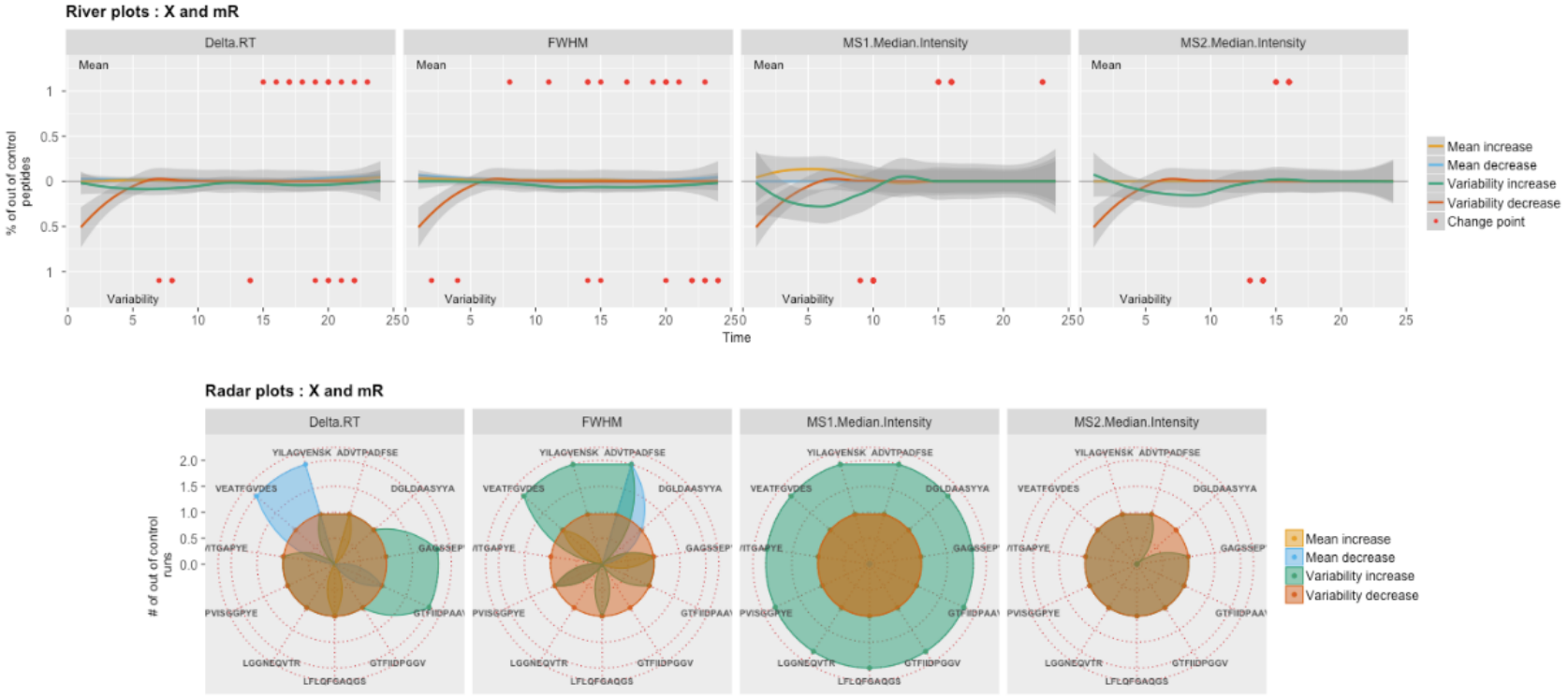


CPTAC Study 9.1 Site 56A



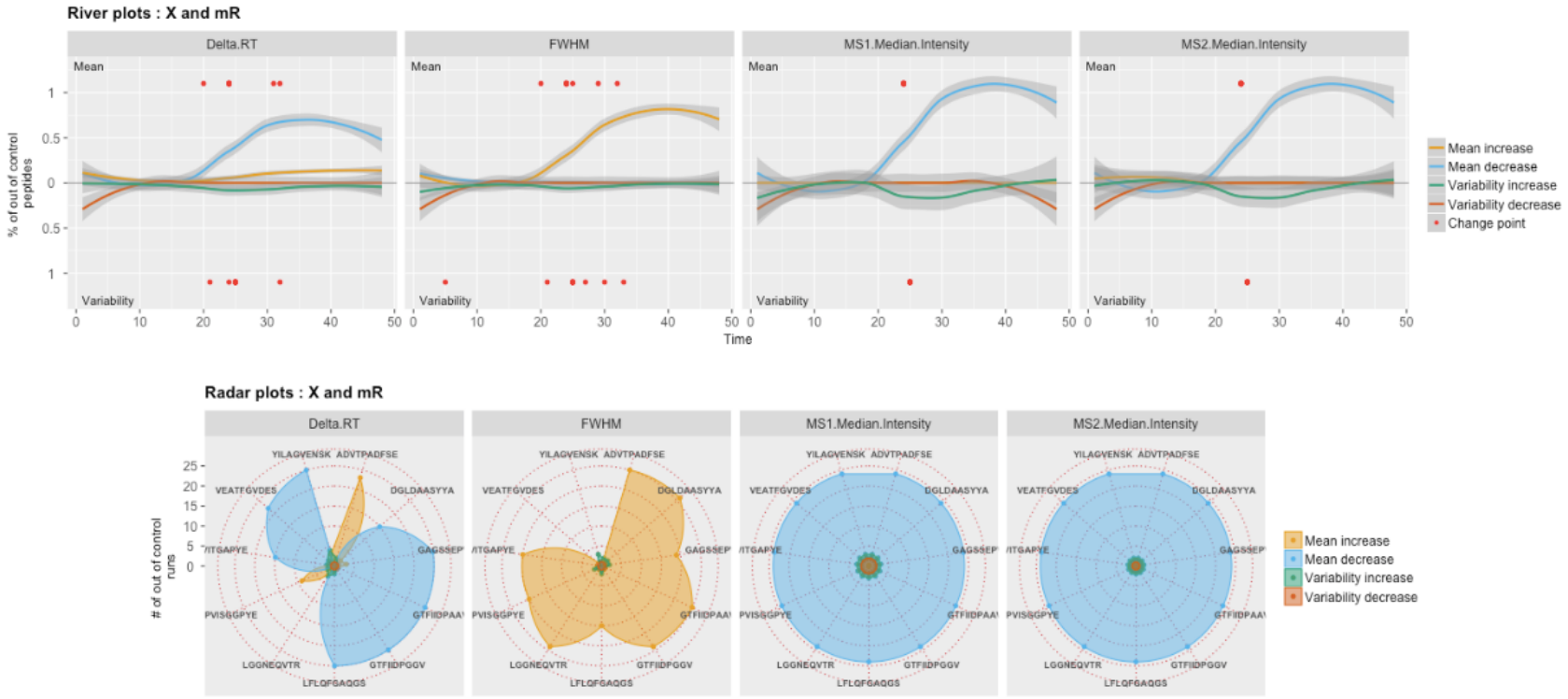
Variance inflation could occur during the equilibration phase of a new LC column

DDA Data: iRT peptides



Most of the peptides are stable and system is acceptable

DIA Data: iRT peptides

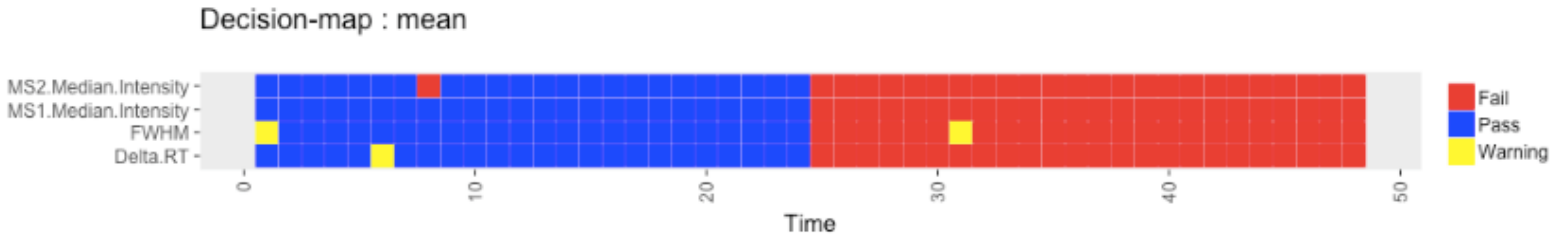


Most of the peptides are unstable and system is unacceptable due to changes in all metrics monitored.

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

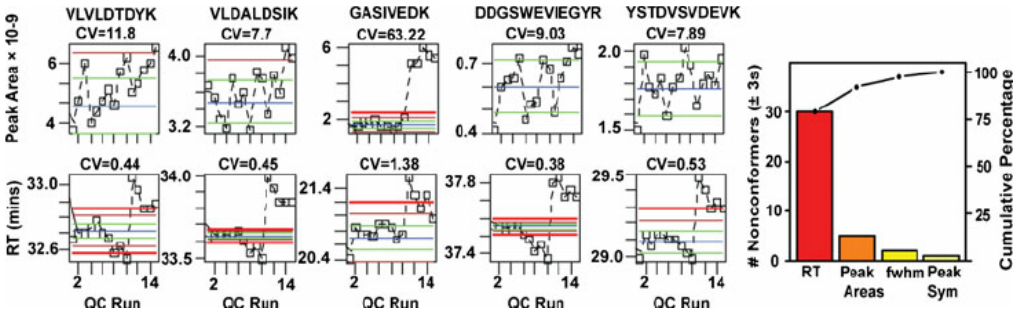
Change point is expected to be around 20th run.

DIA Data: iRT peptides

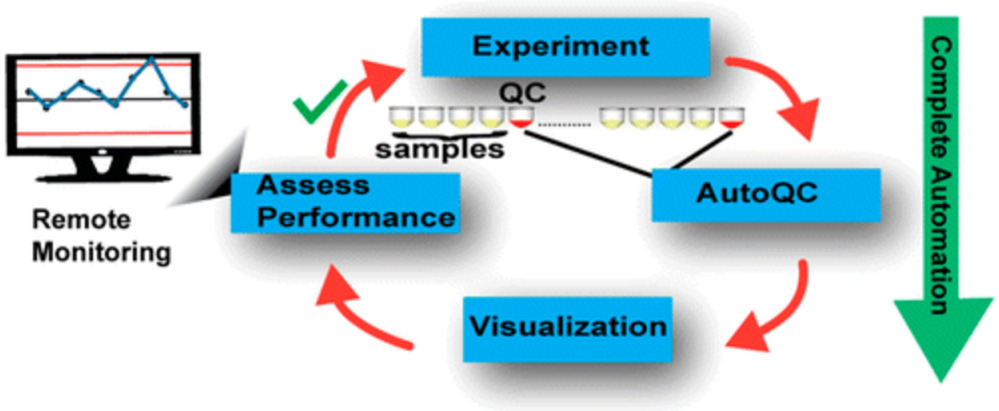


System is not able to pass predefined performance criteria (decision-map)

SPC applied to mass spectrometry proteomics



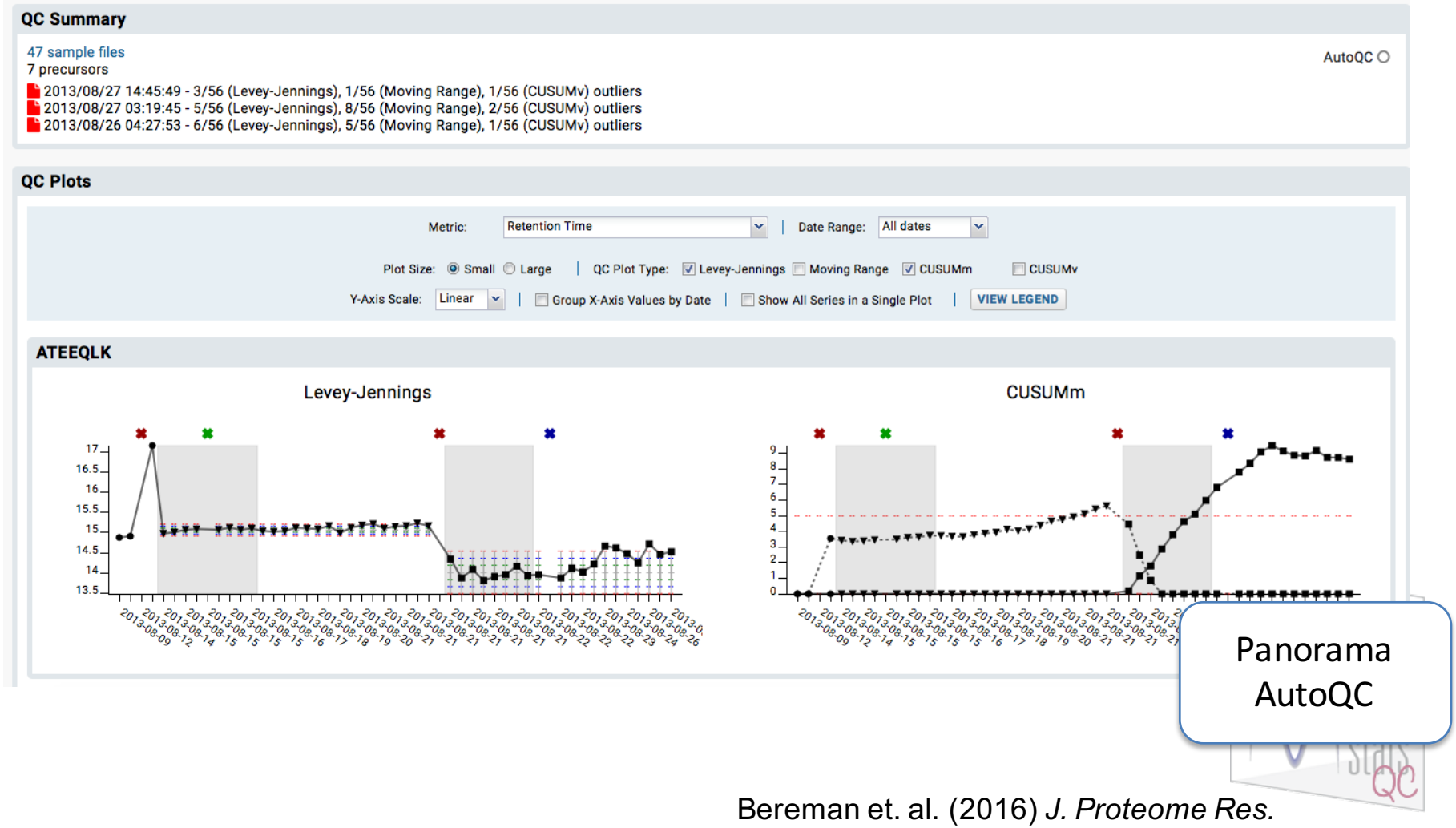
SProCop



Panorama AutoQC

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

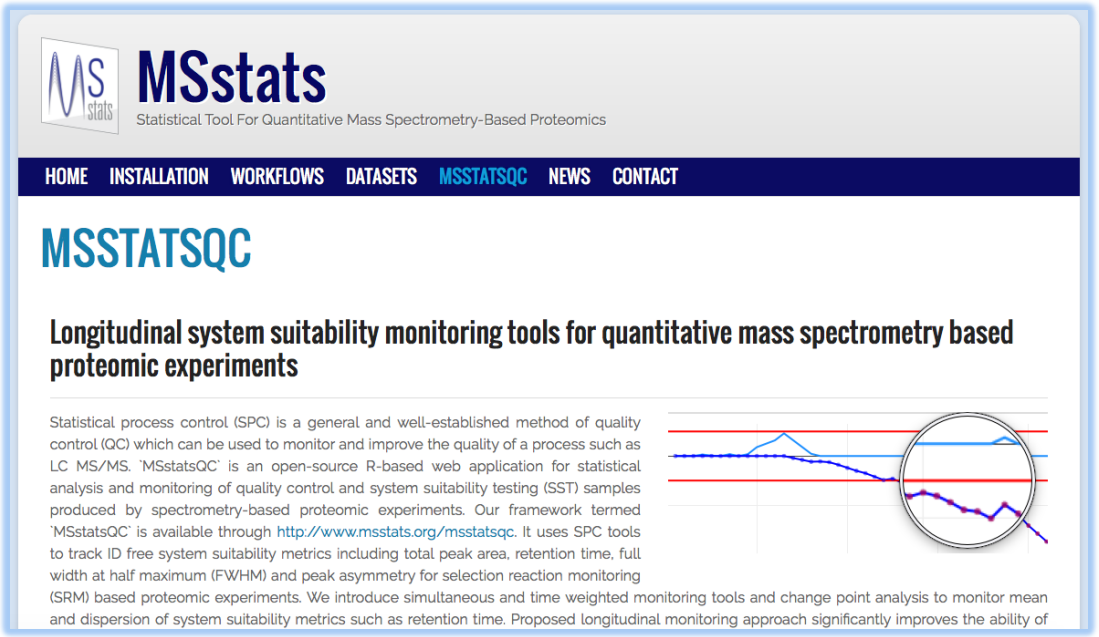
SPC applied to mass spectrometry proteomics



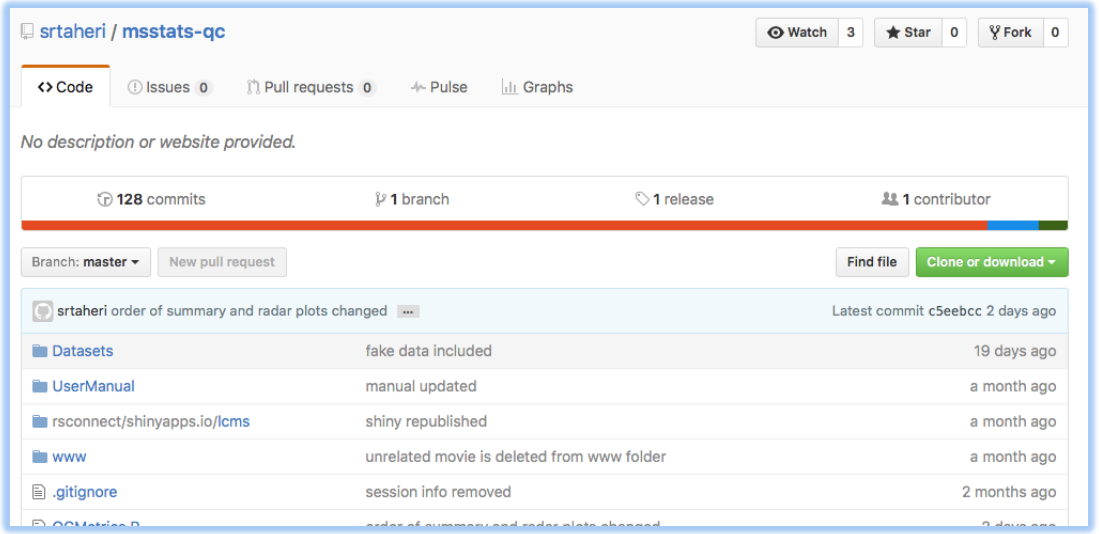
Panorama AutoQC

Bereman et. al. (2016) *J. Proteome Res.*

msstats.org/msstatsqc



- News about MsstatsQC
- Example datasets
- Related publications



- 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

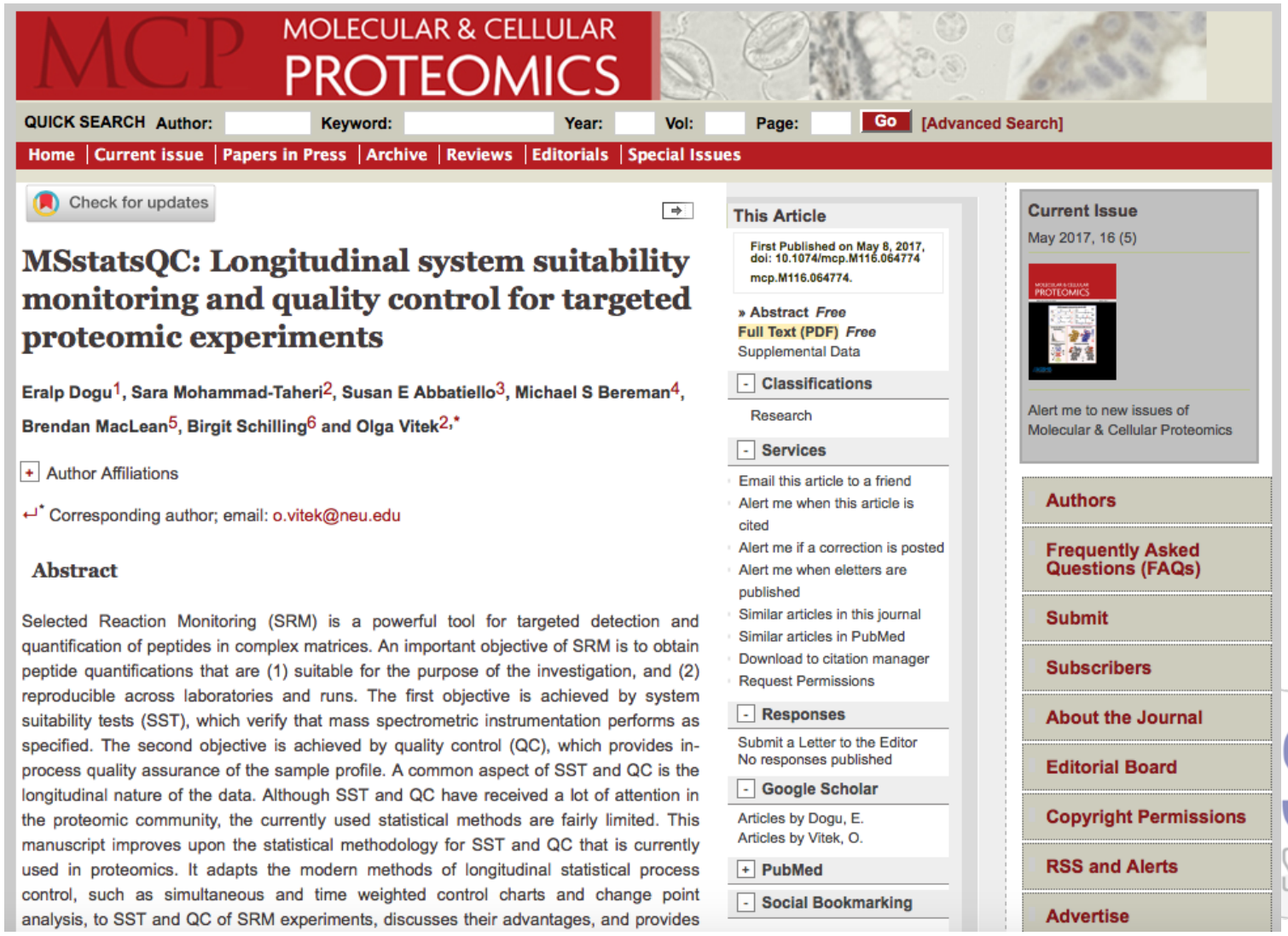
Buck Institute

Birgit Schilling

Thermo Fisher

Sue Abbatiello

Related work



MCP MOLECULAR & CELLULAR PROTEOMICS

QUICK SEARCH Author: Keyword: Year: Vol: Page: Go [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 Dogu[1], Sara Mohammad-Taheri[2], Susan E Abbatiello[3], Michael S Bereman[4], Brendan MacLean[5], Birgit Schilling[6] 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