Analysis of Large Scale MRM Studies Using Skyline

Michael Schirm, PhD
Associate Director, R&D Proteomics

Skyline User Meeting
June 04, 2017

Overview

- CRO providing mass spec based proteomics services since 2001
- Clients: Pharma, Biotech or Government based
- ProteoCarta: MS-based proteomics platform
  - Unbiased LC-MS/MS
  - Targeted LC-MRM/MS
- Specialized in development and analysis of highly multiplexed large scale MRM studies
- Data analysis was very tedious and time consuming:
  - Manual inspection of every peak was required to ensure data quality
  - Automated peak integration had high error rate

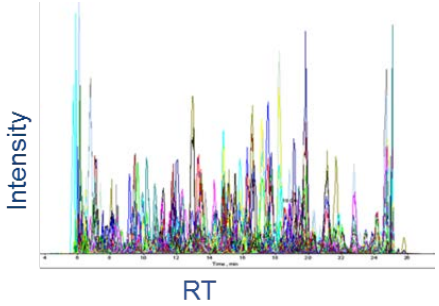
MRM Analysis Workflow

Protein Samples

Denaturation, R/A, Digestion

Spiking of SIL

MRM Analysis



PROCESS:

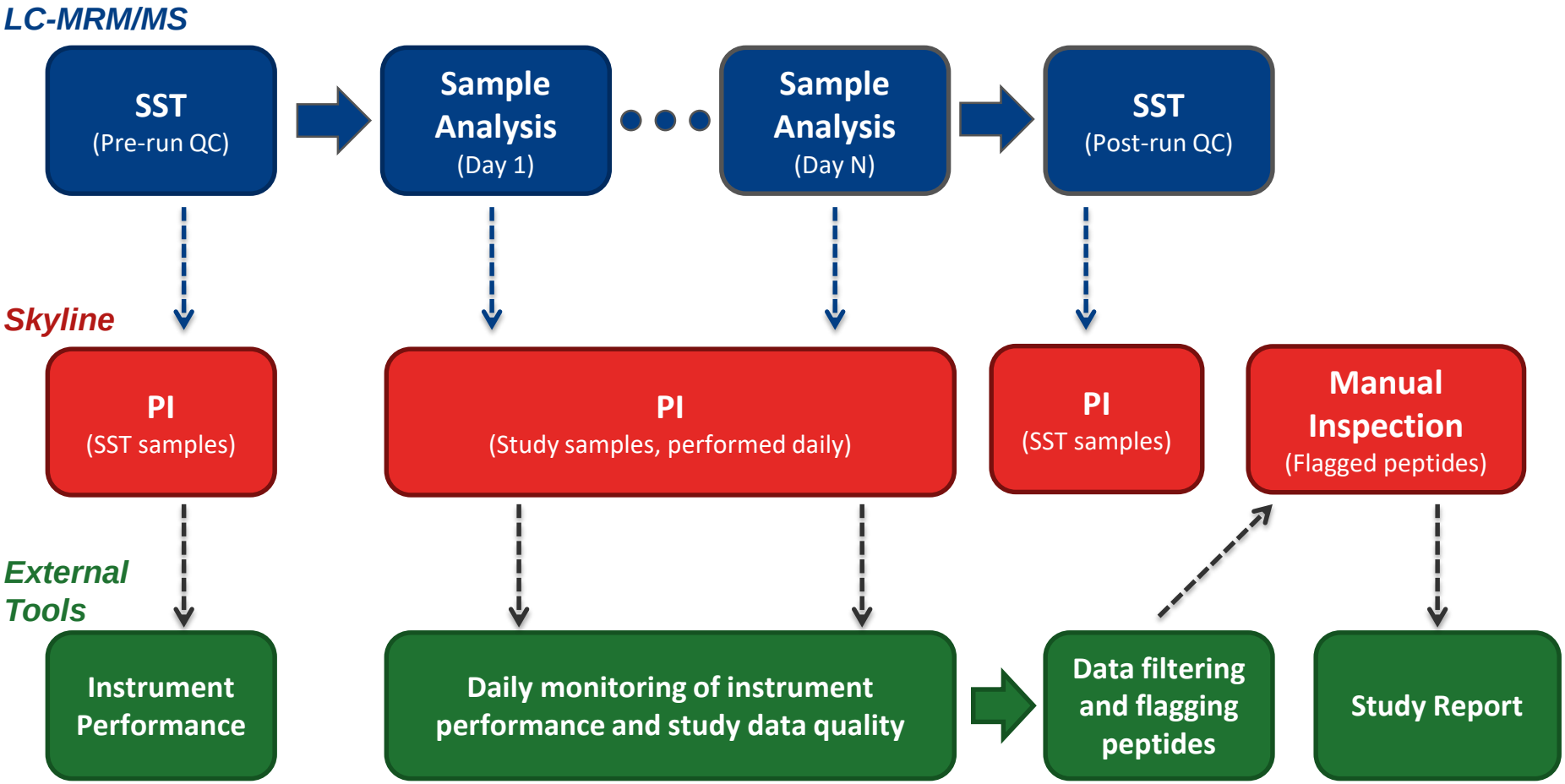
- Study samples (n= 50 to 2000)
- PQCs (pooled matrix) to assess process reproducibility

- Samples are denatured (e.g. TFE)
- R/A (optional)
- Digested with trypsin O/N at 37ºC

- crude SIL

- 4 transitions per peptide; 2 Light (L) + 2 Heavy (H)
- Monitoring of ≤625 peptides/30 min run
- Hundreds of thousands of peaks need to be integrated

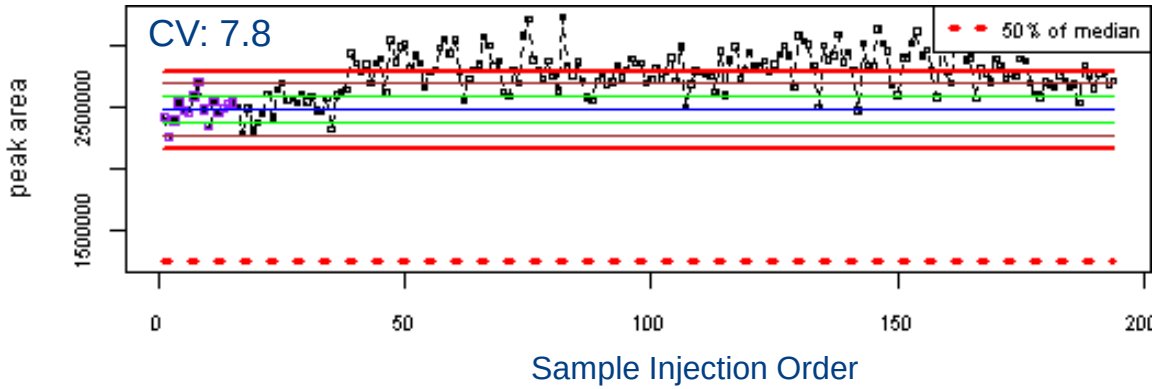
New Data Analysis Workflow



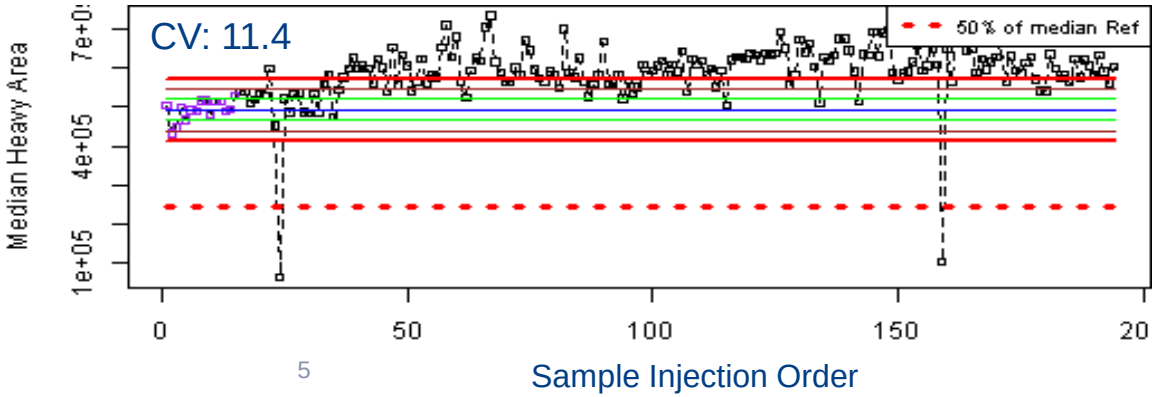
Instrument Performance Monitoring

- Raw data is imported daily into Skyline for peak integration
- Data is exported into a printable report (modified version of the SProCoP external tool)
- Various parameters are evaluated against pre-set criteria:

1. Peak area of the 5 ISP in each sample

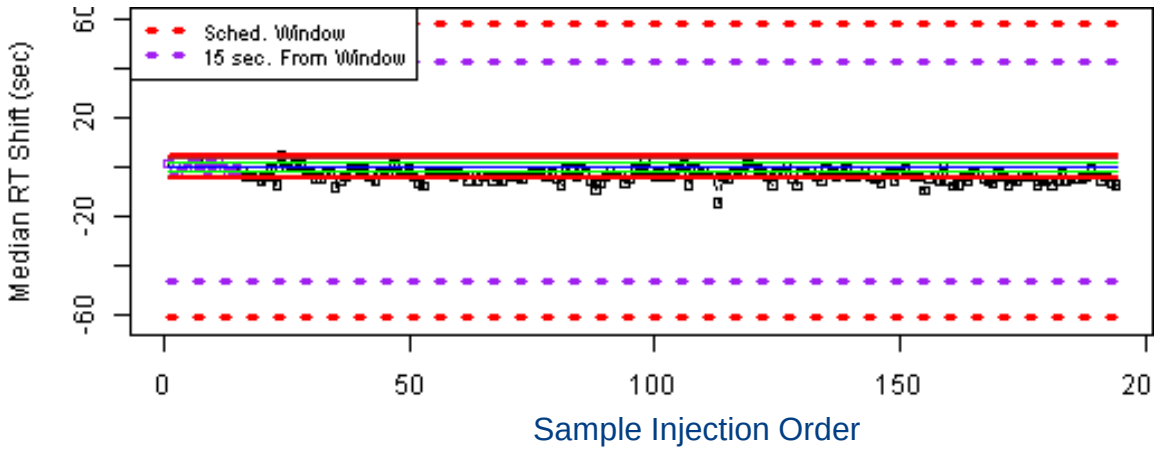


2. Median Peak Area of the SIL in each sample

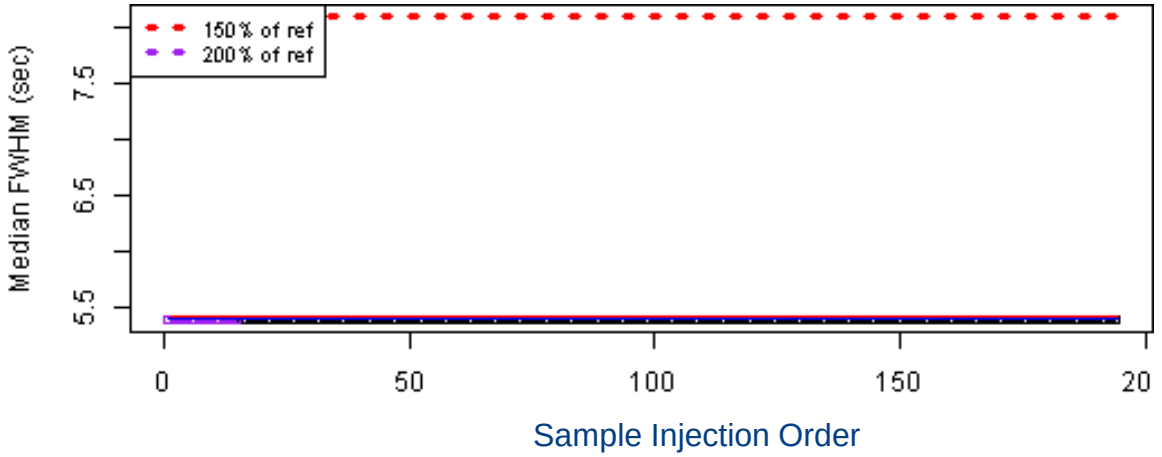


Instrument Performance Monitoring (cont'd)

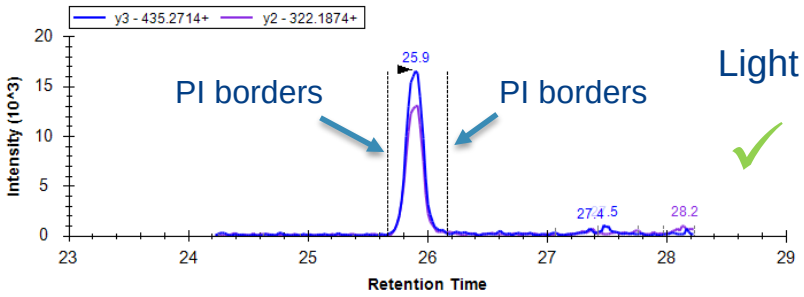
3. Median RT shift

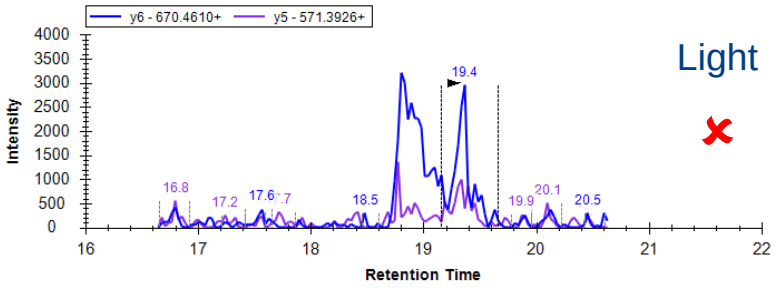


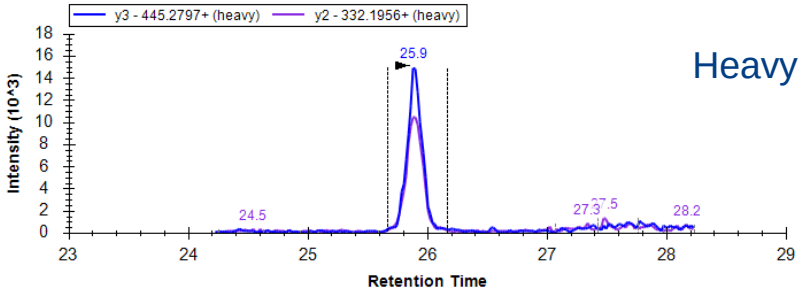
4. Median Peak Width

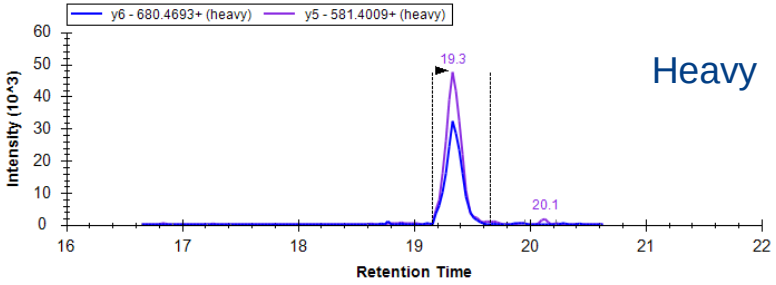


Data Filtering









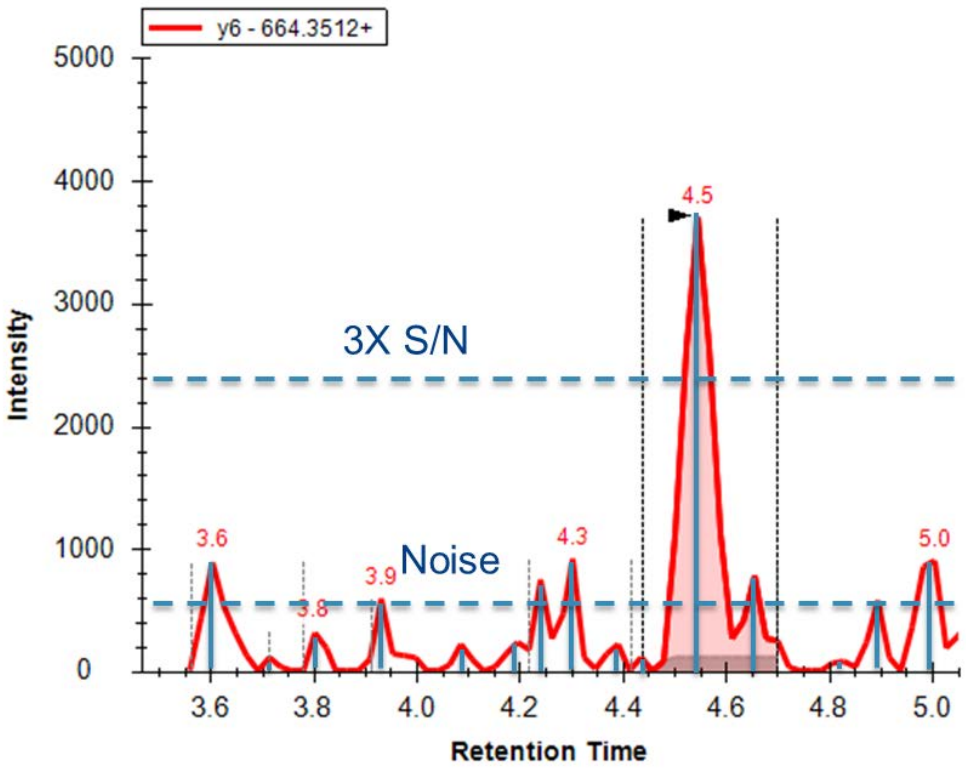
External tools (developed in-house):

- S/N ratio Light > 3
- PA ratio of H transitions matches PA ratio of L transitions
- H and L transitions co-elute

→ Peptides not meeting all 3 criteria are filtered out

S/N Ratio Determination

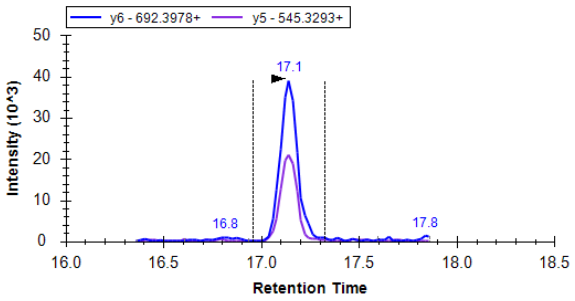
- Chromatograms are exported from Skyline for all transitions across multiple samples (≥ 20 samples)
- In-house developed external tool used to determine height of all peaks and noise spikes
- Median peak height calculated and threshold set at 3X S/N
- Data imported back into Skyline: Custom peptide annotation field “Noise”

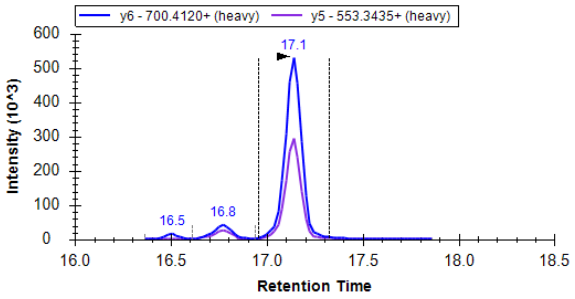


Manual Inspection: Interferences

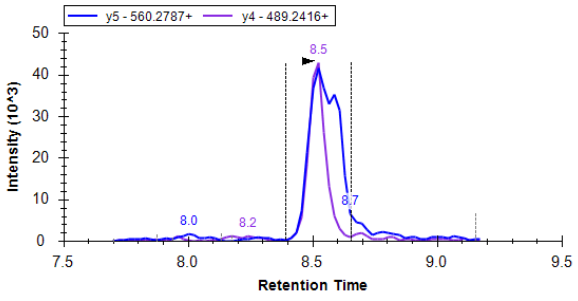
- List of peptides with potential interferences are exported using an in-house developed plugin
- Criteria for flagging peptides (must apply to ≥10% of study samples):
  1. PA ratio of H transitions do not match PA ratio of L transitions
  2. H and L transitions do not co-elute

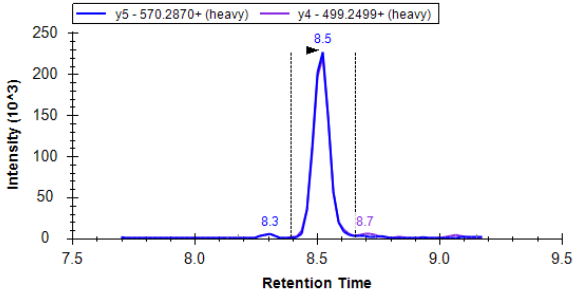
No change



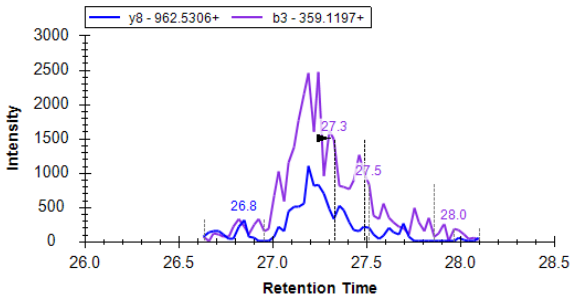


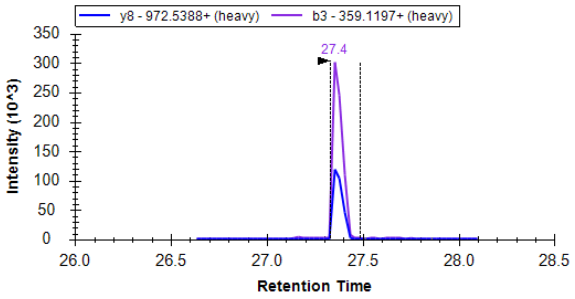
Exclusion of transition





Exclusion of peptide

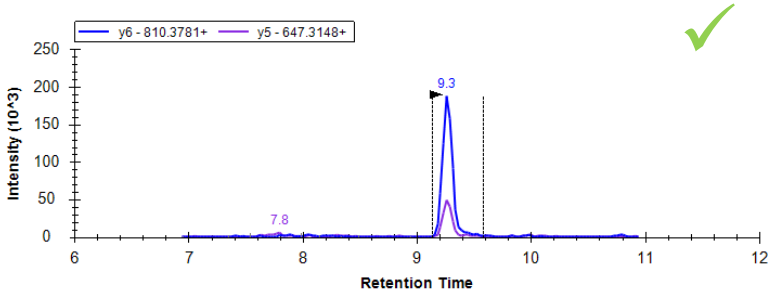


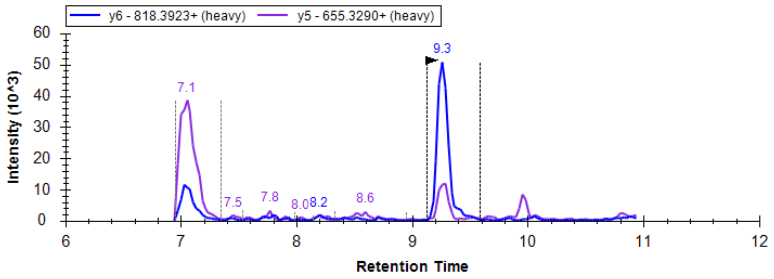


Manual Inspection: Incorrect Peak Integration

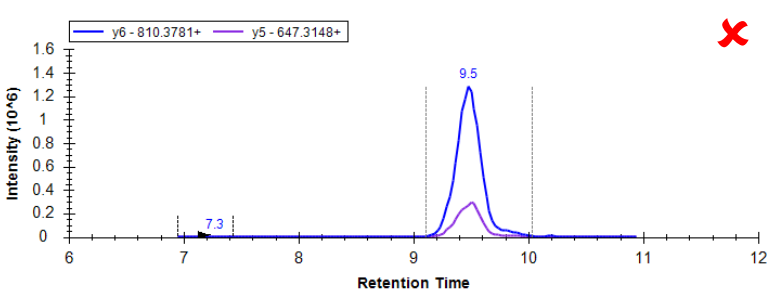
- List of peptides with potentially incorrect peak integration is exported using an in-house developed plugin
- Criteria for flagging:
  1. Peptides with significant Heavy PA variability (2 x median CV)
  2. Peptides with significant variability in peptide elution order

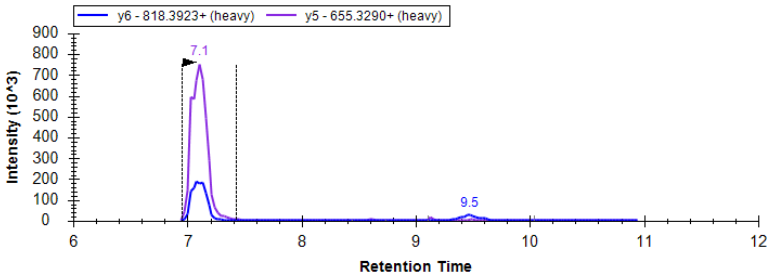
Sample 1





Sample 2





Reporting of Study Data

- An excel-based Study Report is exported using an in-house developed plugin
- Results are summarized over multiple tabs

Summary of Results tab (Study Quality):

Detection Rate

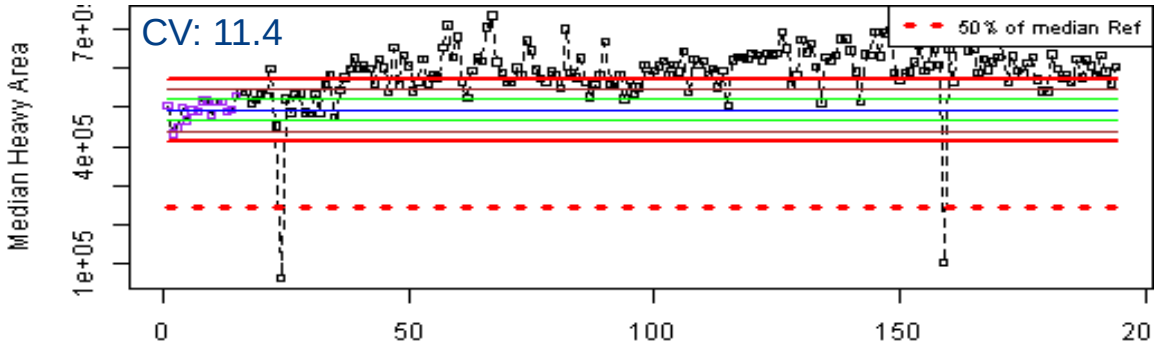
		Overall		Detection Rate by % of study samples							
		Detected		90% - 100%		50% - 90%		10% - 50%		>0% - 10%	
	Monitored	#	%	#	%	#	%	#	%	#	%
Proteins	139	113	81.3%	58	41.7%	15	10.8%	20	14.4%	19	13.7%
Peptides	582	347	59.6%	165	28.4%	53	9.1%	66	11.3%	61	10.5%
Transitions	1,164	680	58.4%	316	27.1%	106	9.1%	132	11.3%	122	10.5%

Reproducibility of 5 ISP and pQCs

Depletion Batch	Peak Area CV ISP (pQCs and study samples)							Peak Area CV pQCs		
	ISP-1	ISP-2	ISP-3	ISP-4	ISP-5	Median		L/H Ratio	Light	Heavy
1	8%	10%	11%	10%	9%	9.8%		7.6%	14.6%	11.4%
Overall	8%	10%	11%	10%	9%	9.8%		7.6%	14.6%	11.4%

Reporting of Study Data (cont'd)

Median PA of the SIL in each sample

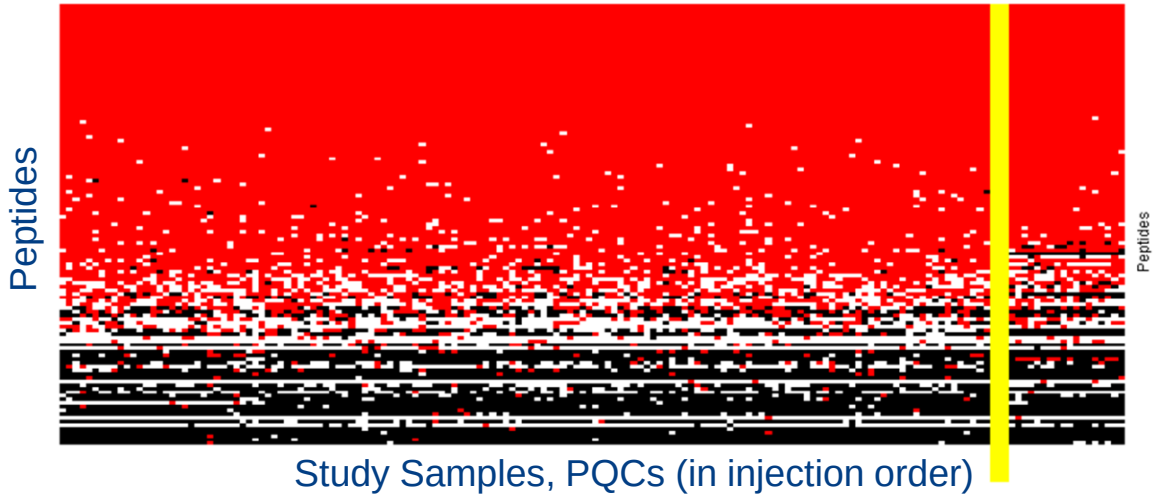


Heatmap across all samples

Detected

NA

<3 S/N



Reporting of Study Data (cont'd)

CAPRION

Filtered PA Ratio Results tab:

Protein.Name	Peptide.Modified.Sequence	Fragment.Ion	Sample #1	Sample #2	Sample #3	Sample #4	Sample #5	Sample #6
Protein A	ASLTIDEK	y3+	9.83410023	14.6949006	19.9243407	11.2337561	9.18223317	10.4752962
Protein A	ASLTIDEK	y6+	9.7358007	13.439119	17.7922937	12.0973524	10.0329146	12.6817619
Protein A	LQHLETELTHDIITK	y2+	2.07349952	2.83072198	4.15701409	2.89761766	2.0608139	3.13178998
Protein A	LQHLETELTHDIITK	y7+	1.96794371	3.37761295	3.96532196	2.70663988	2.12940889	2.66311158
Protein B	LSITFTYDLK	y6+	9.34168192	13.116458	18.5343573	11.2549811	8.95644314	11.8578946
Protein B	LSITFTYDLK	y7+	9.92424696	12.4944031	18.5707892	11.9228943	8.61060008	12.5523645
Protein B	SASLHLPK	y2+	8.761437	12.7165395	16.0234417	9.30092358	NA	10.9246092
Protein B	SASLHLPK	y6+	8.58945664	14.4568927	17.8575282	9.81343416	NA	11.7541854
Protein B	SFSTFADLSFSTEEAPLK	y3+	12.5298775	19.2551303	26.4536931	14.1550018	10.9527736	18.0541911
Protein B	SFSTFADLSFSTEEAPLK	y9+	13.1324994	17.6384837	25.8739611	15.8639186	10.0905355	16.9059759
Protein C	SESLEQEAATER	y7+	1.29434153	1.07994852	1.0398361	0.74750276	0.75952582	0.97628816
Protein C	SESLEQEAATER	y8+	1.31836867	0.99262946	1.1621354	0.86303515	0.85085488	1.02078465
Protein C	DSEFYIYAR	y3+	0	0.03717173	0	0.02919364	0.0305092	0.01713836
Protein C	DSEFYIYAR	y6+	0	0.04079924	0	0.02347657	0.03722721	0.01680939
Protein D	EFSSTSSTLTK	y6+	0.23905573	0.14096781	0.15754895	0.09226716	0.16609643	0.09507662

- A value of 0 is reported if both L transitions are not detected >LOD
- A value of NA is reported for a peptide if signal >LOD but for which other criteria is not met (co-elution or transition ratio criteria)

Summary

- A Skyline peak integration workflow in combination with in-house developed external tools was developed and is used for:
  - Assessment of instrument performance prior to and after sample analysis
  - Monitoring of instrument performance during sample analysis
  - Analysis of raw data (peak integration, flagging of peptides and data filtering)
  - Study report generation

- This workflow allows for analysis of large datasets within 2 days instead of weeks

Acknowledgements

- Carey Sheu
- Laetitia Cortes
- Laura McIntosh
- Lorella Di Donato
- Marija Mentinova
- Oliver Gingras
- Rudolf Guilbaud
- Seydina Lo