ABRF 2018
Annual Meeting
April 23-25 2018

Progress Establishing a Benchmarking and QC Resource for Shotgun Proteomics Analysis

Emily Chen[1], Alexandre Rosa Campos[2], LeeAnn Higgins[3], Achim Treumann[4], Theresa McLaughlin[5], Allis Chien[5]

[1]Columbia University, New York, NY; [2]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA; [3]University of Minnesota, St Paul, MN; [3]Newcastle University, Newcastle, UK; [5]Stanford University, Stanford, CA

Introduction

We report preliminary results from Phase 2 of a study to investigate and promote intra- and inter-laboratory reproducibility of quantitative proteomic LC-MS/MS analyses. The availability of Standard Reference digests and cloud-based repositories for large raw datasets have created new possibilities for sharing and mining raw data from various instruments in labs all over the world. Past studies of proteomic performance metrics have focused on the retrospective evaluation of collected data. In this study we seek to identify data processing tools, including ID-free quality metrics, to support a proactive approach, with the goal of allowing researchers to benefit from this Community Sourced Dataset and to catch and correct instrument problems *before* sample data is acquired.

Traditionally, little if any information on chromatographic performance is reported for shotgun proteomics studies. In 2010 NIST published a suite of Quality Control metrics for LC/MS/MS proteomics experiments, identifying chromatographic performance among the significant contributors to variability[1]. Numerous studies including Phase I of this ABRF-WIN initiative show that liquid chromatography has a significant influence on proteomics data quality and reproducibility.

Materials and Methods

In Phase 2 of this study we employed Indexed Retention Time Internal Standard Peptides (**iRT** peptides) from Biognosys (Schlieren, Switzerland) to facilitate method transfer and evaluation of reproducibility. Per the Biognosys website, "the **iRT Standard** contains eleven non-naturally occurring synthetic peptides in a pooled mix. Peptides have been carefully optimized for stability, sensitivity and even retention time spacing over the gradient."[2] Table 1, provided by Biognosys, lists the retention time index and expected precursor ion for each internal standard.

27 laboratories submitted raw files for Phase 2 of the study, out of a total of 52 requesting samples.

Instrument platforms used for data collection included the Orbitrap Elite, Fusion, Lumos, Velos, Q-Exactive, and Triple TOF.

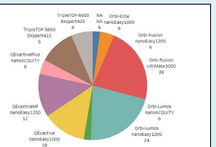

Fig 1. Instrumentation at Participating Laboratories

Each participant received 2 samples: the iRT peptide internal standard mixture and a HeLa cell lysate digest (Thermo Fisher) with instructions to prepare solutions of internal standard only and HeLa lysate spiked with internal standards. The target concentration for the iRT-HeLa mixture was 0.1ug HeLa lysate per injection, and participants were instructed to adjust the volume and concentration of the injection solutions to suit their instrumentation and sample injection apparatus.

Table 1. The iRT peptide internal standards.

	Standard	Q1	iRT
1	LGGNEQVTR	487.257	-24.92
2	GAGSSEPVTGLDAK	644.823	0.00
3	VEATFGVDESNAK	683.828	12.39
4	YILAGVENSK	547.298	19.79
5	TPVISGGPYEYR	669.838	28.71
6	TPVITGAPYEYR	683.854	33.38
7	DGLDAASYYAPVR	699.339	42.26
8	ADVTPADFSEWSK	726.836	54.62
9	GTFIIDPGGVIR	622.854	70.52
10	GTFIIDPAAVIR	636.869	87.23
11	LFLQFGAQGSPFLK	776.93	100.00

Table 2. Analytical Workflow.

	LC/MS/MS Acquisitions
1	Clean source, calibrate MS, wash column
2	iRT alone: WIN 45 minute method (2 INJECTIONS)
3	HeLa + iRT Mix, WIN 130 minute method (3 INJECTIONS)
4	iRT alone, WIN 45 minute method (1 INJECTION)
5	HeLa + iRT Mix - Lab LC method (3 INJECTIONS)
6	iRT alone, WIN 45 minute method (1 INJECTION)

Data acquisition followed the workflow in Table 2. The first two injections were iRT-only using the method described in Table 3. Next, three injections of the iRT-HeLa-mix using the method described in Table 4, followed by 1 iRT-only injection. Participants had the option of running another 3 iRT-HeLa mix injections using their own standard laboratory gradient for comparison, with the caveat that the MS acquisition time could not exceed 130 minutes. The gradient tables assume solvent A is 100% H_2O + Formic Acid. Participants were instructed to acquire data dependent product ion spectra only, to adjust necessary parameters specific to their instrument setup (e.g. solvent percentages, flow rate, equilibration time) but to maintain the MS acquisition time and gradient profile.

Effort needed to incorporate the iRT peptides was kept to a minimum. Participants were directed to the vendor's website for information about the iRT peptide standards; however, no example chromatograms were included in the protocol, and there was no expectation that participants would try to reproduce a particular chromatographic elution profile.

All mass spectra were analyzed with MaxQuant software version 1.5.5.1. MS/MS spectra were searched against the *Human* Uniprot protein sequence database (version January 2017), the concatenated iRT peptides and GPM cRAP sequences (commonly known protein contaminants). Precursor mass tolerance was set to 20ppm and 4.5ppm for the first search where initial mass recalibration was completed and for the main search, respectively. Product ions were searched with a mass tolerance 0.5 Da. The maximum precursor ion charge state used for searching was 7. Carbamidomethylation of cysteines was searched as a fixed modification, while oxidation of methionines and acetylation of protein N-terminal were searched as variable modifications. Enzyme was set to trypsin in a specific mode and a maximum of two missed cleavages was allowed for searching. The target-decoy-based false discovery rate (FDR) filter for spectrum and protein identification was set to 1%. In addition, second peptide search and dependent peptide search were enabled. Spotfire version 7 and R version 3.2.2 were used for data visualizations.

Table 2. WIN 45 min method for iRT-only

Time	Flow rate	%B [100% ACN]	%B [95% ACN]	%B [80% ACN]
0	300	2	2.1	2.5
2	300	2	2.1	2.5
5	300	10	10.5	12.5
20	300	30	31.6	37.5
22	300	90	94.7	112.5
29	300	90	94.7	112.5
30	300	2	2.1	2.5

Table 3. WIN 130 min method for iRT-HeLa

Time	Flow rate	%B [100% ACN]	%B [95% ACN]	%B [80% ACN]
0	300	2	2.1	2.5
2	300	2	2.1	2.5
5	300	5	5.3	6.3
20	300	10	10.5	12.5
90	300	20	21.1	25
105	300	30	31.6	37.5
106	300	90	94.7	100

Results and Discussion

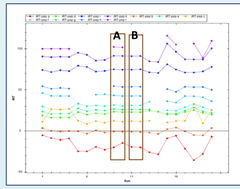

Figure 2. Tracking relative retention time for several groups of technical replicates on different systems. Box A highlights runs where all 11 iRTs were detected, while Box B highlights runs where an iRT is missing, warranting further investigation.

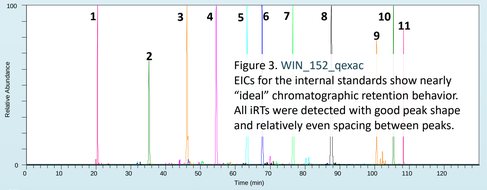

Figure 3. WIN_152_qexac
EICs for the internal standards show nearly "ideal" chromatographic retention behavior. All iRTs were detected with good peak shape and relatively even spacing between peaks.

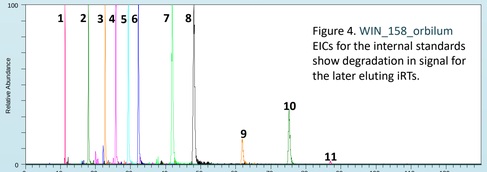

Figure 4. WIN_158_orbilum
EICs for the internal standards show degradation in signal for the later eluting iRTs.

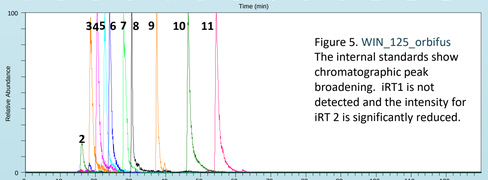

Figure 5. WIN_125_orbifus
The internal standards show chromatographic peak broadening. iRT1 is not detected and the intensity for iRT 2 is significantly reduced.

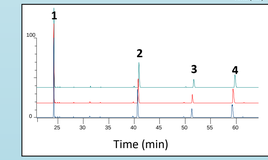

Figure 6. WIN_149_qexachf
The internal standards show subtle retention time drift increasing with retention time.

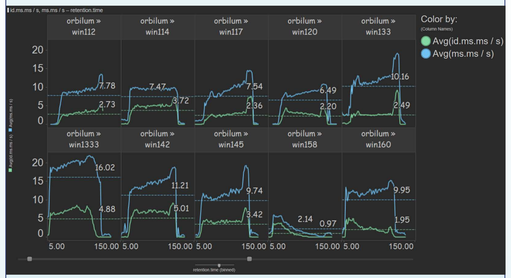

Figure 7. Plots of the average number of identified MS/MS per second (green) and average number of MS/MS per second (blue). Peptide identification efficiency decreases as the gap between the 2 lines increases.

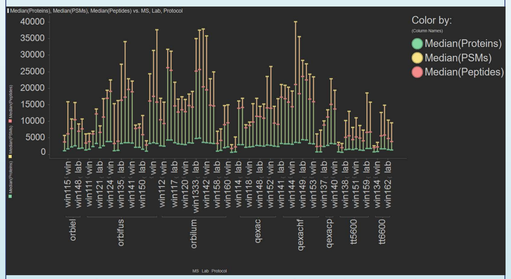

Figure 8. Plots of the median number of proteins identified per lab-MS-Protocol (green), median number of peptide spectral matches per lab-MS-Protocol (yellow), and median number of peptides identified per lab-MS-Protocol (red).

Figure 9. MaxQuant Dependent Peptide search. After the initial FASTA database search, a second search with an expanded list of post translational modifications was performed against a limited database consisting of the peptides identified in the first search.

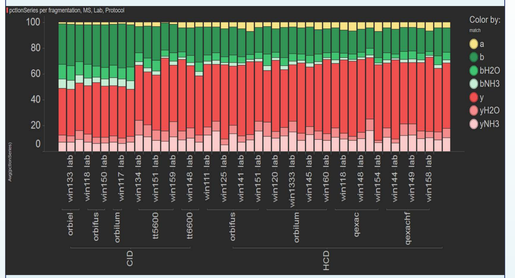

Figure 10. Each bar shows the relative percent of fragment ions in PSMs across instruments for CID and HCD

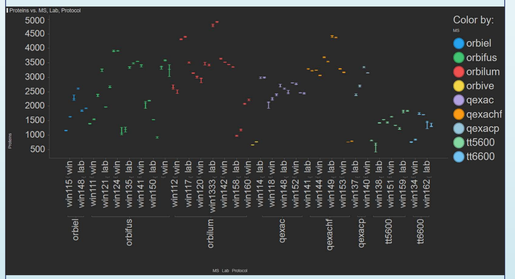

Figure 11. Number of proteins identified per run. Technical replicates of each protocol per lab are grouped to show intra-laboratory reproducibility.

Figure 12. Reproducibility vs Sensitivity. The Number of Protein Groups Identified and the Protein Intensity Coefficient of Variance (95th Quantile) were calculated for each set of three technical replicates. The Protein Intensity CVs are plotted on the y-axis, with the lowest values (signifying best intra-lab reproducibility) at the top. The Numbers of Protein Groups Identified are plotted along the x-axis, with values increasing from left to right signifying greater sensitivity. The size of each point indicates the fraction of proteins consistently identified in all three replicates. For example, a set of three technical replicates with 5,000 total protein groups where 4,000 are identified in all 3 replicates has a 0.80 identification reproducibility score.

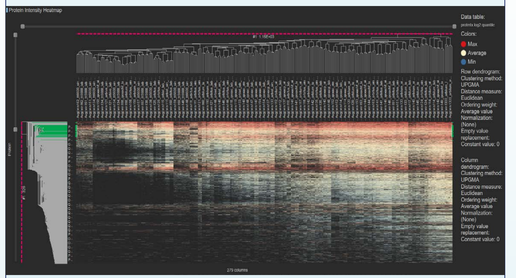

Figure 13. Portion of a heatmap displaying the proteins identified per set of technical replicates from most to least common.

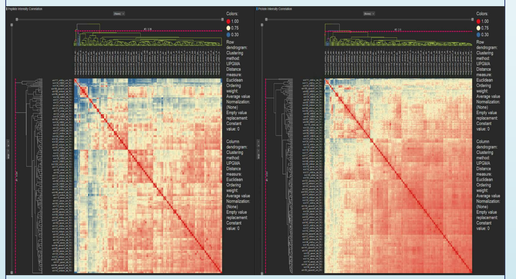

Figure 14. 2D peptide (left) and protein (right) intensity correlation. Peptide or protein intensity data were normalized across all laboratories, then Pearson correlation was calculated pairwise across all runs.

Conclusions

The availability of Standard Reference Materials like the HeLa lysate digest plus large dataset repositories like MASSive at UC San Diego have created new opportunities for method optimization and standardization. However, the heterogeneity of nanoflow LC/MS/MS instrument platforms makes mining this data extremely challenging. Internal standards like the Indexed Retention Time (iRT) peptides provide a simplified suite of analytes to facilitate cross platform comparisons.

An interactive display of this poster's data visualizations and many more is available at the ABRF website on the Workflow Interest Network Research Group page:

https://abrf.org/research-group/workflow-interest-network-win

Acknowledgements

The ABRF Workflow Interest Network would like to thank the following for their contributions to this study:

- All the laboratories and scientists who participated in and provided data for the study
- Nuno Bandeira and the MassIVE team at UC San Diego for use of their platform for data storage and for helpful discussions on data processing.
- Biognosis for providing the iRT internal standards, the QuiC software platform for data analysis, for help shipping the standards, and for assistance with data processing.
- Thermo Fisher for the supplying the HeLa lysate digest standard

References

1. Rudnick PA, et al. Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. Mol Cell Proteomics. 2010 Feb;9(2):225-41
2. Reference sheet for iRT peptides from Biognosis website https://biognosys.com/media.ashx/irt-kit-reference-sheet.xls