ASMS workshop 2023

Data Independent Acquisition Interest Group

Data Independent Acquisition: After the Acquisition
Tue, Jun 06 5:45pm - 7:00pm (Central) in Ballroom C

Presiders: Lindsay K. Pino (lpino@talus.bio); Lukas Reiter (lukas.reiter@biognosys.com)
INCOMING PRESIDER: Qin Fu (Qin.fu@cshs.org) replacing Lindsay Pino

Invited Panelists

Fu, Qin <Qin.Fu@cshs.org>
Devon Kohler <kohler.d@northeastern.edu>
Qing Zhong <qzhong@cmri.org.au>
Tenzer, Stefan <tenzer@uni-mainz.de>

Summary of Discussion

The DIA workshop began with brief introductions from each panelist, describing how they work with DIA data after it's been acquired, from processing raw files all the way through statistical testing and biomarker discovery. We then moved into Q&A to spark discussion.

Panelists all agreed that it is crucial during DIA studies to run QC, whether it's an external sample or an internal surrogate, even better to run both! There was also agreement that the best QC that one can do is simply looking at the data often and babysitting the instrument while it is acquired. As nano LC can still be challenging, sometimes looking at RT/iRT can be useful to spot LC issues. Panelists and audience admitted that analyzing PTM DIA is very tough and not a regular thing despite publications showing its merits, so further work on processing PTM DIA data is needed before it can be adopted wider.

Despite (or perhaps due to?) the plethora of software options for processing DIA, panelists emphasized that picking the "right" analysis software is still difficult, especially how to know what's "right". The panelists mentioned using an LFQBench approach with hybrid proteomes (samples composed from two or more proteomes combined in known ratios) several times as a great way to assess the system, the methods, and the analysis software, because it benchmarks not only detections but quant.

Some of the biggest hurdles for DIA were surprisingly logistical. For example, the size of data! Especially on new instruments, like the timsTOF and Astral, files are large so storage, transfer, analysis has been a logistical bottleneck in the way of processing and exploring the data. Some panelists said that using cloud-based services, while “infinitely” scalable, rack up high costs due to the transfer of data files up and down from the cloud. For statistical testing after processing, reading in the quant matrix into i.e. R or Python dataframes can quickly cause memory issues as well.

Also comparability of data across cohorts and different setups is challenging. There are multiple computational methods to remove batch effects. However, choosing the right one is challenging. And sometimes the more complicated methods don’t perform better compared to the simpler ones.

Yet another challenge is the most optimal statistical analysis of the data. Due to the layered character of bottom up proteomics data statistical analysis can be challenging. However, this is not necessarily an exclusive challenge of DIA data but rather of all bottom up proteomics data.