

## ASMS Workshop 2022: DIA-MS and getting to more biology

**Organizers: Jennifer Van Eyk, Cedars-Sinai Medical Center and Michael MacCoss, Univ Washington**

### Goal and Overview

The goal of this workshop is to instruct and discuss with scientists / researchers who are currently applying data-independent acquisition-mass spectrometry (DIA-MS), data acquisition and data analysis workflows that maximize the reproducible quantification of the proteome in a variety of samples.

The program will use a combination of didactical lectures, hand-on tutorials and focused discussion panels to maximize interactions between workshop participants and to drill down into what is required to obtain reproducible DIA-MS data sets. This includes focusing on system suitability, QC, maximizing MS data acquisition parameters and establishing transparent data analysis. The heart of the workshop is focused on current and upcoming data analysis pipelines. Here we propose a hands-on tutorial that will allow workshop participants to compare current peptide- and spectrum-centric analysis packages using a model DIA-MS data sets generated on the same sample set acquired on 3 different instrument using DIA-MS. These model data sets will allow participants to dig into the robustness of the data using either Peptide- and Spectrum-central data analysis approaches. This session will be followed up by a discussion on data handling of DIA-MS data sets and introduce emerging data analysis packages that are currently being developed. Furthermore, we will concentrate on the logistics required for generate reproducible data in large cohorts/studies as and emerging approaches in DIA-MS, specifically its use for quantification of protein post-translational modifications. Finally, the group will discuss the next steps in application of DIA-MS and what will be required to push this method to new frontiers. All speakers should be present for the full program to promote interaction. "Speak and leave" will not acceptable.

### Workshop Focus

- Main focus: This workshop focus on user-focused learning and discussion on DIA-MS applications in proteomics specifically on the technical requirements for the generation of reproducible and scalable DIA-MS data sets. It will emphasis data acquisition and data analysis including peptide- and spectrum-centric approaches, exploration of the current state-of-the-art DIA-MS application in post-translational modifications and large scale studies.
- The workshop includes didactical lectures, hands-on tutorial and discussion panels to explore the current challenges of DIA-MS accompanied by discussion on solutions. A highlight is a hands-on session centered on data analysis using a choice of a model DIA-MS data sets from the same samples run on 3 different vendor instruments that the participants can analyze using peptide- and spectrum-centric data analysis package. The model DIA-MS data set will be already generated and analyzed by experts prior to meeting to allow the participates to compare and contrast their data analysis workflows. Comparison to several emerging data anlasyis approaches will be presented.
- The workshop participants should be DIA-MS scientists beit bioinformatics, biologists, clinician and/or mass spectrometrists.
- The workshop may be virtual or in-person.

### Suggested Schedule.

#### *Day 1: 9-12 morning session: DIA-MS introduction.*

The goal of this session is to introduce the workshop participate to differences in DIA-MS data acquisition and data analysis while setting up comparison of peptide- verses spectrum-centric data analysis of DIA-MS generated data. 30-45 mins lectures with 1 hr discussion.

- Introduction to DIA concepts and drive for reproducible large scale datasets. Comparison of different data acquisition methods for Q-Orbitrap, Q-TOF, IMS-TOF, etc. 45 min. (Brian Searle, ISB)
- Peptide-Centric approaches for data analysis (e.g. Open-SWATH, Skyline, EncyclopeDIA, Spectronaut). 30 min. (Lindsay Pino, UPenn)
- Spectrum-Centric Approaches for data analysis (e.g. Pular-X, DIA-Umpire). 30 min. (Alexey Nesvizhskii, UMich)
- Small group discussion around current challenges and issues in DIA-MS specifically around robustness and scalability. 1 hr

**Day 1: 1-5 afternoon session: Digging into the nitty gritty of software and analytics around DIA-MS.**

The goal of this session is for workshop participants to examine DIA-MS data sets using the two primary approaches and to understand the underlying assumptions and limitations of peptide- and spectrum-centric data analysis approaches and how new emerging methods may address some of these challenges. This session is comprised of two parts: The first part is a 3 hrs. hands on tutorial using pregenerated data sets followed by 60 min small group discussion. The second part is on new and emerging approaches and includes 30 min talks on their approach and results from analyzing the same data set used in the tutorials followed by a 1 hr panel discussion.

- *Hands-on tutorial looking at Peptide-Centric data in Skyline.* 60 min. (Brendan MacLean, UW)
- *Hands-on tutorial looking at Spectrum-Centric data.* 60 min. (Niveda Sundaraaman, CSMC)
  - Both tutorials will use a pre-generated model DIA-MS data sets from ThermoFisher Orbitrap, Sciex 6500 Triple TOF and or Bruker TimsTOF instrument will be provide to the participants. Prior to the workshop, this data set will be generated on the same samples by organizers and analyzed so existing data set can be compared. Each participate can choice to run data set via peptide- or spectrum-centric data analysis using different existing tools.
- Small group discussion around results from the different data analysis approaches with report to the general group. 1 hr.
- *New and emerging approaches.* In this session, each new approach will outline the novelty and the pros and cons compared to peptide- and spectrum centric approaches based on showing the data generated on the same model data set as above.
  - New Approach: DeepSearch 45 mins (Gautam Saxena, DeepDIA)
  - New Approach: AI 45 mins (Robin Park, Scripps)
- Panel discussion on current challenges in data acquisition and data handling (30 mins). The focus is on establishing the issues around misinterpreting spectra and establishing what should be the gold standard. 30 mins (lead by Hanes Roste, UT (not sure))

**Day 2: 9-12 morning session: Establishing Reproducible DIA-MS Research**

The goal of this session is for workshop participants to explore the study design and infrastructure that can assist in generating reproducible (large scale) DIA-MS data sets. 3 diadictal lectures (20 mins. each)

- Sample Preparation and automation. 20 min. (Rene Robinson, Vanderbuilt,)
- Setting up systems suitability, process control and QC for sample preparation. 20 mins. (Julia Robins, UW)  
The hope for a full system automation, sample tracking, data analysis and batch correction. 20 mins. (Sarah Parker, CSMC)
- Small group discussion (1 hr) around where there are challenges and potential solutions based on series of targeted questions. Questions will include:
  - What are criterial for quantification if based on a single peptide and then extract those criteria to if you have to quantify 20,000 peptides.
  - What are the criteria for comparing DIA-DA data sets obtained using different acquisitions or MS instruments (e.g. boxcar/overlap vs traditional acquisition on Thermo. E.g. Triple TOF vs Thermo vs Tims TOF instruments).
  - What is the strategy for using DIA-MS on PTMs that can be isolated in different DIA-windows. (e.g. mono-, di- verse tri-methylation).
  - What is the strategy for using DIA-MS on PTMs that cannot be isolated in different DIA-MS windows (e.g. citrullination).
- Discussion and brain storming by full group (1 hr)(Jenny Van Eyk, CSMC)

**Day 2: 1-5 afternoon session: Pushing DIA-MS into Large Scale Reproducible Science**

The goals of this session is to provide real life examples with the lectures focusing on the real Q and reproducibility data and highlight the lessons learnt. There will be a directed and moderated directed Q and A on what would you change now that you have done this large scale study (moderator. Lisa Jones (30 mins lecture with 10 min Q and A for each PI)

- 100s mouse models of neurological disease. 30 mins. (Mike MacCoss, UW)
- ISPC answer in solving ALS or large scale plasma projects. 30 mins (Jenny Van Eyk or one of her Trainees, CSMC)
- Plasma and AD biomarkers. 30 mins. (Rene Robinson, Vanderbilt)
  
- Panel discussion (all instructors and participants involved) will summary needs and next steps in DIA-MS application. 1 hr. (Lead by Sarah Parker, CSMC)

**At the end of the course, the participants should have a good understanding of:**

- Challenges and unmet need in the application of DIA-MS for protein quantification
- Existing and emerging data analysis pipeline and their pros and cons.
- New emerging areas of DIA-MS application including post-translational modifications
- Key concepts and innovations in area of large scale applications of DIA-MS including setting up automation, systems suitability and QC.