An ABRF Study to Evaluate Data-Independent Acquisition for Protein Quantification in Core Facility Settings

ABRF Proteomics Research Group
https://abrf.org/research-group/proteomics-research-group-prg

Identification of certain commercial equipment, instruments, software or materials does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

The Proteomics Research Group

Through the promotion of broad participation and scientific excellence, the PRG aims to raise awareness, knowledge and education about modern methods of protein analysis.

Current members:

- Pratik Jagtap (Chair) — University of Minnesota
- Laura Herring — University of North Carolina at Chapel Hill
- Joanna Kirkpatrick — Leibniz Institute on Aging, Germany
- LeRoy Martin III — Waters Corporation
- Mukul Midha — Institute for Systems Biology
- Benjamin Neely — National Institute of Standards and Technology
- Brett Phinney — University of California Davis
- Baozhen (Paul) Shan — Bioinformatics Solutions, Inc.
- Paul Stemmer — Wayne State University
- Yan Wang — University of Maryland
- Allis Chien (EB-liaison) — Stanford University

- *Contact prg.abrf@gmail.com*

Goal of Study

Our goal is to eliminate barriers to adoption, demonstrate where DIA is across platforms and cores.

In other words, what does DIA mean in a core as it relates to productivity and data quality/useability.

Goal of Study

Specific Goals

- Provide baseline methods across platforms
- Create and distribute a test sample that can benchmark performance
- Recruit diverse platforms and skill levels
- Collect data with goal of making public
- Analyze data along with industry partners

Intended Benefits of Study

Participant Benefits

- Develop working DIA methods
- Defined sample for self-evaluation
- Use available software to process data
- Compare your work anonymously to others in the area

Community Benefits

- Multi-platform multi-laboratory DIA data on the same sample set
- Anonymous DIA data made publicly available to help algorithm, workflow, and application development and benchmarking
- Serve as demo dataset for new-comers

ABRF Study Timeline

Sample

Blank 25 fmol 100 fmol

HeLa digest spiked with four non-endogenous proteins and iRT

Data Acquisition

- Acquisition parameters were provided
- Total DIA time < 24 h
- Given enough sample to generate library OR spectral library available for download

Software

Skyline

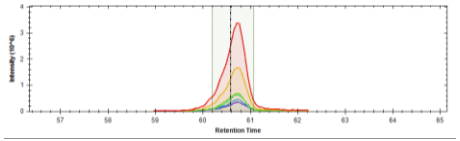
PEAKS

OpenSWATH

TPP

Participants had the option of using trial licenses from Spectronaut, Scaffold-DIA, and PEAKS DIA

Data analysis & interpretation



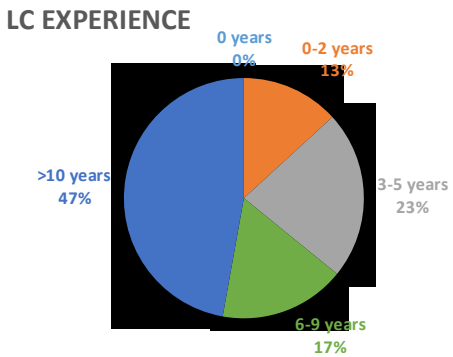
- PRG members analyzed data
- Participants encouraged to analyze as well

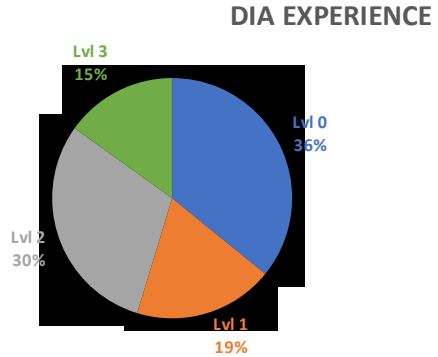
October 2018 | November 2018 | December 2018 | January 2019 | February 2019 | March 2019

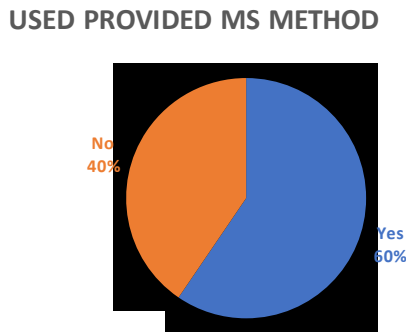
Study Participants: 63 labs, 20 countries, 16 US States

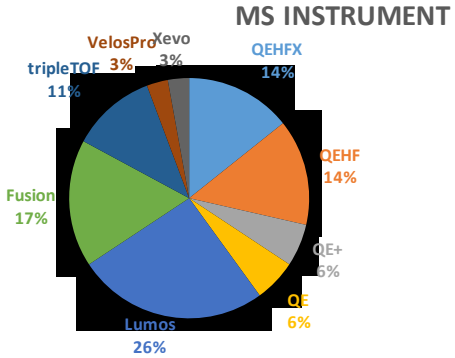
Study Participants

- 40 (63%) participants deposited data
- 35 data sets used for prelim analysis
- 53 survey responses (84%)
- Experience was broad
- Majority used provided acquisition method
- Majority of MS platforms were Thermo





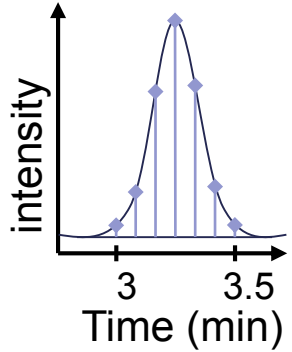


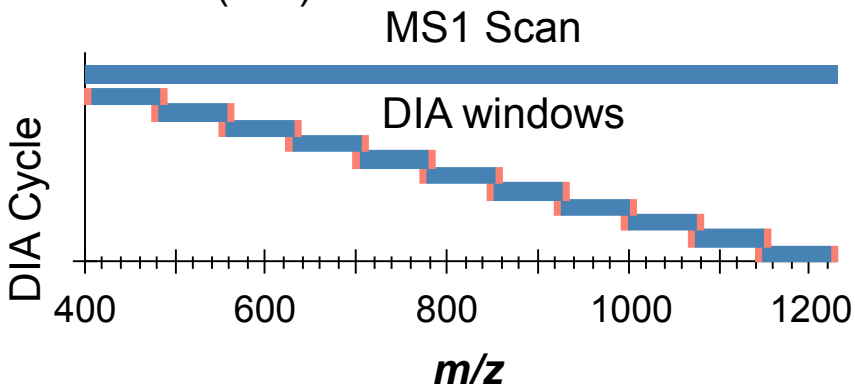


Study Method

Goal: create a base DIA method across platforms.

- Not the best, but standard starting method
- LC: two-step gradient lasting 110-130 minutes
- DIA: try to be at 3.5 sec cycle to be roughly 7-10 dppp if peaks are 30 sec at base
- 1 Da overlapping windows from 400-1200 m/z
- Window width was dependent on instrument scan speed



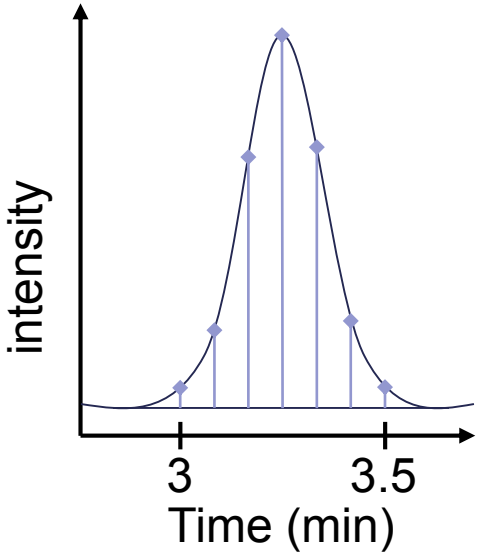


windows x MS2 acquisition time = cycle time

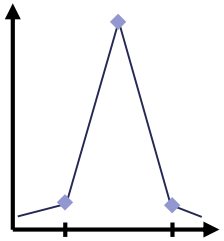
40 (21 *m/z* width, 400-1200 *m/z*, 1 Da overlap) x 60 ms = 3.5 s

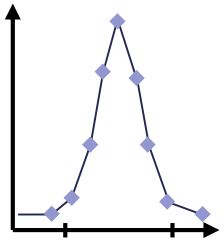
available at https://github.com/neely/PRG2018 or https://www.lcmsmethods.org/

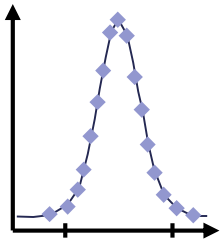
DIA Acquisition



ex. 30 s peak width at base
3.5 s cycle will collect ~8 DPPP







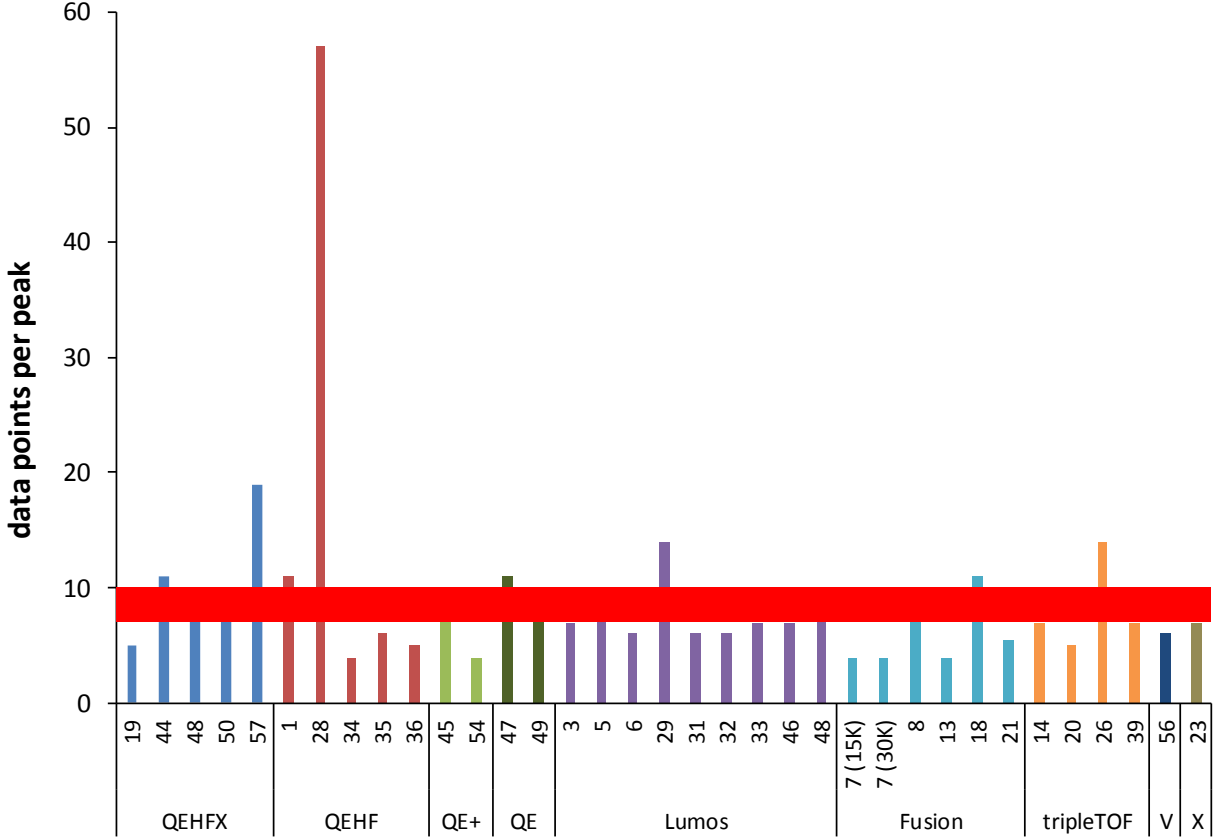
Data Points Per Peak (DPPP)

<7 DPPP = under sampling

7-10 DPPP = optimal sampling

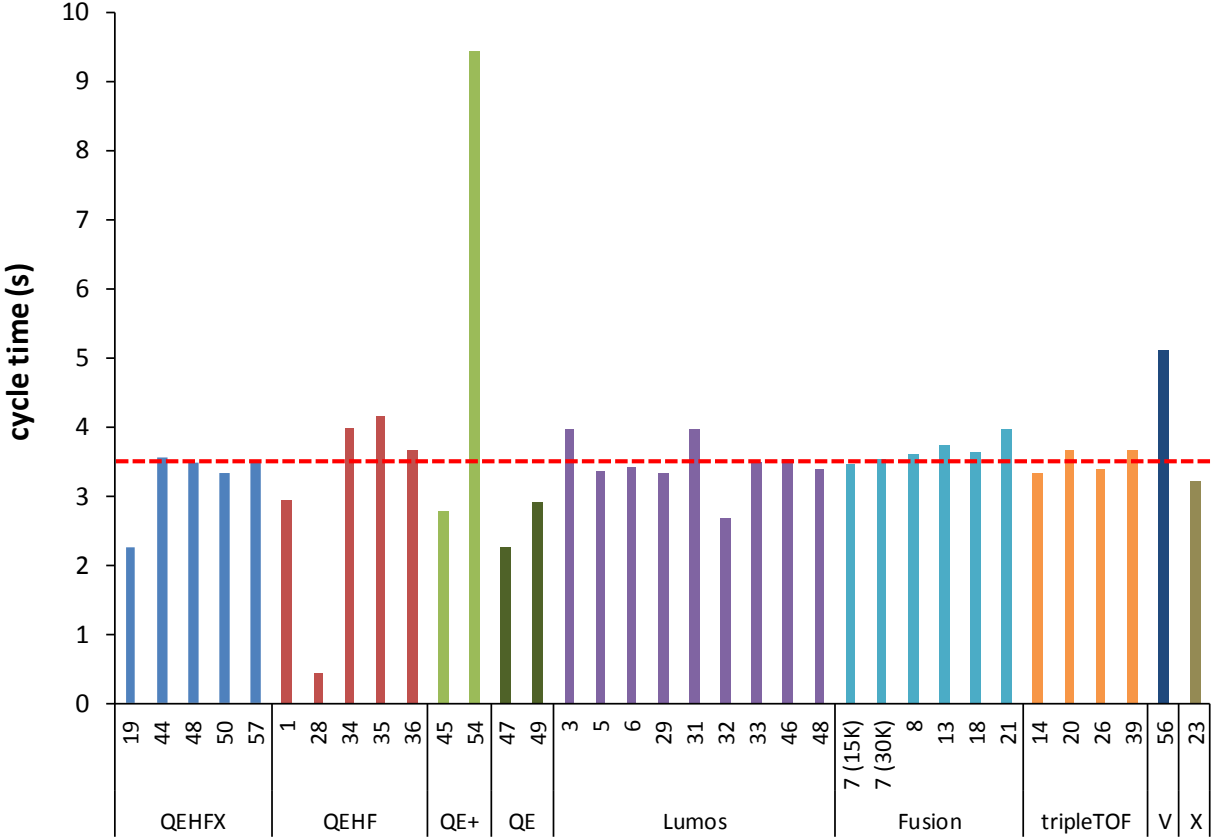
>10 DPPP = over sampling

Performance of Participants – DPPP



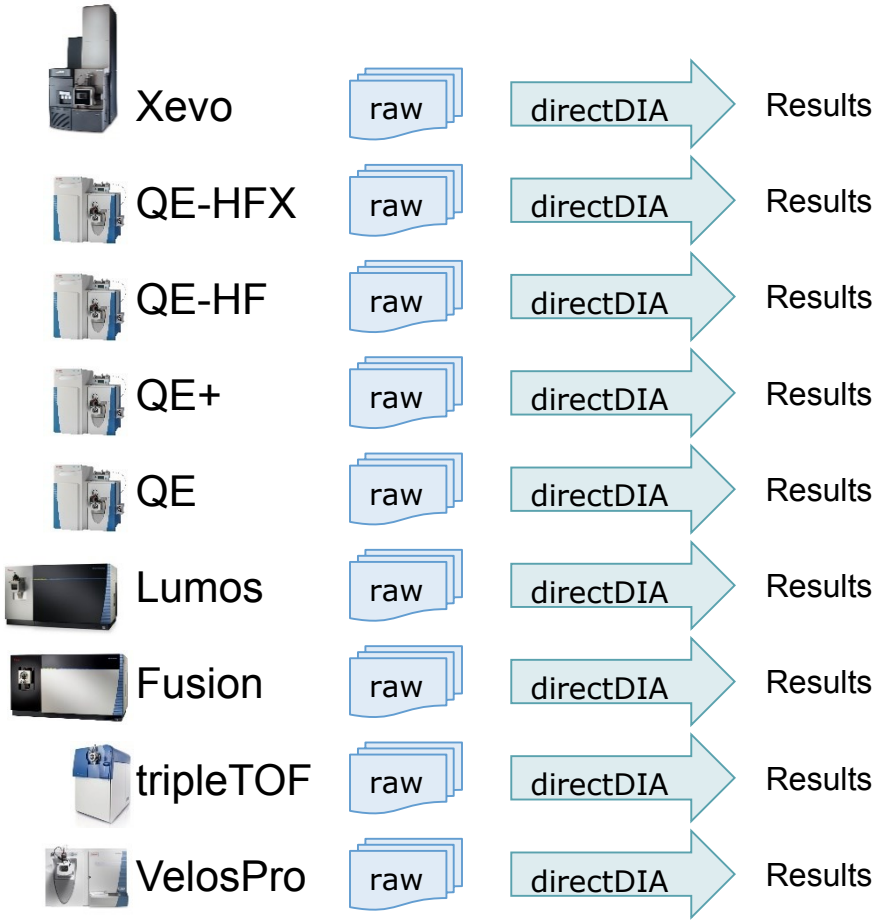
- Most labs achieved a satisfactory DPPP (7-10)
- After removal of outlier, average DPPP was 7.8
- Considering difficulty of predicting cycle time in trap based instruments, and diversity of platforms, this is surprisingly good.

Performance of Participants – cycle time



- Similar to DPPP, most labs achieved the target cycle time of 3.5 sec
- After removal of three outliers, average cycle time was 3.42 sec

Evaluating Results

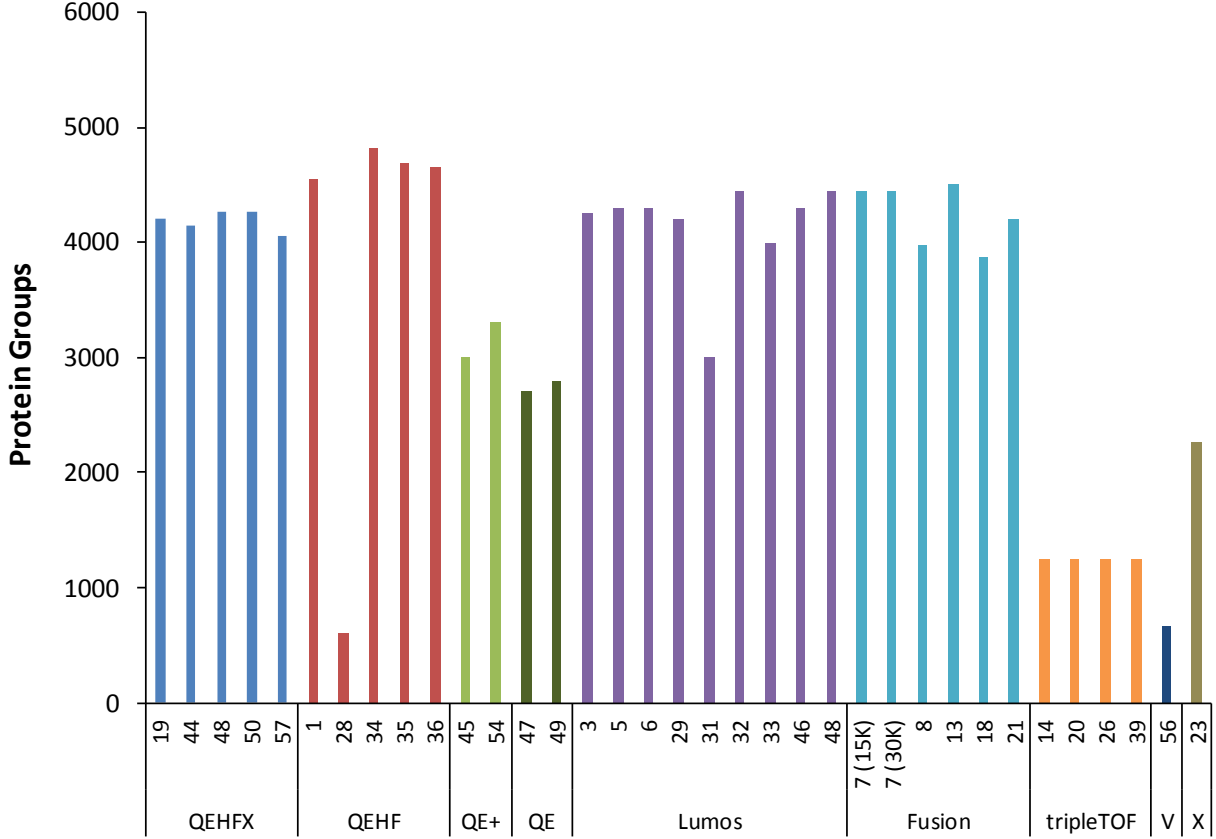


- ~3 TB of raw files (35 data sets)
- A library free approach was used by platform to remove bias towards a single participant library
- Default search settings were used

We used Spectronaut X with support from Biognosys, but this is not an endorsement. The PRG had experience with Spectronaut and available expert assistance.

The PRG is aware that TripleTOF data performs much better when using the Pan Human Library, versus using directDIA and the Human UniProtKB SwissProt + varsplic database.

Performance of Participants



- 3500 protein groups identified on average (600 to 4820)
- Mostly consistent within platforms across users
- Running a two hour gradient of a tissue lysate is not optimal for all instruments (nor their intended application).

Spiked proteins

Sample

Blank 25 fmol 100 fmol

HeLa digest spiked with four non-endogenous proteins and iRT

ABRF-1: beta-galactosidase; 1024 aa; 116.5 kDa

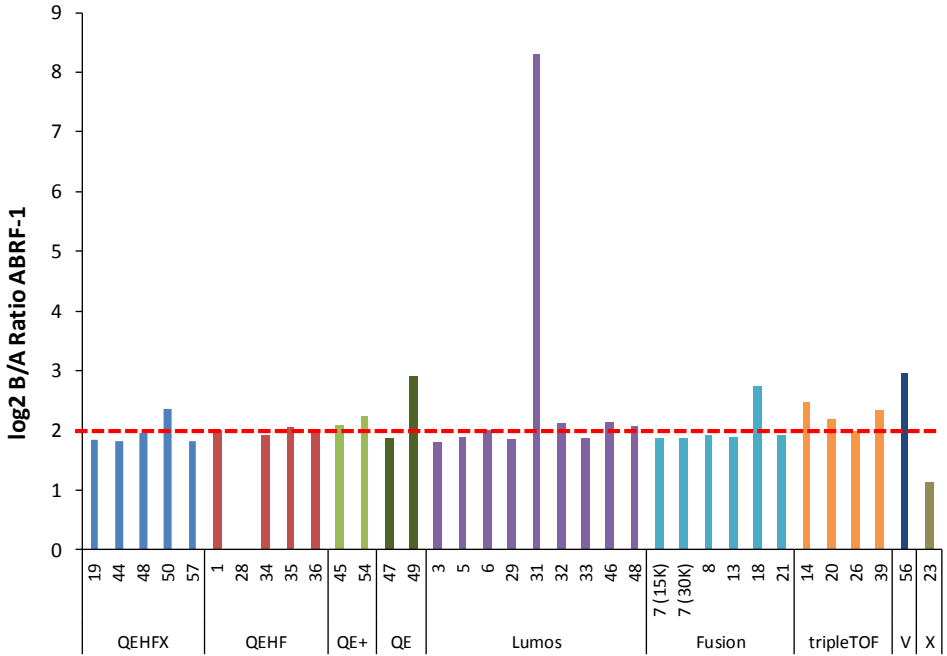
ABRF-2: lysozyme C; 147 aa; 16.2 kDa

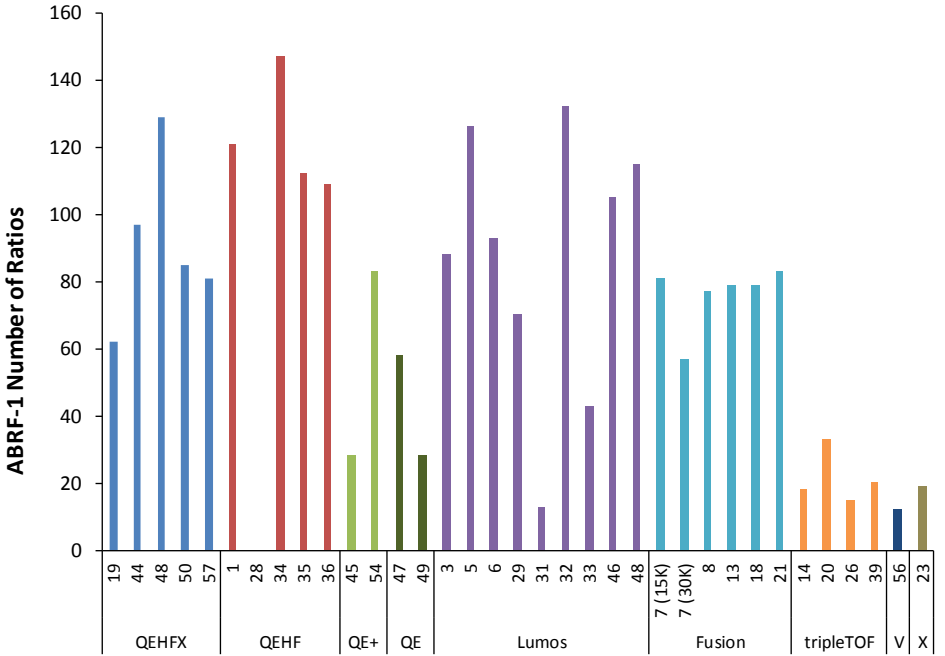
ABRF-3: glucoamylase; 640 aa; 68.3 kDa

ABRF-4: Protein G; 185 aa; 20.1kDa

Sample A: 25 fmol/µg HeLa digest
Sample B: 100 fmol/µg HeLa digest
Sample C: blank (just HeLa digest)

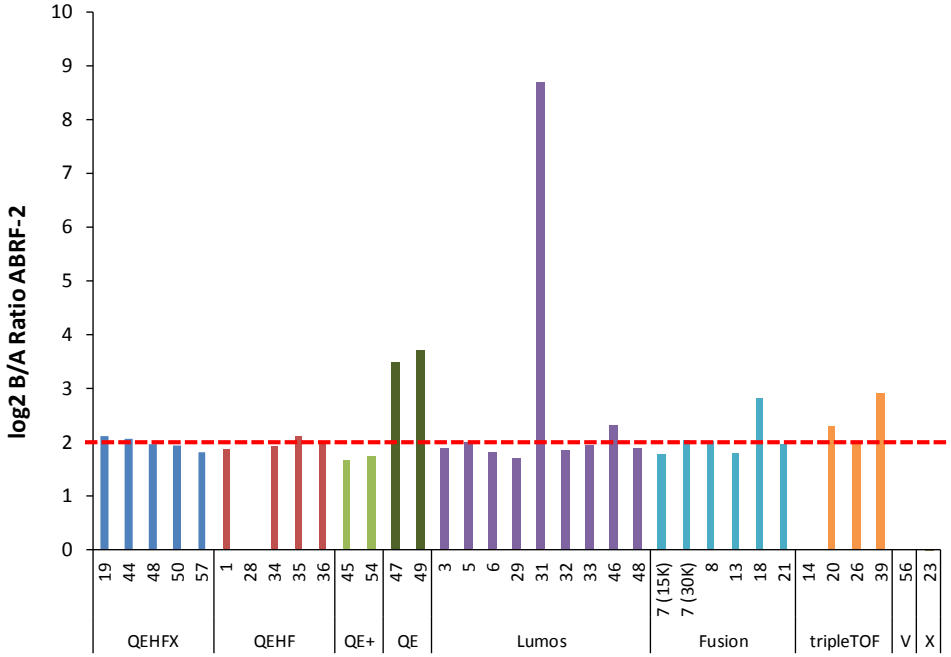
Performance of Participants – ABRF-1 beta-galactosidase

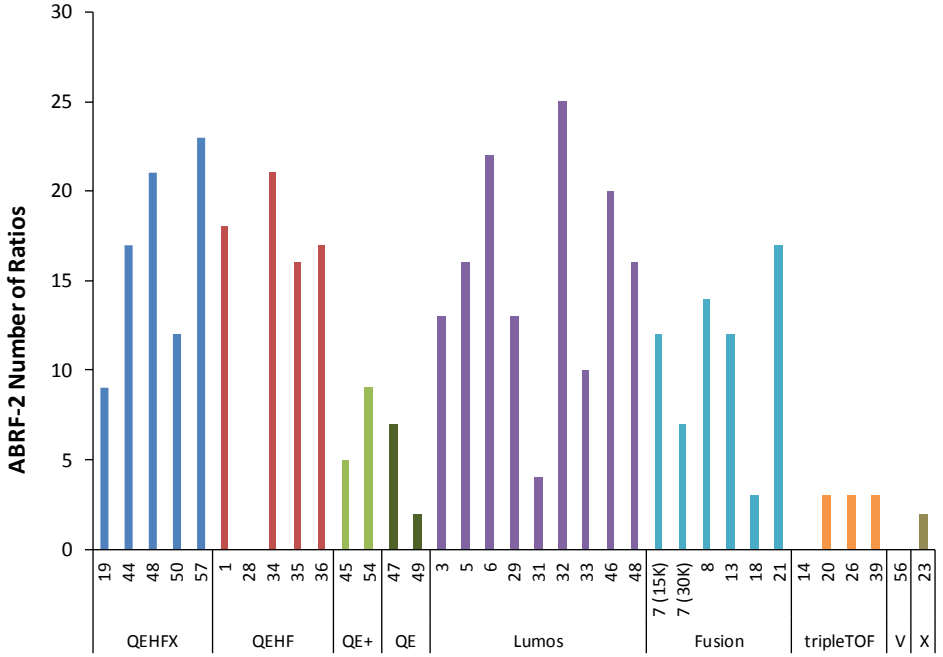




- Majority of labs correctly measured a 4-fold change (83% were very close)
- The number of measured ratios (peptide elution groups) varied from 147 to not detected

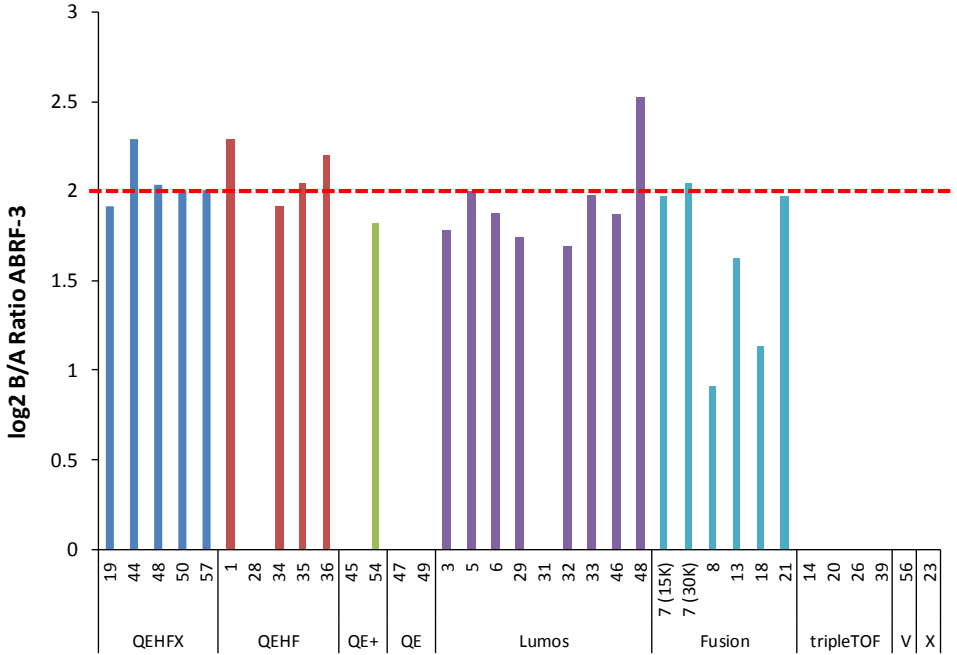
Performance of Participants – ABRF-2 lysozyme C

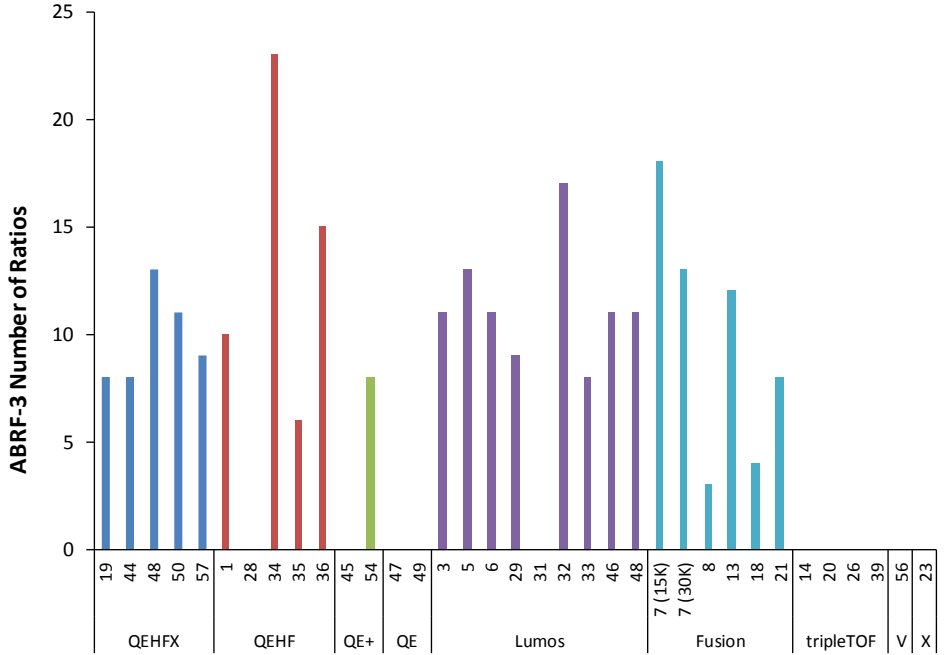




- Majority of labs correctly measured a 4-fold change (69% were very close)
- The number of measured ratios (peptide elution groups) varied from 25 to not detected

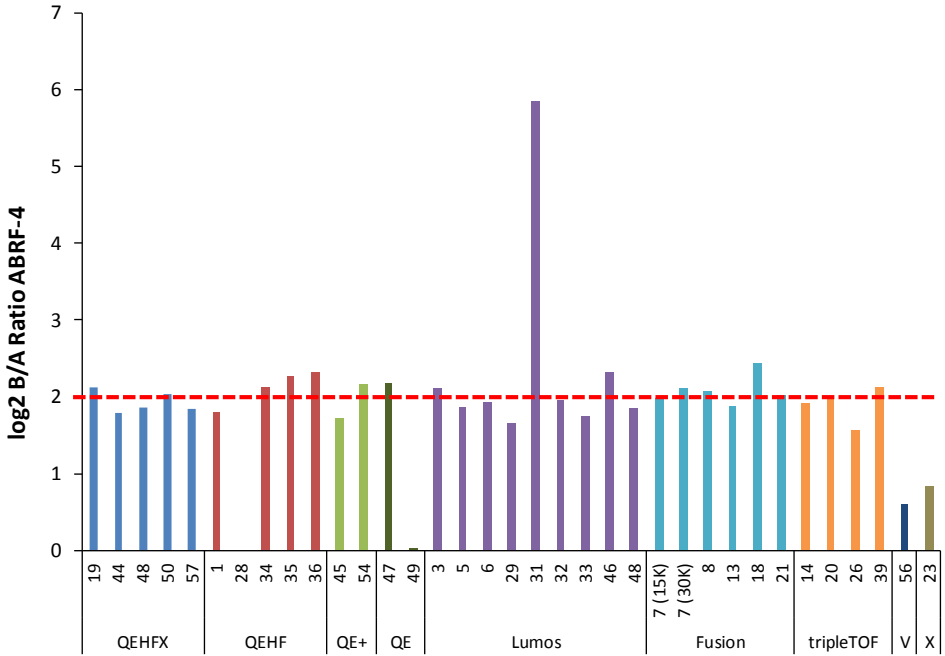
Performance of Participants – ABRF-3 glucoamylase

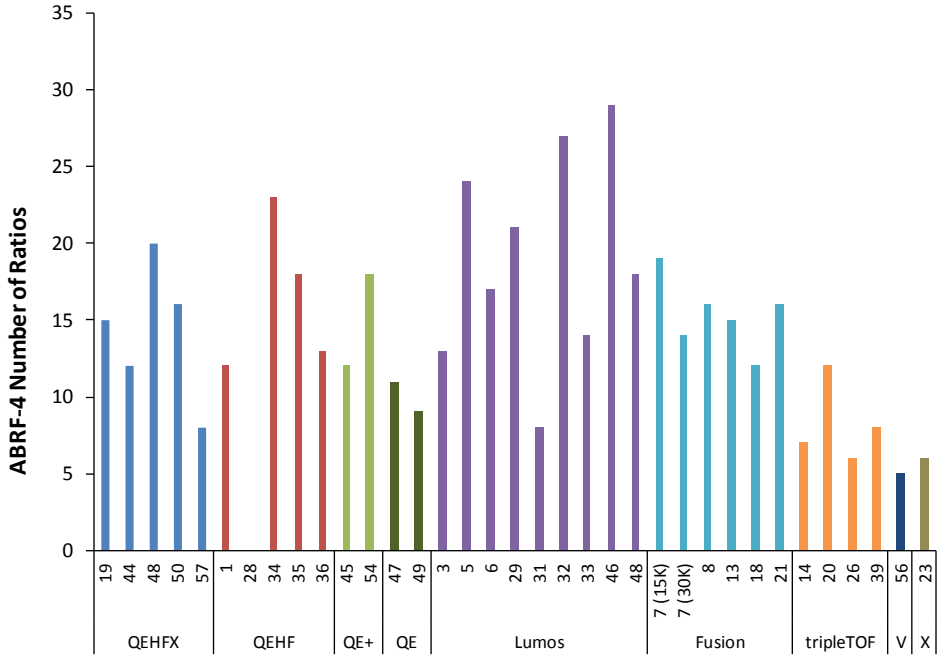




- Most labs correctly measured a 4-fold change (54% were very close)
- The number of measured ratios (peptide elution groups) varied from 23 to not detected
- This protein was the hardest to detect in the study

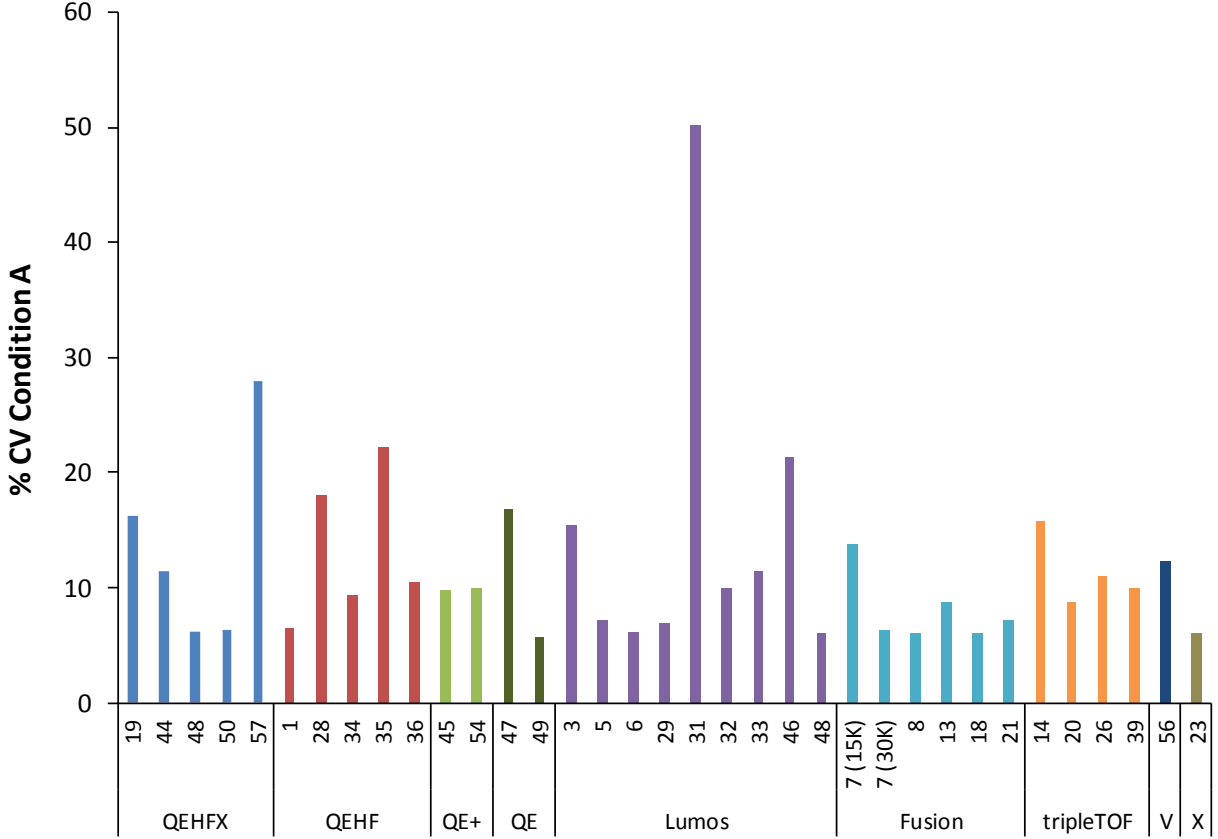
Performance of Participants – ABRF-4 protein G





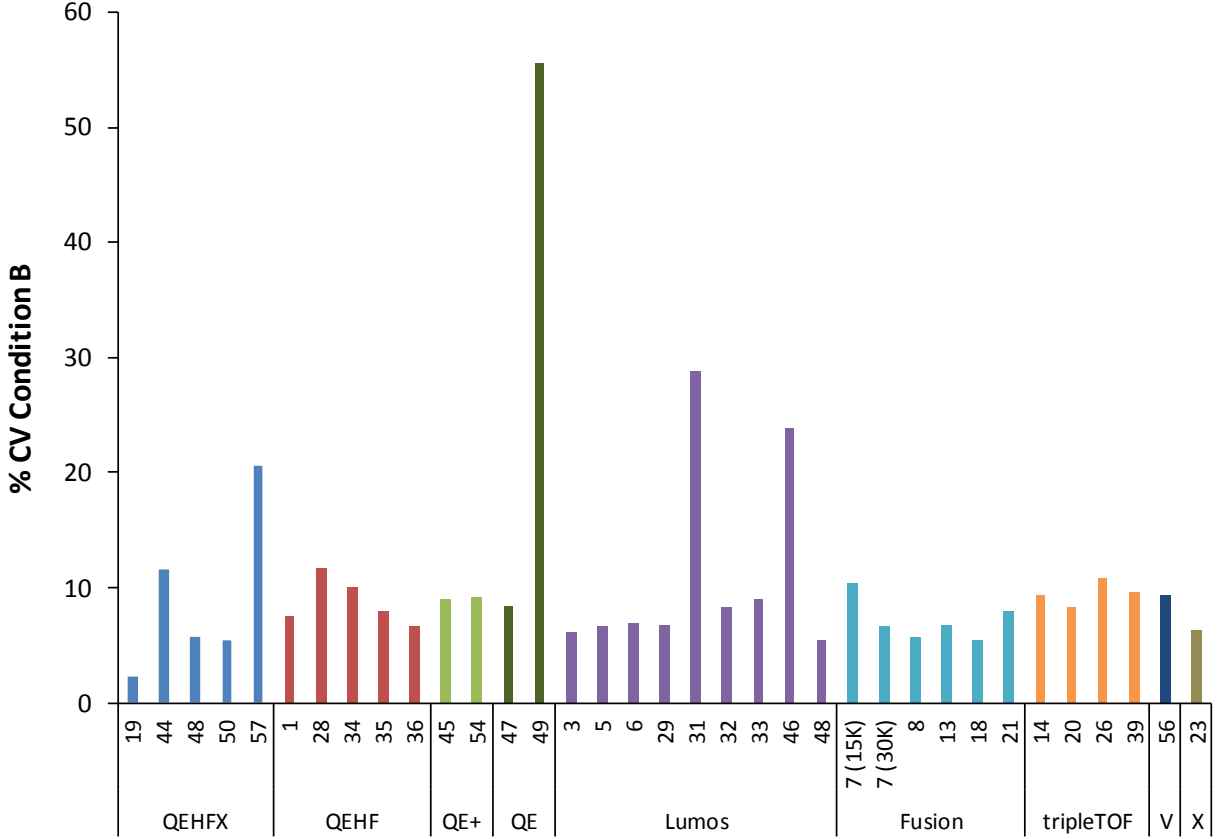
- Majority of labs correctly measured a 4-fold change (83% were very close)
- The number of measured ratios (peptide elution groups) varied from 29 to not detected
- Interesting to note much fewer elution groups versus ABRF-1, yet similar overall performance

Performance of Participants



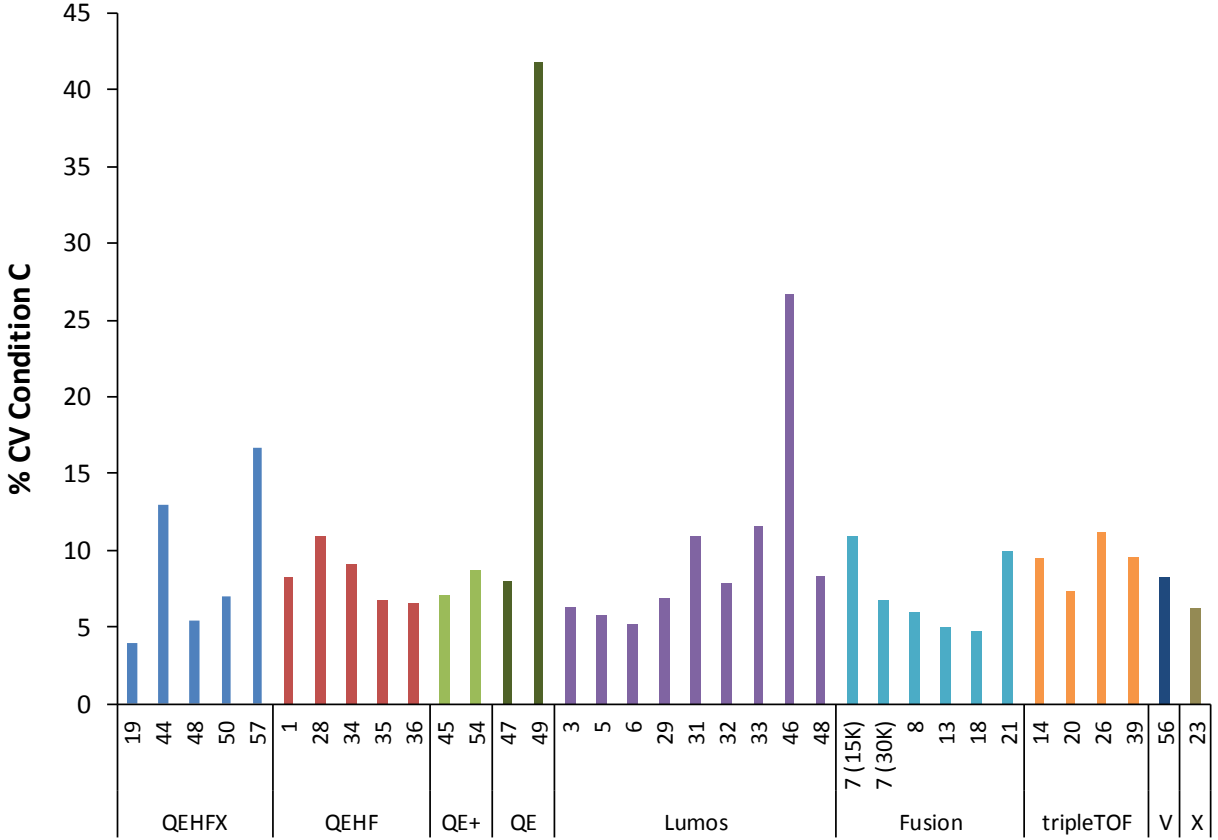
- Sample A was 25 fmol of non-endogenous spikes
- Across participants, average CV = 12.1 %

Performance of Participants



- Sample B was 100 fmol of non-endogenous spikes
- Across participants, average CV = 10.6 %

Performance of Participants



- Sample C was 0 fmol of non-endogenous spikes
- Across participants, average CV = 9.7 %

Future Plans and Data Availability

- Continue to look into the nuanced results of the study
- Some trends are consistent within platforms, and that users were consistent across performance metrics
  - if you had “good” DPPP and “good” high protein IDs, then you likely did well at everything else
- Summarize results into manuscripts
- Anonymize raw data and make available before June via MassIVE
- Alert software makers (commercial and non-commercial) to data availability to help with development and education

Questions?