Proteomics Research Group - 2018 DIA Study
Information for Participants

Since this document contains login information for the FTP, please do not share or distribute.

Introduction

Thank you for participating in the [ABRF PRG's 2018 DIA study](). The goals of this study are broadly two-fold: (1) allow you to run DIA methods on a standardized sample and (2) generate DIA data from a single sample across laboratories, expertise levels and instrument platforms. This document will provide instructions on how to set up your LC and MS, and upload the resulting data. Once the data has been uploaded you will be eligible to receive extended commercial licenses for some software (see relevant instructions). **This document will continue to be updated and the current version will be on the FTP where you will deposit your data (information below).**

You have received enough sample to run each sample in triplicate, which should be less than 24 hours of instrument time. There will be additional sample available to generate data for a spectral/chromatogram library, though this isn't required*. This document will provide you with methods to acquire data, but please email [prg.abrf@gmail.com]() if you have any questions.

*The PRG will be generating raw data that can be used to create a spectral library. Users that are interested in data analysis can download this from the FTP site.

Also, you have received a participant ID. **Do not share your participant ID with anyone unless you want to be known.** If you have lost your ID or have specific questions that require PRG to know your ID, please contact Paul Stemmer ([pmstemmer@wayne.edu]()) who is the only person with the information. You and your data will remain anonymous from all other PRG members, other participants, and the public.

The minimum we are asking for from you is to upload your raw data (de-identified with your participant ID). This will be uploaded to an FTP. This anonymous data will be available to all PRG members, participants and participating software vendors. This will allow for "real world" data to be used during a proposed DIA workshop at the March 2019 ABRF Meeting in addition to being analyzed by the PRG for the March 2019 ABRF Meeting. We expect to make this data publicly available as soon as possible.

If you wish you may also analyze the data, or better yet, attend the proposed ABRF workshop on March 23 in San Antonio and we can analyze it together.

If you have questions or concerns, please email prg.abrf@gmail.com.

Sample Preparation

You have received three samples labeled A, B and C. These are HeLa digests spiked with Biognosys iRT peptides and digests of four non-endogenous proteins at three levels: 25 fmol/μg or 100 fmol/μg of Hela digest or with no spiked-in proteins, in random order. The total digested protein amount in each vial is 10 μg or 20 μg for nanoflow or microflow users, respectively. We suggest bringing the samples up in 0.1% formic acid. For nanoflow applications 1 to 2 μg on column should suffice whereas microflow may require 4 to 8 μg on column. We are expecting you to know the appropriate loading capacity for your system, but we are available to consult if need be (prg.abrf@gmail.com). You will have enough material to run each sample in replicate, with the remainder available for whatever you wish (see comments on library generation earlier).

LC Gradient

The MS methods given below have included LC methods, but due to the diversity of LC systems as well as nano- or micro-flow applications, the PRG is assuming users can set up their gradients appropriately. Broadly, we assume you are doing a two-stage linear gradient lasting 110-130 min. Below are some options, or feel free to use a gradient you prefer for a complex mammalian tissue lysate like HeLa. ***If you use your own LC gradient there will be a csv in your upload folder where you can paste your gradient and mobile phase information.***

(all % are acetonitrile, ACN, so adjust based on your mobile phase compositions)

130 min gradient
equilibrate
5-25% ACN over 100 min
25-40% ACN over 20 min
40-90% over 10 min
Hold at 90% for 5 min
90-5% over 1 min
re-equilibrate

Acquisition Settings

The PRG has generated MS methods for this study. If you have never used DIA we feel these are excellent starting points, but also acknowledge you will be able to improve them as you grow more comfortable with your specific setup's performance.

When constructing a DIA experiment, the MS2 mass range, number of MS2 windows, MS2 window width and time spent acquiring data all contribute to determining your instrument's cycle time and therefore how many data points you will acquire over a peptide's elution peak. For this study we have specified static MS2 window widths covering 400-1200 *m/z*, with a 1 Th overlap. We encourage you to learn more about other window strategies, but think these are adequate initial settings.

We are assuming a 30 sec peak width at base and therefore these methods all have 3.5 sec cycles to achieve between 7 and 10 data points per peak. This cycle time is slightly longer than some recommendations, and if you have shorter peaks they will be undersampled. Although not required, if you would like advice on how to generate a 3, 2.5 or 2 sec cycle method please email us at prg.abrf@gmail.com.

Method Files:
Lumos/Fusion
> PRG2018_Lumos_130min_15kMS2_62_14mz.meth
> PRG2018_Lumos_130min_30kMS2_40_21mz.meth

QE-HFX
> PRG2018_QE-HFX_145min_30kMS2_40_21mz.meth

Waters
> PRG2018_SYNAPT_17_65V.luc (right-click and save-as)

SCIEX
> PRG2018_TTOF_117min_80_11mz.dam

Table 1. Settings for Thermo platforms. Check back soon for more!

Platform	Lumos 30k MS2	Lumos 15k MS2	QE-HFX
Gradient	130 min	130 min	145 min
Default z	4	4	3
(S-Lens)*/ion funnel RF	(60) 30	(60) 30	40
Resolution FS	120,000	120,000	120,000
AGC FS	1.E+06	1.E+06	3.E+06
mode	profile	profile	profile
Max Inj FS	20	20	20
Scan Range	399-1200	393-1200	399-1200
Iso Width	21	14	21
Number of Segments	40	62	40
window range	399-1200	393-1200	399-1200
NCE**	HCD30	HCD30	HCD30
Resolution MS2	30000	15000	30000
Scan Range	200-2000	200-2000	200-2000
AGC MS2	1.00E+06	1.00E+06	3.00E+06
Max Inj MS2	60	30	60
mode	profile	profile	profile
Parallelization	OFF	OFF	OFF
cycle time (sec)	3.5	3.5	3.5

no lock mass used

* 60 if Fusion rather than Lumos because then S-lens rather than IF

** adjust to prefered energy

(a note on cycles, for the Lumos when using these settings a 3 second cycle would be 34 windows in the 30k MS2 method and 54 windows in the 15k MS2)

Table 2. Settings for Sciex platforms.

Platform	TripleTOF6600
Gradient	117 min
TOF-MS Mass Resolution	≥35,000
Mode	Profile
Accumulation time	250ms
Scan range	400-1200
# Fixed isolation Windows	80
Mass range of windows	11 m/z
Fragmentation	CID
Collision energy (CE)	Rolling
Collision energy Spread (CES)	15V
Ion source	Duo Spray
TOF-MS/MS Mass Resolution	≥20,000
Mode	Profile
Accumulation time	40ms
Scan range	100-1500
Cycle time (sec)	3.5

Table 3. Settings for Waters platform.
Xevo G2-XS settings (SONAR)
Synapt G2-Si Settings (HDMSE)
Quad window 28Da
Quad scan 400-1200
Acquisition 50-2000
CE Profile for SONAR 20-35V, HDMSE see lookup table

If you decide to use a different method, there will be a csv file in your upload folder where you can provide information about your acquisition settings.

Acquisition Order
In order to minimize batch effects please run samples in the following order: (A1, B1, C1), (B2, C2, A2), (C3, B3, A3) (with a blank run between each set).

Data Deposition
The raw files will be deposited onto a private ftp site following these directions.

1. Download FileZilla Client: https://filezilla-project.org/download.php or CyberDuck: https://cyberduck.io/
2. Use the following information to connect:
> Host: massive.ucsd.edu
> Username: PRG_DIA
> Password: PRGDIAstudy
> Port 21
3. Find the folder with your assigned participant ID and move the raw files into the folder.
4. Please label your raw files according to the file naming instruction (below).
5. Email Paul Stemmer (pmstemmer@wayne.edu) and let him know your data has been uploaded.
6. If you have any questions about the protocol and data submission, please send an email to prg.abrf@gmail.com.

Name raw instrument data files according to this structure.
ID_instrument_[prg|lab]_S[A,B,C]_R[1-3]

ID - your unique study identification number
instrument - (see table below)
[prg|lab] - whether you used a supplied MS method or your own. If your own, please include a brief table of relevant settings in your upload folder.
S[A,B,C] - sample A, B, or C
R[1-3] - replicate 1, 2, or 3

Ex. Data for participant 007 (license to measure) from second rep of sample A, acquired on a lumos with PRG supplied method:

007_lumos_prg_SA_2.raw

Table 4. Instrument abbreviations for file name:

Instrument	Naming convention abbreviation (use lower case)
Orbitrap XL	orbixl
Orbitrap Velos	orbive
Orbitrap Velos Pro	orbivep
Orbitrap Elite	orbielite
Orbitrap Fusion	fusion
Orbitrap Lumos	lumos
QExactive	qe
QExactive Plus	qeplus
QExactive HF	qehf
QExactive HFX	qehfx
TripleTOF 5600	tt5600
TripleTOF 6600	tt6600
timsTOF	Tims
Synapt G1	syn1

Synapt G2	syn2
Synapt G2-Si	syn2si
Xevo G2-XS	xevo

Survey after Data acquisition

Once you have acquired the data and deposited in the repository, we request that you take a short survey. This will help us to understand and document the specifics on how you acquired the data.
https://www.surveymonkey.com/r/PRG_DIA_2019

Please contact us at prg.abrf@gmail.com if you have any questions.

Data Analysis

We require no data processing from participants. That being said, software vendors will supply extended versions of their products to participants once data has been deposited (see relevant section for instructions), and the PRG will be hosting a workshop at ABRF 2018 on March 23, 2019 in San Antonio, TX that will be focused on DIA data analysis, specifically related to this study. **If you are interested in attending this workshop, we recommend installing demo software on your laptop so that you can use it at the workshop.**

Most software allows for a library independent search (ex. both Spectronaut and Scaffold DIA do not require a library to analyze DIA data). You will still need a protein FASTA file. Here is a protein FASTA file of the four non-endogenous spike protein sequences and with these protein sequences concatenated onto the human SwissProt + isoforms database (release 2018_08):
PRG2018_spikes.fasta (right-click, save-as)
Human SwissProt + isoforms + PRG spikes (right-click, save-as)

Alternatively, the PRG has supplied raw data files (via the FTP) that can be used to generate a library using different software. On the FTP in the "Raw files for library generation" folder are two sub-folders. These contain raw files using the gradient and sample provided for this study. The "DIA Chromatogram Lib Files of test sample with spikes" was acquired using tight-windows (aka gas fractionation) to be used with Scaffold DIA. The "Triplicate DDA data of test sample with spikes" is DDA data that can be used for most other DIA software. [note: within each of these folders is data acquired during our pilot phase which may be interesting since this is a slightly different gradient and different HeLa stock, but *may* still be used.]

After generating the DIA outputs, if you have enough sample for a single DDA injection of each sample, you can choose to make your own instrument specific library. Alternatively, you can combine all your remaining samples into one and perform a DDA run(s) with that sample (in this case there will be ~40 fmol of spiked-in proteins per µg Hela). If you choose to acquire DDA

data to create your own library, it should be acquired with high resolution and in profile mode (*e.g.*, Orbitrap/Orbitrap MS1/MS2). Alternatively, if you wish to generate a chromatogram library, please consult Scaffold DIA's documentation. **Again, this is not necessary and you can download the PRG generated library files from the FTP site**

Finally, if you do elect to analyze your own data (which is great!), we would like to know some very general performance metrics. This includes the number of protein groups measured across all three samples and then questions related to percent coefficient of variation (CV), also known as relative standard deviation (RSD). This can be calculated within each replicate run when a protein was measured three times. The %CV (RSD) is the standard deviation divided by the average (x 100). We will ask how many proteins in sample A had % CV < 20, and then what the % CV was for each of the four spikes in the low (*i.e.*, 25 fmol/μg HeLa) and high (*i.e.*, 100 fmol/μg HeLa) spiked samples. Since you are blinded to these, you will need to know which is low and high (good luck!). If you have problems determining this for your data, please email prg.abrf@gmail.com for assistance.

Once the study is closed in January 2019, PRG members will be analyzing the data with available software, potentially including Spectronaut, Scaffold-DIA, Peaks DIA, Skyline, and others (*e.g.*, OpenSWATH, DIA-Umpire, DIA-NN, etc.). We will be using a resource based library to analyze the data, specifically focused on variability and detection limits of proteotypic peptides from the non-endogenous peptides.

Participant anonymity, software vendors and demo availability
In order to receive extended software demos from the software vendors for the study you must first acquire and deposit your raw data onto the FTP. Following the instructions above, you will have alerted Paul Stemmer (pmstemmer@wayne.edu) that your data is deposited. Once you have done this, you will be added to a list of participants eligible to request software. If you do not wish your participation to be shared with the software vendors, please send an email to prg.abrf@gmail.com stating that you wish to opt out, though you will not be eligible for any extended software demo.

Below are links to specific contacts at companies once you have acquired data (**please do not request via their website or before you have deposited your data**):

Biognosys - Spectronaut
Email Sira Echevarría-Zomeño <sira.echevarria@biognosys.com> with your name, institution sponsored email and your group and institution.
Ex.
Joe Bob
jbob@nrnih.gov
Bob Lab/NR-NIH

Proteome Software - Scaffold DIA
Email Nick Vincent-Maloney <support@proteomesoftware.com> with your name, institution sponsored email, and your group and institution.

Bioinformatics Solutions Inc. - PEAKS X
Email Katherine Tran <ktran@bioinfor.com> with your name, institution sponsored email and your group and institution.

Please note that the list of eligible participant names will be provided to all vendors with the disclaimer that it not be used as a marketing tool. (Your name will be known, independent of your study ID). Also, once the participant has requested software the subsequent relationship is between participant and vendor, and the PRG is not responsible for any follow-up marketing or software specific support. The deposited data will remain anonymous (only one member on the PRG will know what Study IDs belong with what participant).

If you are interested in attending the [ABRF DIA workshop](), we recommend installing demo software on your laptop so that you can use it at the workshop.

Link to an upcoming workshop at ABRF 2019: https://conf.abrf.org/sw1-data-independent-acquisition-practice-foundations-and-resources-implementing-dia

Registration Link: https://conf.abrf.org/registration-information

Please contact us at prg.abrf@gmail.com if you have any questions.