Association of
Biomolecular Resource Facilities

Research • Technology • Communication • Education

iPRG-2020 Proteome Informatics Research Group Study on Metaproteomics

Dear iPRG-2020 Study Participant,

Thank you for your interest in this year's Proteome Informatics Research Group (iPRG) study. This letter provides the instructions needed to access the data files, complete your analysis, and submit your results. This the first iPRG study in which hints to answering the questions will be made available in stages, enhancing the interactive nature of the study and enabling participants to update their submission after receiving the additional information. In the first phase participants will be blinded to the species composition, while in the second phase the species will be given. First results should be returned by **Friday, June 26, 2020** and final submissions by **Friday October 2, 2020** to qualify for inclusion in the study report that will be presented at the next ABRF Annual Meeting in Boston, March 7-10, 2021.

Study Overview

Metaproteomics – the characterization of proteins expressed by microbiomes [1] – presents a range of technical challenges, from sampling to data processing and interpretation. In this study we are investigating where we are as a field with respect to metaproteomics data analysis workflows. This year's study is a bit different than iPRG studies in the past in that participants will not need to identify or quantify specific peptides or proteins, but rather are asked to deduce the organisms or taxa in a metaproteomics sample ("What species are represented in the sample?") and what biological phenomena have taken place ("What interactions took place between the species in the mixture?").

The goal of the study is to learn where we are as a field with respect to analysing and reporting metaproteomics data. The participants, therefore, have a high degree of freedom and may use any peptide and protein identification software, metaproteomics analysis, sequence database (or *de novo*), or a combination of several methods. Participants will have immediate access to both MS1 and MS2 data and can report their findings in any format they like, including plain text, as long as they are well described and human-interpretable. As in

the previous iPRG study, we welcome submissions as R Markdown or Jupyter notebooks, but they are not essential.

There are many different pipelines and software tools available for metaproteomics data analysis, often building on or incorporating standard strategies for peptide and protein identification from bottom-up proteomics data. Many of these tools can be found on **https://bio.tools** and several were recently reviewed in a paper by Sajulga and co-workers [2], including UniPept 4.0, a tool for visualization of metaproteomics data (Figure 1).

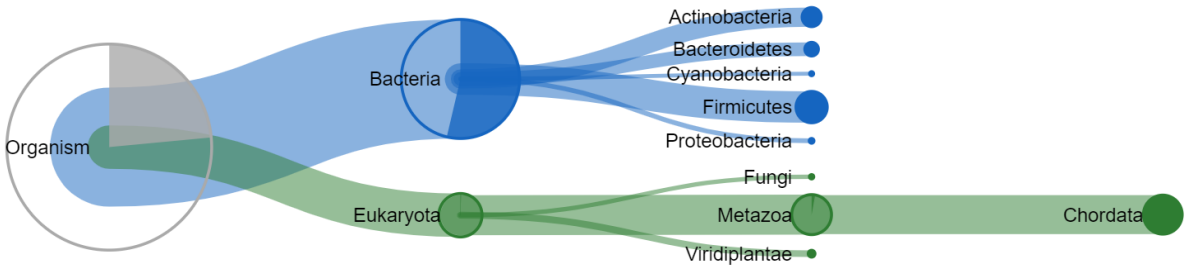

Figure 1. UniPept view of human gut microbiome proteomics data from Verberkmoes *et al*. [3]. Note that this is *not* related to the iPRG-2020 study data.

In Phase 1, participants are provided raw data only. The first task is to determine the taxonomical composition of the sample. In the first phase, participants may use any sequence database(s) they wish or identify peptides *de novo*. The second task is to estimate the relative abundance of the taxa in the sample. The final and most challenging task is to describe observed biological phenomena. In Phase 2, starting July 27, we will provide a sequence database that covers the species present in the sample. The participants will then be asked again to estimate the relative abundance of the species and describe any biological phenomena observed, such as interactions between species.

Deliverables

Phase 1 (answers due **Friday, June 26, 2020**):
Submit the following:

1.1 A detailed explanation of the steps used for the analysis, including why the specific sequence databases were selected, how they were assembled, how spectra were matched/assigned and how the taxa were identified.

1.2 A list (as a text file or table) of the taxa identified in the sample.

1.3 Relative abundance of the difference species (such as numbers of PSMs, distinct peptides or proteins).

1.4 Description of any biologically interesting phenomena you observed (such as biological pathways, functional groups or proteins).

Phase 2 (answers due **Friday October 2, 2020**):

In the second phase, **the participants will be provided with a protein sequence database of the species present in the sample**. Interested participants will then be asked to:

2.1 Provide a detailed description of how you performed the analysis, after being provided with the FASTA sequence databases covering the species present in the sample.

2.2 Provide list as in 1.2 above of the species or taxa identified in the sample, along with metrics of their relative abundance such as number of PSMs, distinct peptides or proteins.

2.3 Describe any biologically interesting phenomena you can observe (such as biological pathways, functional groups or proteins).

Raw LC-MS/MS data files (see details below) from a Thermo Fisher Scientific Orbitrap Fusion Lumos instrument are available on **https://cpm.lumc.nl/export/iPRG2020/**. In the second phase, FASTA protein sequence databases will also be provided at the same location. The results should be submitted via e-mail to the study anonymizer, Sue Weintraub (**weintraub@uthscsa.edu**). The anonymizer will not share the identities of the participants with anyone, including other members of the iPRG. Please do not disclose which results are yours unless you no longer wish to be anonymous.

Sample Preparation and LC-MS/MS Data Acquisition

Four biological replicates of a metaproteomic sample were analysed. The sample contains proteins from several biological species, but no spiked-in protein standards. All proteins were reduced and then alkylated with iodoacetamide prior to digestion with trypsin. Each digest was analysed by capillary LC-MS/MS on a Thermo Scientific Orbitrap Fusion Lumos mass spectrometer. On-line HPLC separation was accomplished with a Thermo Scientific/Dionex RSLC NANO HPLC system: column, PicoFrit™ (New Objective; 75 μm i.d.) packed to 15 cm with C18 adsorbent (Vydac; 218MS 5 μm, 300 Å); mobile phase A, 0.5% acetic acid (HAc)/0.005% trifluoroacetic acid (TFA); mobile phase B, 90% acetonitrile/0.5% HAc/0.005% TFA; gradient 3 to 42% B in 30 min; flow rate, 0.4 μl/min. Precursor ions were acquired in the Orbitrap in centroid mode (scan range, *m/z* 300 – 1500; resolution, 120,000); data-dependent higher-energy collision-induced dissociation (HCD) spectra of ions in the precursor scan were acquired at the same time in the ion trap ("top speed;" threshold to

trigger MS2, 50,000; quadrupole isolation, 0.7; charge states, 2+ – 5+; dynamic exclusion, 30 sec; normalized collision energy, 30%). The manufacturer recommended mass tolerances for data generated in this way are 10 ppm for MS1 and 0.6 Da for MS2.

Important note to vendors and commercial laboratories: ABRF imposes strict guidelines on the use of study results for marketing purposes. The guidelines are described in this document:

https://abrf.org/sites/default/files/temp/Resources/research_group_study_participation_guidelines_2010.pdf

We thank you for your support of ABRF and look forward to receiving your results for the study!

Sincerely,

The ABRF Proteome Informatics Research Group (iPRG)

Darryl Davis (Chair) - *Janssen Pharmaceuticals, Horsham, PA*
Tony Harvey - *Thermo Fisher Scientific, Austin, TX*
Michael Hoopmann - *Institute for Systems Biology, Seattle, WA*
Lukas Käll - *Royal Institute of Technology, Stockholm, Sweden*
Samuel Payne - *Brigham Young University, Provo, UT*
Yasset Perez-Riverol - *European Bioinformatics Institute, Hinxton, UK*
Susan T. Weintraub - *Univ. of Texas Health Science Center at San Antonio, TX*
Magnus Palmblad (EB Liaison) - *Leiden University Medical Center, The Netherlands*
Pratik Jagtap (*ad hoc*) - *Univ. of Minnesota, St. Paul, MN*
Benjamin Neely (*ad hoc*) - *National Institute of Standards and Technology, Charleston, SC*

References

1. Zhang X and Figeys D Perspective and Guidelines for Metaproteomics in Microbiome Studies. J. Prot. Res. 2019;18(6):2370-2380, doi:10.1021/acs.jproteome.9b00054

2. Sajulga R, Easterly C, Riffle M, et al. Survey of metaproteomics software tools for functional microbiome analysis. bioRxiv 2020.01.07.897561, doi:10.1101/2020.01.07.897561

3. Verberkmoes NC, Russell AL, Shah M, et al. Shotgun metaproteomics of the human distal gut microbiota. ISME J. 2009;3(2):179-189, doi:10.1038/ismej.2008.108