HIPC Data Standard for:
Mass Spectrometry-based Metabolomics, Proteomics and Lipidomics

The HIPC Data Standards Working group proposes the following data standard for mass spectrometry-based proteomics, metabolomics and lipidomics experiments:

Proteomics:

- Raw Data (format depends on vendor): The samples should be renamed so that the sample names correspond to ImmPort experiment sample identifiers.
- Protocol: free-text document including vendor name, machine model, description of setting used to pick peaks (including search parameters, software name, version number)
- Protein sequence database (FASTA format): sequence names in this file should be UniProt ID
- Protein intensities (csv format): this file should provide the protein intensity measurements. One column should be labeled "Protein", and another should be labeled "Intensity". Each row will provide data for a single protein with the protein column providing the protein name (UniProt ID or PRO short label, matching with the protein sequence database), and the intensity column providing the corresponding relative intensity value (arbitrary units). This file may contain other columns if desired. There is no requirement for retention time, m/z or z values. It is allowable to submit a single file that includes results from multiple experiment samples. In this case, one column should be labeled "Protein", while multiple additional columns will indicate the intensities associated with each experiment sample. These column names should include both "Intensity" and the ImmPort experiment sample identifier. A template detailing the requested format is available at ImmPort.

Metabolomics (and Lipidomics):

- Raw Data (format depends on vendor): The samples should be renamed so that the sample names correspond to ImmPort experiment sample identifiers.
- Protocol: free-text document including vendor name, machine model, description of setting used to pick peaks (including search parameters, software name, version number)
- Relative intensities or concentrations (csv format): this file should provide the metabolite (or lipid) intensity measurements. One set of columns will identify the metabolite (or lipid), and should include columns for retention time, m/z and z. For metabolites, HMDB ID and PubChem ID (where available) should be included, and RefMet IDs may also be included. Lipids should be identified by IUPAC or common names (https://www.lipidmaps.org/data/classification/lipid_cns.html) and PubChem ID (where available). A second column labeled "Intensity" will provide the corresponding relative

intensity value (arbitrary units). It is allowable to submit a single file that includes results from multiple experiment samples. In this case, multiple columns will indicate the intensities associated with each experiment sample. These column names should include both "Intensity"and the ImmPort experiment sample identifier. A template detailing the requested format is available at ImmPort.