Features of GlycoPAT2.0

Target release date: April 15, 2021

Final testing and documentation is in progress

Overall goal: Glycoproteomics software to handle standard proteomics LC-MS/MS workflow, particularly for complex mixtures (like human blood plasma)

What to expect:

- *OS:* Open-source, platform-independent with many user-friendly Graphical User Interfaces (GUIs). Tested extensively using Win PCs (Intel, AMD) and Intel-based clusters

- *Computer* requirements: Can run on a single CPU with 16GB RAM (minimum), but better to have >6 core-64 GB RAM configuration. As the program is capable of parallel processing, more cores will ~proportionally reduce computational time

  - *Types of analysis*: peptide, N-glycopeptides, O-glycopeptides (data-dependent acquisition studies only)

- *Input file formats*:
  - **a.** *MSDATA expt data*: .mzML (.RAW files are converted using Proteowizard package);
  - **b.** *Protein sequence file*: .fasta, .txt;
  - **c.** *Glycan PTM modifications*: provided in SGP1.0 format (IUPAC-condensed to/from SGP1.0 converter provided)
  - **d.** *Non-glycan PTM definitions*: .txt

- *Fragmentation modes*: predefined CID, HCD, ETD, EThcD, ETCiD fragmentation rules are available. These can be modified for additional fragmentation modes

- *RawData pre-*processing:
  - **a.** *Glycopeptide quantitation*: Performs label free quantitation based on XIC evaluation of MS1 chromatogram
  - **b**. *Monoisotopic mass adjustments*: Based on custom Averagine model

- *Search space*:
  - **a.** *Protein*: Handles either 1 single protein or whole human proteomics (~20,000 proteins).
  - **b.** *N-glycan search library (default)*: 1180 members, including 270 compositions plus isomers.
  - **c.** *O-glycan search library (default)*: 27 carbohydrate structures, including glycan isomers.
  For *glycan-structure focused glycoproteomics*, all isomers are scored and consensus structure(s) are identified

- *MS/MS workflows handled*:
  - **a.** *Product-dependent*: Specifically optimized for HCD product-dependent MS workflows, i.e. HCD(pd)-> CID IT plus EThCD OT. cES (combined Ensemble Score, $0<cES<1$) will cover metrics from all fragmentation modes

b. *Independent spectrum analysis*: Can score individual MS/MS spectrum independently. The output is individual ensemble scores (0<ES<1)

- *Result evaluation and visualization*: Integrated browser with DrawGlycan-SNFG for viewing results from both product-dependent and single MS/MS mode workflows. Tools available to accept/reject/refine GlycoPAT2.0 derived automated results. Output saved in .pdf, .xlxs, .mat format.

- *gpAnnotate:* A single MS/MS spectrum annotation tool for integrating SNFG sketches

- *Testing*: Tested primarily using Thermo Orbitrap instruments. But should work with other MS instruments also.