Instructions Manual

2024

GlycoGenius

The Ultimate Glycomics Data Analysis Tool

Version 1.2.7

GUI Version 1.0.6

Table of Contents

1. General Information

GlycoGenius is a Python package that was developed at the *Universidade Federal do Rio de Janeiro* (UFRJ), Brazil and at the University of Groningen (UG), the Netherlands to facilitate the data analysis of glycomics. The program enables the identification, characterization as well as quantification of released N- and O-glycans, GAGs and glycoforms of specific glycopeptides that were acquired through full scans (without fragmentation data) or Data Dependent Acquisition (DDA) mode fragmentation using liquid chromatography (LC) or capillary electrophoresis (CE) hyphenated with mass spectrometry (-MS) instrument setups. GlycoGenius is applicable for data obtained in positive or negative ionization mode.

This Python package is capable of performing a plethora of functions, including:

- Accessing the contents of MzML and MzXML mass spectrometry data (through Pyteomics);
- Automatically calculating the noise level in a per-retention-time and per-m/z-range basis;
- Building a library of glycan compositions, allowing for different adducts and derivatizations strategies;
- Tracing extracted ion chromatograms/electropherograms (EICs/EIEs) of every glycan in the library;
- Filtering any non-specific traces on the EIC/EIE, (which avoids the picking of non-monoisotopic peaks as well as peaks with incorrect charge counts, possible misassignments);
- Peak-picking the features in the chromatograms/electropherograms, identifying different peaks for separate analysis, allowing to analyze isomers;
- Quantifying the area-under-curve (AUC) of each identified peak;
- Providing a score for the isotopic peaks based on theoretical peaks intensity calculated on the abundance of isotopes of the atoms;
- Providing a score for the identified peaks shapes based on Gaussian fit;
- Analyzing MS2 spectra, identifying possible fragments with a series of customizable options and scoring the spectra as a whole based on the percentage of the total-ion intensity of the spectra;
- Outputting Excel files, with a combined sheet that contains all information, as well as separated sheets that distinguishes different samples, glycans, adducts and peaks;
- Allowing integrations with other software data analysis platforms:
  - Creating a transition list compatible with Skyline;

    - Creating a Metaboanalyst compatible file to do statistical analysis of the results.
- Visualizing all the EICs/EIEs traced, along with all the fitting data;
- Automatically arranging data based on quality thresholds, signal-to-noise ratio or maximum PPM error;
- Drawing SNFG-compliant glycan cartoons;
- Calibrating MzML and MzXML spectra files;
- Other minor features.

GlycoGenius has not been tested with datasets that contain the following conditions:

- Data Independent Acquisition (DIA) fragmentation mode;
- Analysis of other glycoconjugates, such as glycosphingolipids;
- Analysis of glycoproteomics data.

Please note that GlycoGenius has not been tested extensively on macOS. There may be other conditions not matching the ones described above that have not been extensively tested.

The program is completely written in Python and benefits from computers with a high number of CPU cores (up to 64 threads can be used).

This user-guide is A5 size and has been designed to be printed as a handbook.

2. Installation and Running GlycoGenius

GlycoGenius is a package available as a Command-Line Interface (CLI) or Graphical User Interface (GUI) at the Python Package Index (PyPI) and can be installed using pip:

Windows:

There are two ways to install GlycoGenius on Windows:

1. Download the latest version of the executable for the GUI, built with PyInstaller from GlycoGenius_GUI GitHub from the link:
   https://github.com/LoponteHF/GlycoGenius_GUI

2. Installing directly from PyPI:
   a. Install the latest version of Python from python.org;
   b. Go to your preferred terminal (i.e. Powershell or Command-Prompt) and type 'pip install glycogenius' or 'pip install glycogenius_GUI';
   c. Type 'glycogenius' or 'glycogenius_GUI' in the terminal to start using.

Linux and other operational systems:

1. Most Linux distributions come with Python installed, otherwise install python 3 from the relevant repository (usually by typing '[sudo] apt install python' for Ubuntu/Debian, '[sudo] dnf install python' for Fedora, for example, or by using the package manager available for you OS distribution);
2. Type 'pip install glycogenius' or 'pip install glycogenius_GUI' at the terminal and wait for the installation to finish;
3. Run GlycoGenius by typing 'glycogenius' or 'glycogenius_GUI' on the terminal.

3. The Graphical User Interface (GUI)

3.1 The Main Window (**Fig. 1**)

Please note that all settings in GlycoGenius are provided with a description, for this the user needs to hover over the checkbox/button/dropdown/entry field.

1. **Workflow:** helps to analyze the data, there are five essential steps;
   a. Select files (for more information see *Section 3.2*)
   b. Set parameters (for more information see *Section 3.3*)
   c. Generate a new library (c.1), import an existing library (c.2) or check information on how the imported library was built (c.3)
   d. Start the analysis
   e. Save the results into Excel files (for more information see *Section 3.4* and *5.2*)
2. **Dropdown menu:** provides access to all loaded samples;

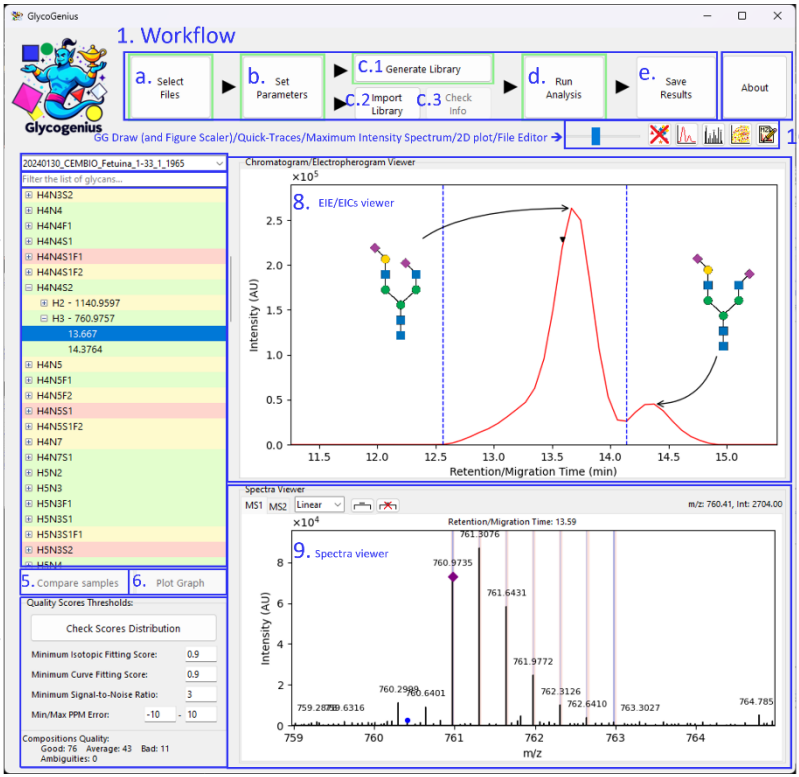


Fig. 1: *Main window of GlycoGenius*

3. **Filter/Search field:** allows to filter/search for specific glycan compositions (e.g. N2) or features (e.g. good, average or bad);
4. **Glycans/Chromatograms/Electropherogram list:** displays all identified glycans with their corresponding chromatogram/electropherogram. Additionally, the Base Peak Chromatogram/Electropherogram (BPC/BPE) is provided per sample (for more information see *Section 3.5*);
5. **Compare Samples:** allows comparing chromatograms/electropherograms of a specific glycan between all your samples (for more information see *Section 3.6*);
6. **Plot graph:** allows comparing the abundance of your glycans within a sample or compare a single glycan between multiple samples, regardless of their quality;
7. **Quality Control Parameters:** allows setting four parameters to control the quality threshold of your glycans. These include *Isotopic Fitting Score, Curve Fitting Score, Particle-per-Million (PPM) error* and Signal-to-Noise (*S/N) ratio*. The settings can be changed and a direct visualization will appear:
   a. If the glycan **fit all** thresholds, it is considered to be a ***"Good" glycan***, indicated **green** in the list;

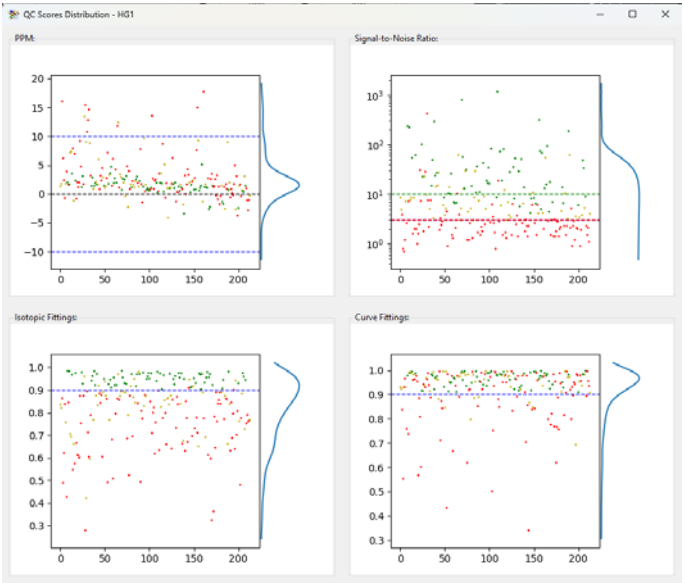


Fig. 2: *Quality Criteria Scores distribution window. Analytes that meet all criteria are indicated in green, if one of the criteria is not met it is illustrated by yellow and if two or more are not met, they are displayed as red. Dotted blue lines indicate the set values. Green dotted line indicates S/N >9 and red dotted line S/N >3*

b. If the glycan **fails one** of the thresholds, it is considered to be an ***"Average" glycan***, indicated with **yellow** on the list;

c. If the glycan **fails two or more** of the thresholds, it is considered to be a ***"Bad" glycans***, indicated with **red** on the list;

The distribution of the four quality control parameters can also be plotted to determine which values fit best to your dataset (**Fig. 2**). On this window, you can also model the electrophoretic migration of your analytes, according to Barroso A et al., 2015 (EXPERIMENTAL);

8. **Chromatogram/Electropherogram:** displays the EIC/EIE of the glycan that has been selected on the left side (**Fig. 1**, *4. Glycans/Chromatogram/Electropherogram list*). The EIC/EIE is fully interactive, that allows to zoom in using different methods and to select any retention/migration time (RT/MT) that is desired. Multiple EICs/EIEs can be plotted from the same sample by holding down the *Shift button* while selected the desired glycans. Using the arrows on the keyboard the user can navigate from left to right in the EIC/EIE.
9. **Spectra Viewer:** displays the MS1 (full Scan) spectrum for a specific RT/MT that is selected in the *8. Chromatogram/Electropherogram window*. If the data contains MS2 data, precursors will be indicated with a purple diamond on top of the peak (See **Fig. 1**). To access the MS2 spectra, the user should click on the purple diamond and it will be displayed on the "MS2" pane. The y-axis can be scaled linearly (default), by the square-root or logarithmically. The user can measure distances between spectra peaks by using the ruler function, available at the buttons on top of the spectra viewer, besides the scaling drop-down menu.
10. **Additional Function Buttons:**
    a. **GG Draw** allows GlycoGenius to draw the SNFG-compliant glycan cartoon of the compositions

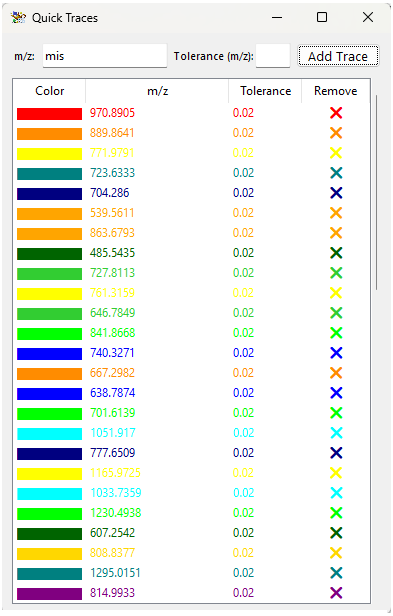
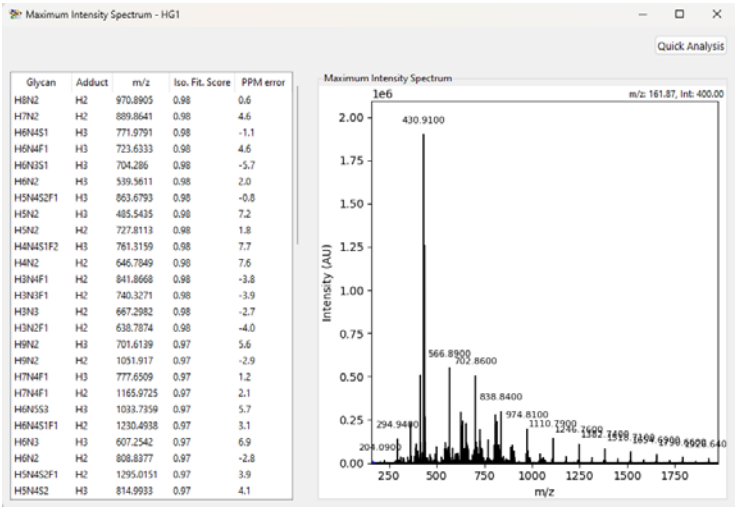


Fig. 3: *Quick Traces window. Here you can insert a desired m/z and a tolerance to quickly trace on your sample.*



Glycan	Adduct	m/z	Iso. Fit. Score	PPM error
H8N2	H2	970.8905	0.98	0.6
H7N2	H2	889.8641	0.98	4.6
H6N4S1	H3	771.9791	0.98	-1.1
H6N4F1	H3	723.6333	0.98	4.6
H6N3S1	H3	704.286	0.98	-5.7
H6N2	H3	539.5611	0.98	2.0
H5N4S2F1	H3	863.6793	0.98	-0.8
H5N2	H3	485.5435	0.98	7.2
H5N2	H2	727.8113	0.98	1.8
H4N4S1F2	H3	761.3159	0.98	7.7
H4N2	H2	646.7849	0.98	7.6
H3N4F1	H2	841.8668	0.98	-3.8
H3N3F1	H2	740.3271	0.98	-3.9
H3N3	H2	667.2982	0.98	-2.7
H3N2F1	H2	638.7874	0.98	-4.0
H9N2	H3	701.6139	0.97	5.6
H9N2	H2	1051.917	0.97	-2.9
H7N4F1	H3	777.6509	0.97	1.2
H7N4F1	H2	1165.9725	0.97	2.1
H6N5S3	H3	1033.7359	0.97	5.7
H6N4S1F1	H2	1230.4938	0.97	3.1
H6N3	H3	607.2542	0.97	6.9
H6N2	H2	808.8377	0.97	-2.8
H5N4S2F1	H2	1295.0151	0.97	3.9
H5N4S2	H3	814.9933	0.97	4.1

Fig. 4: *Maximum Intensity Spectrum and Quick Analysis window. This window will have an aggregated spectrum from your sample, which comprises the highest intensity every m/z achieves. You can quickly check your sample for glycans by loading a library file and clicking on Quick Analysis, on the top-right side of the window. Every glycan on the list to the left can be selected to highlight Isotopic peaks and can be copied to clipboard by pressing CTRL+C.*

identified in the analyzed sample directly at the *8. Chromatogram/Electropherogram Viewer* (See **Fig. 1**). When toggled on, it will check it will generate the glycan cartoons the first time you select a glycan on the identified glycans list (up to 2 minutes per glycan). Subsequent times you select the glycans, it will be instantaneous. Once the glycans are drawn, you can drag them around, scale them using the slider besides the GG Draw button and you can double-click them to access a glycan cartoon browser where the user can choose from a variety of different structures for a given composition, with filtering options for glycans classes, branching numbers and other features. **The first composition drawn is chosen randomly from glycans within the class analyzed**;

b. **Quick-Traces** menu, which allows you to quickly trace a desired (or multiple, semicolon separated) *m/z* value in your sample (**Fig. 3**). Additionally, you can type in "mis" in the *m/z* entry field to import the identification list acquired from the Maximum Intensity Spectrum quick analysis (See below);

c. **Maximum Intensity Spectrum (MIS)** creates an aggregated single spectrum from your data (**Fig. 4**), which allows you to quickly check for

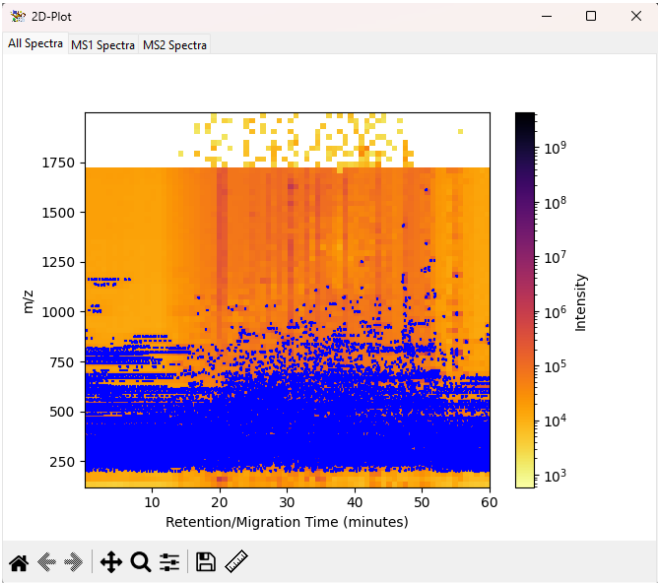


Fig. 5: *2D heatmap where the RT/MT is displayed on the x-axis and the mass/charge ratio (m/z) on the y-axis. MS2 data is indicated with blue circles and blue lines connected to it, highlighting the precursors and points to the MS2 spectrum associated with it.*

the presence of glycans (either manually or using a library to quickly detect them using GlycoGenius capabilities). The list can be arranged by clicking on the column titles and can be quickly imported into the Quick-Traces menu by typing in "mis" in the *m/z* entry field;

d. **2D plot** allows to explore your data using a 2D heatmap (**Fig. 5**). The RT/MT will be displayed on the x-axis and the mass/charge ratio (*m/z*) on the y-axis. In the "All Spectra" pane, blue circles will highlight precursor *m/z* values, with blue lines connecting to its respective MS2 spectrum. The user can also select to display only MS1 or only MS2 spectra by selecting the respective panes above. It's also possible to zoom in/out, pane around, reset view, measure distances and save high resolution pictures of the 2D plot using the controls below;

e. **File Editor** (**Fig. 6**) allows you to calibrate your mzML and MzXML spectra files. You build a calibrants list out of individual *m/z* values, standard calibrants or glycans from your library and search for them in one or all the sample. Once the search is done, the mass errors, peak retention time and other information about each calibrant will be displayed on the tree view panel and a plot will be draft showing the mass error, in *m/z*,

over the *m/z* values in your sample. You can remove any calibrants you wish to better calibrate your samples. Clicking on a calibrant will move the *9. Spectra Viewer* to the corresponding calibrant peak for better visualization. Once the calibrants list is tweaked properly, the sample can be calibrated by pressing the "Calibrate File" button, which will then create a new mzML spectra file and a PDF report file in the same folder as the sample was.

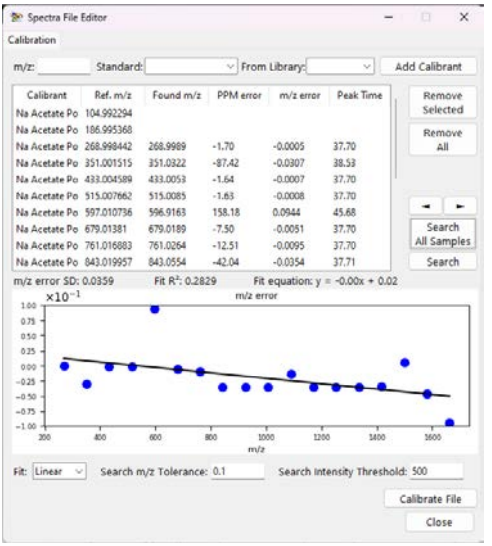


Fig. 6: *File Editor calibration pane.*

11. **About:** provides general information about the version of GlycoGenius, the author and disclaimers.

3.2 The Select Files Window (Fig. 7)

1. **Add files:** allows to add new files to the list by clicking the "*Add File*" button and selecting one or more files. Accepted formats are mzML and mzXML files;
2. Selecting files for analysis displays a list of all the sample files (".mzML" and/or ".mzXML" files) that have been added. The number of files can be seen on the right side, below the "*Remove Selected*" button;
3. **Remove selected:** allows to remove sample files from the list that have been highlighted in *2. Select files for analysis* section.

Optional functions

4. **Load .gg file:** allows the user to select a GlycoGenius analysis file (".gg") from a former analysis that has been performed. This allows for previous completed analysis results to be reviewed. After selecting the file, the loaded file path is displayed on the right side of the button;

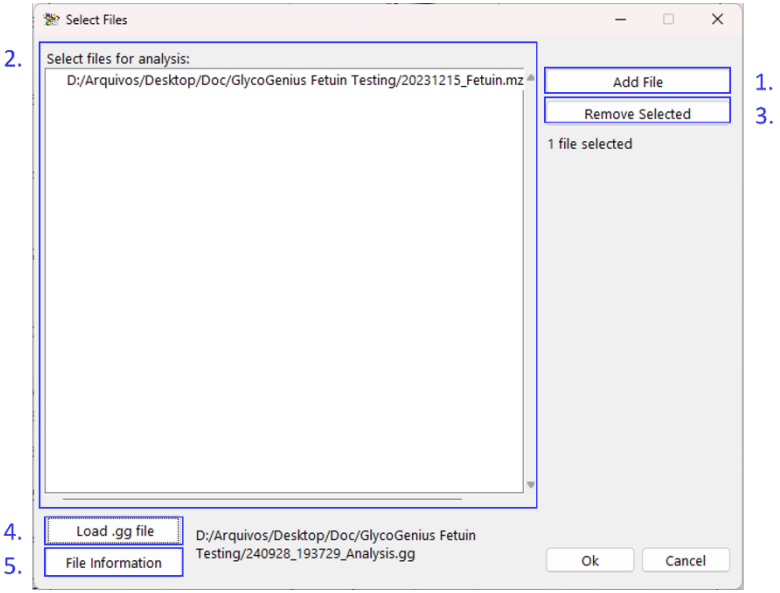


Fig. 7: *Select files window. In this window you are able to select the raw data files to analyze and/or a .gg file to reanalyze.*

5. **File information:** provides insights about the .gg file that is selected via *4. Reanalyze .gg file* (**Fig. 7**).

3.3 Set Parameters Window (Fig. 9)

The **Set Parameters Window** is divided into two sections, the left section is focused on the targeted library (search space) while the right section is focused on which settings should be used for the analysis as well as which resources GlycoGenius should use.

The **Library** section is further divided in different sections; *please note that if a library has already been generated, this step can be skipped and the library can be imported in the Main Window Workflow using Import Library (c.1);*

1. **Library building:** this section contains several options that

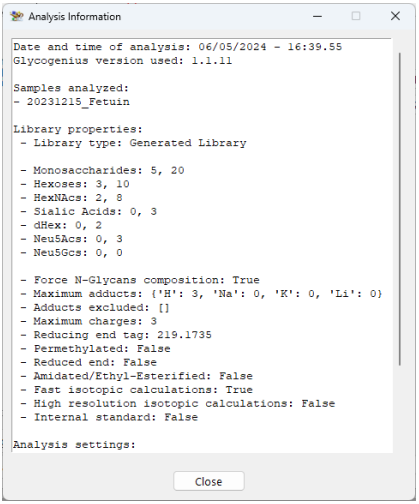


Fig. 8: *Analysis file information window.*

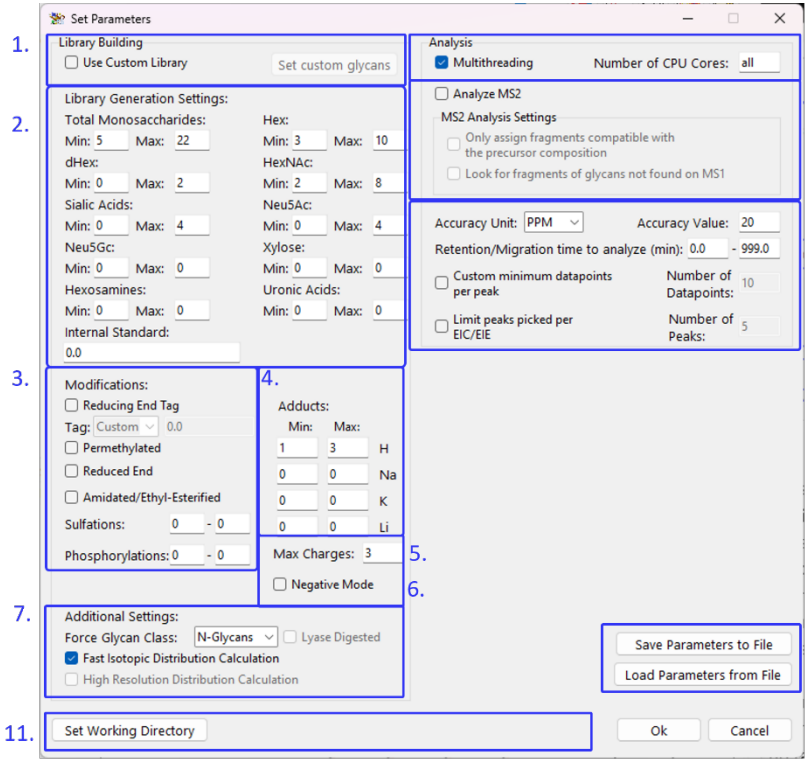


Fig. 9: *Set Parameters window. Here you can set parameters for building a library of glycans and for the analysis itself.*

allows the user to build a search space (or glycan library). The first option is "*Custom Library*". If this box is selected the "*Library Generation Settings*" will be disabled (except for "*Internal Standard*") and the "*Set custom glycans*" will be enabled. By clicking this option, another window will pop-up (**Fig. 10**), allowing the user to input specific glycan compositions for the data analysis of the sample. This function also allows to import the compositions using a text file (.txt), where glycan compositions should be written down in their one letter code (described on the window) either being comma-separated or line-separated within the file;

2. **Library Generation Settings:** allows users to generate a library based on a set of ranges defined per monosaccharide. The library is based upon combinatorial analysis. In the case an internal standard is used, the user can input either the molecular formula of the compound (of which the isotopic envelope will be calculated from), the mass of the compound (of which the

atomic composition will be automatically derived of and the isotopic envelope automatically calculated) or a glycan composition (in this case, the glycan will also suffer all the reducing-end, permethylation and sialic acid modifications that the user can set on field "3. *Modifications*");

3. **Modifications** can be taken into account:
   a. "*Reducing End Tag*" should be selected if the reducing end has been modified with a tag or in the case of glycopeptides (e.g. IgG1 where only one peptide backbone is explored). The user can choose from one of the available tags on the dropdown menu or set it to "Custom" and input the added mass, chemical formula or the peptide sequence of the backbone in the adjacent field. (e.g.: 219.1735 for procainamide or 133.0644 or $C_7H_7N_3$ for Girard Reagent P tag). For peptide backbones, "pep-" followed by the amino acid peptide sequence should be used (e.g.: pep-NK is a dipeptide with an asparagine and a lysine residue).
   b. "Permethylated" check-box can be used if the glycans have been permethylated.
   c. "*Reduced End*" should be checked if the released glycans are reduced without adding a tag.
   d. "*Amidated/Ethyl-Esterified*" should be checked if the sialic acids have been derivatized using Amidation/Ethyl-Esterification procedure (Lageveen-Kammeijer, G.S.M & Haan, N. de,. *et. al.*, 2019).
   e. "*Sulfations*" allows the user to select a minimum and maximum amount of sulfate groups added to the glycans in the library.
   f. "*Phosphorylations*" can be used to add a range of phosphorylations to the glycans generated;
4. **Adducts** allows the user to select the minimum and maximum amount of the different adducts

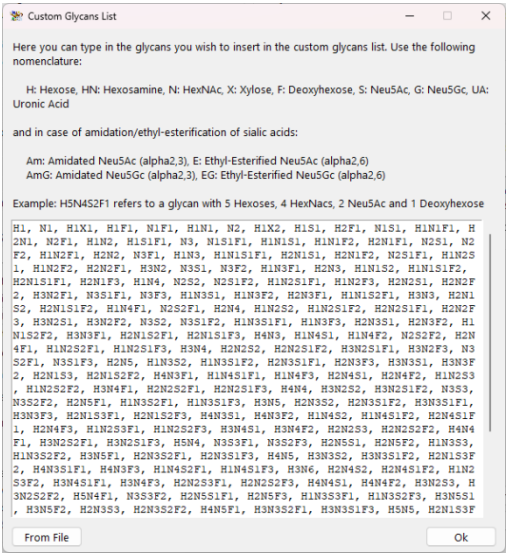


***Fig. 10**: Custom Glycans List window*

available (currently H, Na, K and Li are available as adducts). A combinatorial analysis is used to build all the possible combinations of adducts, within the charge range set on "*Max. Charges*";

5. **Max. Charges** should be set to define the maximum number of charges that are allowed within the search space of your library;
6. **Negative mode** should be selected if the data has been acquired in negative mode;
7. **Additional Settings**:
    a. "*Force Glycan Class*" options enforce rules to avoid putative glycan composition that are not common to have within the biosynthetic pathway of the specified class. The following rules are enforced:
        I. *N*-Glycans:
            i. The number of hexoses (Hex; H) and *N*-acetylhexosamines (HexNAc; N) cannot be smaller than two.
            ii. The number of sialic acids cannot be greater than two times the number of HexNAcs minus two, and cannot be greater than one time the number of Hex minus two; this limits the number of sialic acids to a maximum of two per antenna;
            iii. The number of fucoses (Fuc; F) cannot be equal or exceed the number of HexNAcs;
            iv. The number of xyloses (Xyl; X) can't be bigger than one;
            v. There can't be any hexosamine (HexN; HN) or hexuronic acids (UA).
        II. *O*-Glycans:
            i. The number of sialic acids cannot exceed the number of HexNAcs summed to the number of Hex;
            ii. The number of Fuc can't exceed the number of HexNAcs summed to the number of Hex;
            iii. The number of Hex can't exceed the number of HexNAcs plus one;
            iv. The number of HexNAcs can't exceed the number of Hex plus three;
            v. There can't be any Xyl, HexN or UA.
        III. GAGs:
            i. The sum of HexN and HexNAcs can't exceed the sum of UA and Hex plus one;
            ii. The sum of Hex and UA can't exceed the sum of HexN and HexNAcs;
            iii. Both Hex and UA can't be bigger than zero;

iv. There can't be any Fuc or sialic acids.

Additionally, if "GAGs" is selected, a "Lyase Digested" checkbox becomes available, which will make sure the library generated for GAGs is compatible with GAGs digested by the enzyme Lyase;

b. "*Fast Isotopic Distribution Calculation"* allows the calculation of the theoretical isotopic distribution of the glycans to be based solely on the carbon atoms, and then corrected using an equation. If left unticked, the calculation will be based on all the atoms, and may take a while, but it will be more precise. This will not only generate isotopic peaks evenly spaced based on the hydrogen mass divided by the charge count, but also adjacent peaks (within 0.002 Daltons of each other, due to the accurate mass of the oxygen isotopes). By default, these adjacent peaks are combined into a single peak, as most mass spectrometers don't have the resolution to resolve these adjacent peaks;

c. "*High Resolution Distribution Calculation"* becomes available when "Fast Isotopic Distribution Calculation" is disabled. This option allows the adjacent peaks disabled by default to become available and might be beneficial for ultra-high resolution mass spectrometry, such as Fourier-transform mass spectrometers (*note that the library generation will take much longer*);

The **Analysis Parameters** is also divided in different sections:

8. ***Multithreading****:* whether GlycoGenius should use multiple CPU cores on your computer. If you tick the "Multithreading" checkbox, you can set the number of threads to be used by GlycoGenius. The maximum amount it will use will always be the total amount of threads available minus two. Using higher number of threads can significantly improve the analysis time, if the resources are available. If unsure, leave it on and with "Number of CPU Cores" set to "all";
9. **Analyze MS2** allows to explore the MS2 data, if available. GlycoGenius will build a fragments library (usually several thousands of fragments long) and assign them to a MS2 spectrum, no matter what the putative monosaccharides composition of the precursor is. By enabling the MS2 option, two more options become available:
    a. *"Only assign fragments compatible with the precursor composition"* will not try to assign fragments containing monosaccharides outside the precursor putative composition (e.g.: if the putative composition of the precursor is H5N4S2, no fragments of N1F1 will be assigned in the spectra, even if the mass matches);

b. *"Look for fragments of glycans not found on MS1"* allows *to* look for MS2 spectra whose precursor mass matches glycans in the library that were not identified in the MS1 pass of GlycoGenius; if left unticked GlycoGenius will only look for MS2 spectra of glycans identified on the MS1.

10. **Other analysis settings** allow to set other restrictions to the analysis:
    a. *Accuracy unit/value* allows the user to set a PPM or *m/z* as the tolerance for *m/z* values of glycans searched by GlycoGenius (e.g.: 10 PPM is roughly 0.01 *m/z* at *m/z* 1000);
    b. *RT/MT* can be restricted to the analysis, by setting the beginning and end of the retention/migration time interval;
    c. *Minimum amount of datapoints per chromatogram peak identified* allows the setting of a minimum number of datapoints per peak found (e.g.: if this value is set to 10, chromatogram/electropherogram peaks with less than 10 datapoints will not be picked);
    d. *Maximum number of peaks picked per chromatogram* allows the setting of a maximum number of peaks picked per traced chromatogram/electropherogram (e.g.: an EIC/EIE might have 10 peaks, including noise peaks. If this option is used and the maximum is set to five peaks, it will only save data from the most intense peak and the four peaks that are surrounding it);
11. **Working Directory** allows the user to select the folder where all the files that are created during the analysis will be saved (e.g.: library files, .gg analysis files, Excel results sheets, etc.). The selected directory will be displayed to the right side of the button. **This option is mandatory**.
12. **Save/Load Parameters to/from file** allows the user to save the set parameters to a ".ini" file that can be used later on by loading it, both on the "Load Parameters from File" on this window or on the command-line interface (CLI).

3.4 Save Results Window (Fig. 11)

1. **Unassorted settings**, which allows the user to:
    a. *Include whole composition information*, which combines all the peaks identified for a specific EIC of a glycan into one abundance (area-under-curve, AUC) for quantification of compositions, in addition to the peak-separated information that in general is saved;

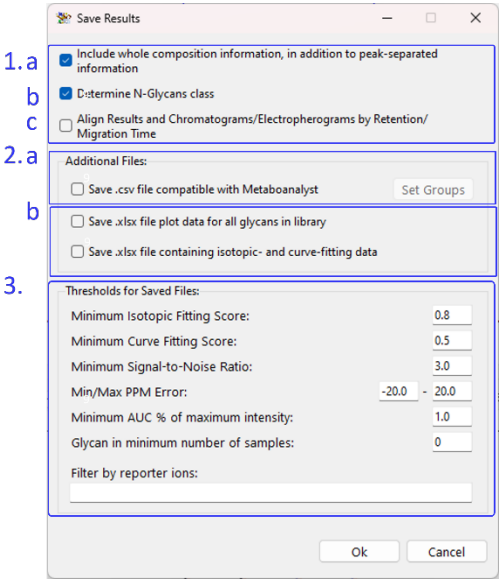


Fig. 11: Save Results window

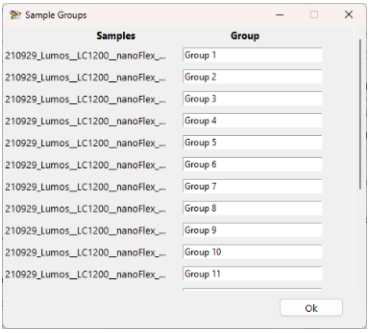


Fig. 12: Sample Groups window

b. *Determine the N-glycan class* (Hybrid, High-Mannose, Complex or Paucimannose) of each identified composition, which allows GlycoGenius to automatically quantify the different fractions;

c. *Align Results and Chromatograms by Retention Time* to provide results and chromatogram/electropherogram files aligned by the RT.

2. Additional files

a. *MetaboAnalyst file* option allows GlycoGenius to output a .csv file that puts the samples in columns and glycans in rows, with abundance information of every glycan for each sample. *Please note that the analyzed samples should be bundled into groups of at least 3 samples to be properly analyzed by MetaboAnalyst.* These groups can be set on the *"Set Groups"* button (**Fig. 12**). If you output this file, it can be used on MetaboAnalyst website to automatically provide many different statistical analyses of your dataset.

b. *Extra excel files* allows the user to choose extra output files, mostly for diagnostic purposes of the analysis (i.e.: you might think that a certain glycan not showing on the chromatograms/electropherograms list should actually be traced, so you can tick this option and GlycoGenius will output an excel file with isotopic fittings calculations and EICs/EIEs traced even for glycans not shown on the list). If you check the "*Save .xlsx*

file plot data for all glycans in library" box, you'll get an excel file with the chromatograms traced for all of the glycans in the library, no matter if they were identified or not. These chromatograms also include:

i. Raw chromatograms: Only the monoisotopic peaks considered, with no checks whatsoever;
ii. Smoothed processed chromatograms: The chromatograms/electropherograms you usually see in the GUI.

If you check the "Save .xlsx file containing isotopic- and curve-fitting data", you'll get an excel file containing all the isotopic peaks considered for scoring and what the score is, for every glycan of every sample in your dataset. You'll also get an excel file containing the curve-fitting data, where you can check the actual data and the gaussian traced to compare it to. **Outputting these extra files is very resources heavy and might take a considerable amount of time, so only do it if you really need to and know what to look for.**

3. **Thresholds for saved files** allows the user to select the score thresholds for the data saved to the Excel files. In addition to the thresholds that are displayed in the main window of GlycoGenius, the user can also choose to exclude peaks from chromatograms/electropherograms if they are below a certain percentage of intensity of the highest intense peak in a given EIC/EIE and to exclude detected glycans that weren't found in a minimum amount of samples. In case the data contains MS2, glycans can also be filtered by reporter ions. For this the user can type a specific *m/z* or monosaccharides formula (e.g.: N1T1 means the reducing end (with tag, if there is one) plus an HexNAc) and only glycans that have one of the reporter ions will be saved to the results file.

3.5 Using the Chromatograms/Glycans List and Chromatogram/Spectra Viewer

The identified glycans are displayed in a tree view pane. Each glycan contains information on identified adducts (for this the user can expand on the glycan by clicking the ⊞ symbol), followed at which RT/MT the identified peaks have been found (**Fig. 1, *Section 3.1***).

Color coding: All the glycans, adducts and peaks are color-coded:

- ***Green:*** Indicates that at least one of the peaks of the glycan meets all the quality control thresholds you have set;
- ***Yellow:*** Indicates that the best peaks within that glycan/adduct have failed one quality control threshold;
- ***Red:*** Indicates that the best peaks within that glycan/adduct has failed two or more of the quality control thresholds.

If the user selects a RT/MT of specific glycan/adduct, the chromatogram for that adduct, where the peak was found, will be displayed in the chromatogram/electropherogram viewer. The vertical dotted lines illustrate how the peak is delimited. By clicking within the peak boundaries, the spectrum for that RT/MT will be drawn on the Spectra Viewer (if you have the mzML/mzXML files loaded as well). The highlighted bars represent the spectra peaks chosen by GlycoGenius for the quantification, with the tolerance window in pink and the picked peak in blue (**Fig. 1, *section 3.1***).

3.5.1 Interaction with the Viewer

The user can interact with the Viewer using several commands:

1. *Pan the view:* by clicking and holding the left-mouse button and dragging the user can pan the view;
2. *Pan on specific axis:* the user can hold the left-mouse button on an axis and drag to pan on that specific axis;
3. *Zooming in*: the user can use their mouse wheel on the plot to zoom in both axis;
4. *Zooming in on specific axis*: the user can use their mouse wheel over an axis to zoom in only on that specific axis;
5. *Zooming into selected region*: the user can hold SHIFT and drag their mouse while holding the left-mouse button to zoom into a specific region;
6. *Right-Click Zoom/Pan on specific axis*: by right-

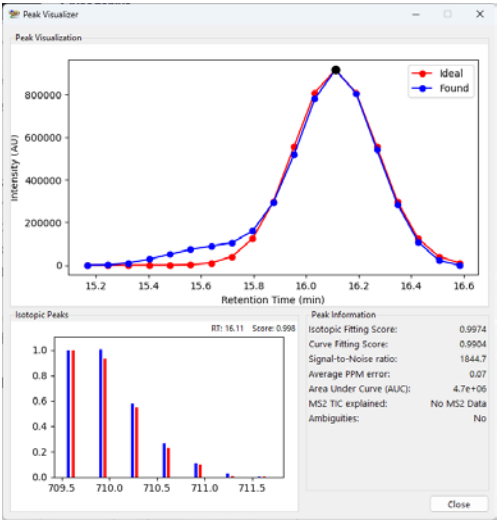


Fig. 13: *Peak Visualizer window*

clicking a specific axis and dragging, the user will be able to stretch the upper boundary of that axis;

7. *Reset zoom*: the user can double-click on any of the axis to reset the zoom on that axis.

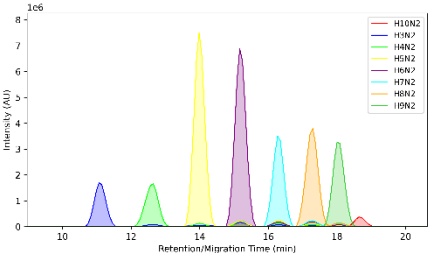


Fig. 14: *Comparing glycans chromatograms/ electropherograms by overlaying them*

3.5.2 Peak evaluation

To further explore the data, the user can double-click on one of the peaks in the chromatogram/electropherogram/ glycan list. This action will pop-up a Peak Visualizer window (**Fig. 13**), allowing the user to check the Curve-Fitting of that peak (e.g. how does the chromatogram peak shape hold against an ideal Gaussian) and the isotopic peaks used for scoring, as well as the theoretical isotopic peaks distribution of that glycan.

The user can also review the scores and the Area Under Curve (AUC) calculated for that peak. In the case MS2 annotation has been performed, the MS2 assigned percentage will be displayed (MS2 TIC explained). In the case of ambiguities (e.g. the glycan shares its mass with another glycan of different composition), the ambiguity will be mentioned and a black diamond shape will appear besides the glycan on the glycans list.

3.5.3 Overlay multiple glycans/adducts

The user can hold CTRL or SHIFT and select multiple glycans/adducts to draw their chromatograms/electropherograms together on the Chromatogram/Electropherogram Viewer (**Fig. 14**). If the user presses ESC, the selection from the glycans list will be disabled, but the last chromatograms selected will remain drawn until a new glycan/chromatogram is selected.

3.6 Comparing Samples

To plot the chromatogram/ electropherogram of a glycan for all the samples, the user can select a glycan (or adduct)

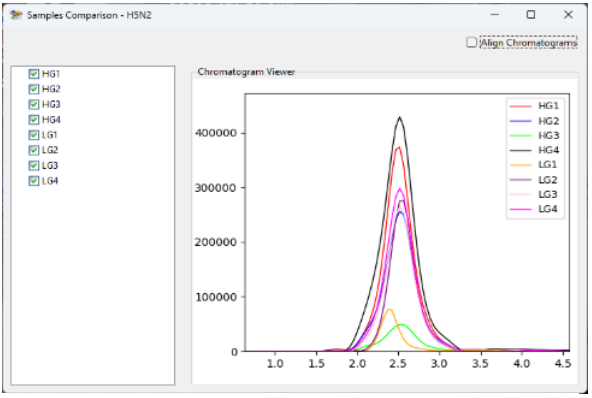


Fig. 6: *Samples Comparison window*

and click the button "*Compare Samples*" under the glycans list. Once the compare samples window loads, a new interface will pop-up (**Fig. 15**). On the left side of the window, the user can choose the samples of which to plot the chromatograms /electropherograms. On the top right the user can find a checkbox for aligning the samples. If checked, GlycoGenius will align the chromatograms of the sample based on glycans that have been identified as "good" on the list. *Please note that it might take up to several minutes to align the samples the first time the user clicks on it with the current thresholds.* Subsequent alignments, with the same quality criteria thresholds, will be instantaneous. The user can zoom in/out and pan around the chromatograms/electropherograms just as with the one in the main window.

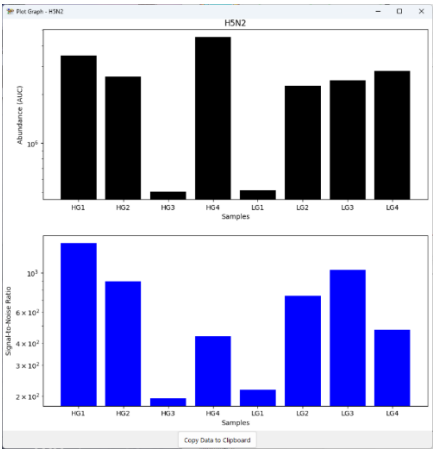


Fig. 7: *Plot Graph Window*

The user can also compare glycans between samples by selecting a glycan/adduct and clicking the *"Plot Graph"* button. This will display a window containing two bar graphs, with the top one being the abundance of the chosen glycan within each sample and the bottom graph shows the Signal-to-Noise ratio of the glycan, regardless of their quality (**Fig. 16**). If the user selects multiple glycans on the glycans list this button also allows to compare these glycans within the selected sample. The user can click the button *"Copy Data to Clipboard"* on the bottom of the window to copy all the data used in the graph to the clipboard, which can then be pasted wherever desired (e.g.: in Excel to make new plots). In addition, if the user hovers over the bars, it provides the name of the sample/glycan and the value of the bar.

3.7 Saving Figures in Publication Level Resolution

Every graph plotted in GlycoGenius can be saved by right-clicking the plot area and clicking "Save Image" (**Fig. 17**).

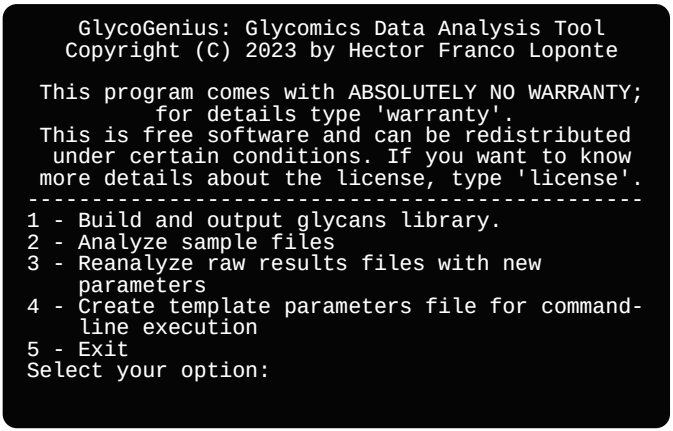
Fig. 8: *Save Image option when right-clicking plots*

The available formats for saving images are SVG, PNG and JPEG. It is recommended to save every picture as SVG, as this format allows for later editing and offers virtually infinite resolution. For example, the graphs can be ungrouped in an SVG picture using PowerPoint, allowing to recolor the graphs or change the format as desired.

Please note that not all graphs perform well as SVG. The 2D plot graph can be very heavy to edit as SVG. For such plots it is recommended to save the image as PNG.

4. Command-Line Interface

4.1 The Welcome Menu



Once you enter the Command-Line Interface (CLI) of GlycoGenius by typing "glycogenius" in the terminal, you'll be greeted with a welcome menu, containing several options for you to choose from:

1. **Build and output glycans library**, which allows you to build a new ".ggl" library file based on various parameters;
2. **Analyze sample files**, which allows you to analyze mzML or MzXML spectra files directly on the terminal;
3. **Reanalyze raw results files with new parameters**, which allows you to open a former analyzed ".gg" file and output new results spreadsheets with different quality criteria thresholds;
4. **Create template parameters file for command-line execution,** which allows GlycoGenius to output a template parameters file (with the same layout as the one that can be saved on the Set Parameters window on the GUI. See Section 3.3.12) that can be pipelined directly to GlycoGenius, in the terminal, for advanced analysis settings;
5. **Exit** for finishing the execution of the CLI.

Additionally, the user can type "license" as an option to get access to the GNU GPL v3+ license used by GlycoGenius or "version" to check the version of GlycoGenius running.

4.2 Build and output glycans library

```
1 - Use a custom glycans list
2 - Generate a library based on possible
    monosaccharides numbers
Select your option:
```

Selecting “Build and output glycans library” will allow you to select from two options:

1. **Use a custom glycans list**, which allows you to input a custom list of glycans, one by one, to be used as a glycan list;
2. **Generate a library based on possible monosaccharides numbers**, which allows you to set the range of monosaccharides to be used on a combinatorically generated library.

Once an option has been chosen, you’ll be able to input the glycans or monosaccharides ranges, as instructed in the CLI and then options that are common for both library generation will be available:

1. **Force class structures?**
   Here you can select whether to force a specific glycan class composition. The options are **'none'/'n_glycans'/'o_glycans'/'gags'.** This works exactly like the “Force Glycan Class” option described at Section 3.3.7;
2. **Type an element to calculate as adduct/Type the maximum number of such adduct**
   Here you can input the elements to be used as adduct, such as ‘Na’ or ‘H’ and the number of each adduct added;
3. **Type the maximum amount of charges**
   Type in the maximum number of charges per glycan molecule to be searched in your sample files;
4. **Do the glycans have a reducing end tag?**
   Type in ‘y’ if your glycans have a reducing end tag, or ‘n’ otherwise. If ‘y’ is chosen, there will be the following options:
   a. **Insert the tag added mass or molecular formula**
      Here you have to insert the added mass of the reducing end tag or the molecular formula of it (ie. 133.0644 or C7H7N3 for GirP or 219.1735 or C13H21N3 for ProA) or 'pep-'+ aminoacids sequence for peptide as tag (ie. pep-NK for the peptide NK);
5. **Is the sample aminated/ethyl-esterified?**
   Whether the sialic acids in the sample were derivatized by amidation/ethyl-esterification;

6. **Are the glycans permethylated**
   Whether the glycans have been permethylated or not;
7. **Are the glycans reduced**
   If no reducing-end tag is selected, you have to answer whether the glycans were reduced or not;
8. **Do you want to include sulfations and/or phosphorylations in your glycans?**
   Whether to include sulfations or phosphorylations in the glycans compositions. If answered positively, the user will be required to input the range of sulfations and phosphorylations;
9. **Insert the path to save the files produced by the script**
   Input the directory path to which the .ggl file generated will be saved to.

Once the library Generation is done, the library files will be saved to the specified directory and can be later on imported by GlycoGenius to be used in an analysis.

4.3 Analyze sample files

Initially, you'll be faced with the same options as 4.2 Build and output glycans library. Once the library parameters are set, you'll be faced with new options:

1. **Do you wish to analyze MS2 data?**
   This option should be answered positively if the user wishes to have MS2 spectra related to glycans annotated. If answered as 'y', the following option will be made available: "Do you want to only output fragments compatible with identified precursor glycan?". This option behaves exactly like *"Only assign fragments compatible with the precursor composition"* described at Section 3.3.9;
2. **What is the accuracy unit you want to input for mz tolerance? Insert the accuracy value for the unit you've chosen**
   Here the user can set the accuracy unit to be used on the data analysis. The options available are 'ppm' or 'mz'. The value for this tolerance unit will be prompted for afterwards;
3. **Insert the start of the retention time interval at which you want to analyze, in minutes/Insert the end of the retention time interval at which you want to analyze, in minutes**
   The retention/migration time analyzed can be trimmed down here, and it will make GlycoGenius analysis much faster. Default values are 0 and 999 mins;

The options available afterwards are options related to the thresholds set for results spreadsheet files, which are:

1. **Insert the maximum PPM error that a detected glycan must have in order to show up in results' table**
   Peaks with absolute PPM error above the set value will be excluded from the results table;
2. **Insert the minimum isotopic fitting score for a glycan in order for it to show up in the results' table**
   Peaks with isotopic fitting score below this value will be excluded from the results table;
3. **Insert the minimum curve fitting score for a glycan in order for it to show up in the results' table**
   Peaks with curve fitting score below this value will be excluded from the results table;
4. **Insert the minimum signal-to-noise ratio that a detected glycan must have in order to show up in results' table**
   Peaks with signal-to-noise ratio below this value will be excluded from the results table;

The last option will be to insert the directory to save the results files to, as shown at Section 4.2.9. Once the analysis is done, the ".gg" analysis results file and the spreadsheets with filtered results will be saved to the specified directory.

4.4 Reanalyze raw results files with new parameters

If this option is selected, you'll be prompted for the path to the folder containing the ".gg" file and for the parameters described at the end of Section 4.3. These parameters will be used to filter out the data and output new results files, which will be saved to the specified directory.

4.5 Create template parameters file for command-line execution

The previous options described allows for library generation, data analysis and results exporting within the CLI, but offer simplified options to the user.

By selecting option '4' on the Welcome Menu, GlycoGenius asks for a directory path to save the template parameters file to and outputs a ".ini" file into the specified folder. This parameters file allows for advanced CLI execution of GlycoGenius and

features all the library building and data analysis settings available at the GUI, with proper descriptions. A copy of the fully-commented template file can be found below:

```
[running_modes]
mode = analysis
;   Select in which mode you want to run Glycogenius:
;   - Use 'analysis' if you want to start a new analysis.
;   - Use 'library' to just generate a new library and
;   stop the execution afterwards.
;   - Use 'reanalysis' to reanalyze an existing .gg
;   analysis file.
;
use_multiple_CPU_cores = yes
number_cores = all
;   Allows to use multiple cores for the processing
;   of the data. If number_cores = all, uses
;   total_cores-2 (ie. if you have a CPU with 20
;   cores, it will use 18 cores).
;
working_directory =
;   Directory to load and save files from script.
;
samples_directory =
;   Directory where sample files are located.
;   -> Only needed in analysis mode.
;
exported_library_name =
;   Choose the name of the library file in library_name.
;   If left blank, library will be exported with the file
;   name 'date_time_glycans_library.ggl'.
;   -> Only needed in library mode.
;
file_for_reanalysis =
;   Indicate the path to the .gg analysis file.
;   -> Only needed in reanalysis mode.
;

[library_building_modes]
mode = generate_library
;   Select in what mode you want to build a library:
;   - Use 'import_library' if you want to use an existing
;   .ggl library file. Ignore common_library_building_
;   settings if importing a library.
;   - Use 'custom_library' if you want to build a library
;   from a list of glycans.
;   - Use 'generate_library' if you want to generate a
;   new library using combinatorial analysis.
;
export_library = no
exported_library_name =
;   Exports the library you generated to the working
;   directory to use in future analysis without having
;   to build a new library. Also creates an excel file
;   containing a human-readable version of the library
;   generated and a file compatible with Skyline's
;   transition list model. Choose the name of the
;   library file in library_name. If left blank,
;   library will be exported with the file name
;   'date_time_glycans_library.ggl'.
;
import_library_path =
;   Indicate the path to an existing .ggl library file
;   or to where and with which name you'd like to save
;   your exported library. If left blank, library will
;   be exported as 'date_time_glycans_library.ggl' at
;   the working directory folder.
;   -> Only needed in import_library mode.
;
custom_glycans_list = H3N2, H5N2, H5N4S2F1
;   Input the glycans directly, comma separated, or
;   indicate the path to a text file containing
;   the list (line or comma separated).
;   Monosaccharides accepted: Hexoses (H),
;   HexNAc (N), Acetyl Sialic Acid (S or Am and E
;   if you have lactonized-ethyl sterified glycans),
;   Glycolyl Sialic Acid (G), Deoxyhexose (F). Case
;   sensitive.
;   -> Only needed in custom_library mode.
;
total_monosaccharides = 5, 22
hexoses = 3, 10
hexosamines = 0, 0
```

```
hexnacs = 2, 8
xyloses = 0, 0
sialic_acids = 0, 4
uronic_acids = 0, 0
fucoses = 0, 2
neu5ac = 0, 4
neu5gc = 0, 0
;   Specify the minimum and maximum monosaccharides
;   amounts for generating a library.
;   -> Only needed in generate_library mode.
;

[common_library_building_settings]
force_class_structure = none
;   Used to force some monosaccharides compositions
;   compatible with structures of N-glycans, O-glycans
;   or GAGs.
;   Options:
;   - none
;   - n_glycans
;   - o_glycans
;   - gags
;
max_adducts = H3
adducts_exclusion =
;   Indicates the desired adducts and their maximum
;   amount. H3Na1 means a maximum of 3 Hydrogens and
;   a maximum of 1 Sodium per adduct combination.
;   Case sensitive. Doesn't work with complex adducts,
;   such as NH4. Set adducts in 'adducts_exclusion' to
;   avoid using specific adducts. Comma separated list.
;
max_charges = 3
;   Limits the maximum amount of calculated charges
;   for each glycan. Set to a negative value if you
;   want to do negative mode analysis [EXPERIMENTAL].
;
reducing_end_tag = 133.0644
;   If a reducing end tag is added to the glycans,
;   insert its added mass or molecular formula here.
;   If no reducing end tag is added to the glycans,
;   set this value to 0. Procainamide: 219.1735 or
;   C13H21N3; Girard Reagent P: 133.0644 or C7H7N3
;   (deprotonated, neutral). Can also be used with
;   a peptide by inputting 'pep-' + the peptide
;   sequence (ie. 'pep-NK' for the peptide NK).
;
permethylated = no
;   If the sample was permethylated, set this
;   parameter to "yes". Doesn't take into account
;   partial permethylation.
;
reduced = no
;   If the sample doesn't have a tag and the glycans
;   had their reducing end reduced, set this to "yes".
;
aminated_ethyl_esterified = no
;   Use if the sialic acids were derivatized with
;   amination (alpha2,3) and ethyl esterification
;   (alpha2,6). Lactonized Acetyl Sialic Acid will
;   be identified as 'Am' and Ethyl Esterified Acetyl
;   Sialic Acid will be identified as 'E'.
;
min_max_sulfation_per_glycan = 0, 0
;   Minimum and maximum amount of sulfation
;   substituents per glycan.
;
min_max_phosphorylation_per_glycan = 0, 0
;   Minimum and maximum amount of phosphorylation
;   substituents per glycan.
;
lyase_digested = no
;   If analyzing GAGs, set whether or not Lyase was used
;   to digeste the polysaccharides.
;
fast_iso = yes
;   Allows you to calculate the isotopic distribution
;   of glycans based only on carbon isotopes (fast,
;   innacurate) and corrected artificially or on all
;   the atoms isotopes (VERY SLOW, very accurate).
;   If not used, library building may take many hours
;   depending on size.
;
high_resolution_isotopic_dist = no
;   If not used, doesn't clump isotope peaks
```

```
;   together, meaning that you'll have more than one
;   isotopic peak in a 1 Da interval. Only use this
;   if you have fast_iso off. Useful when analyzing
;   very high resolution data, such as data acquired
;   on FT mass spectrometers.
;
internal_standard_mass = 0.0
;   If using an internal standard, insert its mass,
;   chemical formula or glycan formula here for
;   GlycoGenius to calculate its area.
;   If a glycan formula is used, reducing end modifications
;   and permethylation (if selected) are applied to the
;   internal standard.
;   This also allows the script to output a normalized
;   metaboanalyst compatible file.
;

[analysis_parameters]
analyze_ms2 = yes
force_fragments_to_glycans = yes
unrestricted_fragments = no
;   Allows to analyze ms2 data, as well. Fragments
;   identified will be associated with each glycan.
;   You can choose to filter identified fragments by
;   monosaccharides compositions, in order to avoid
;   reporting fragments that aren't compatible with
;   detected precursor. If unrestricted_fragments is
;   used, it searches for glycans in every ms2 scan,
;   regardless if the glycan was found in full scan.
;   This will take a bit longer.
;
accuracy_unit = ppm
;   Determines the units of mz tolerance to be used
;   by the script. Options: 'ppm' or 'pw'.
;   'ppm' = Particles per Million, where 10 ppm is
;   around 0.01 mz tolerance at mz 1000, 'mz' = Fixed
;   mz tolerance from centroid, 0.01 mz means it
;   tolerates a 0.01 variance in mz
;
accuracy_value = 10
;   The value for the accuracy_unit parameter. You
;   can use a broader accuracy value and then filter
;   raw data using max_ppm, but this may lead to
;   false positives.
;
ret_time_interval = 0, 999
;   The minimum and maximum retention time, in MINUTES,
;   used for various portions of the script. A shorter
;   interval of ret_time makes the script run faster,
;   so try to trim your sample as much as possible,
;   if you know when your analytes are leaving the
;   column.
;
custom_min_points_per_peak = no
number_points_per_peak = 5
;   If used, set the minimum number of datapoints to
;   consider a chromatogram peak part of the raw
;   dataset. If left on False it calculates
;   automatically.
;
limit_peaks_picked = yes
max_number_peaks = 5
;   If used, picks only the most intense peak on the
;   EIC and up to [max_number_peaks]-1 other peaks
;   closest to it. Warning: This may reduce the range
;   of your results.
;

[post-analysis/reanalysis]
filter_ms2_by_reporter_ions = N1T1, 366.14
;   Set reporter ions to hide MS2 spectra that don't
;   contain them. Can be based with glycans formula
;   (with T being the reducing end of the glycan,
;   including possibly the tag, if used) or an mz.
;
align_chromatograms = yes
;   If enabled, will align the assignments and drawn
;   processed EICs of the different samples. The
;   alignment is highly dependent on the features
;   identified and their inherent quality, so things
;   will change with different quality thresholds.
;
auc_percentage_threshold = 1
;   Allows you to suppress from the analysis peaks
;   that are of the specified percentage (from 0% to
```

```
;   100%) area under curve related to the most intense
;   peak area within the same adduct (ie. if biggest
;   peak has a area under curve of 100 and
;   auc_percentage_threshold is set to 1%, every peak
;   with an auc of 1 and below will be suppressed)
;
minimum_samples = 0
;   Filter out glycans that weren't found in at least
;   a certain number of samples, set by 'minimum_samples'.
;
max_ppm_threshold = 10
;   Maximum PPM error for data curation. If value is
;   greater than equivalent accuracy_value, data won't
;   be filtered by this criterion, as it was already
;   filtered during processing by accuracy_value.
;
isotopic_fitting_score_threshold = 0.9
;   Minimum score of the isotopic distribution fitting
;   in order to consider a mz peak viable.
;
curve_fitting_score_threshold = 0.9
;   Minimum score for the chromatogram peak curve
;   fitting to a gaussian to consider a viable peak.
;
signal_to_noise_threshold = 3
;   Minimum signal-to-noise ratio to consider a
;   chromatogram peak viable. Can be reapplied on
;   raw data reanalysis.
;
output_compositions_analysis = yes
;   If used, also plots data related to the whole
;   composition of each identified glycan in the
;   analysis, in addition to the peak-separated data.
;
output_metaboanalyst_file = no
metaboanalyst_groups = CONTROL, TREATED
;   Here you set up whether or not you want to output
;   a .csv file to be used for plotting data using
;   metaboanalyst. If you want that, you must specify
;   your sample groups, comma separated. Sample
;   groups specified must be present in sample
;   filenames for proper identification. If none is
;   set, samples are defaulted to "ungrouped". Case sensitive.
;
output_fittings_data = no
;   Allows to output files with the fittings data to
;   check scoring criteria. Defaulted to 'no' as
;   these files will be big. Only use it if you really
;   need.
;
output_plot_data = no
;   Allows to output data plotting files for all the
;   EICs drawn by the program. If set to 'no', it will
;   still output the found glycans EIC.
;
```

Once the file parameters have been set, the user can pipeline the file to GlycoGenius CLI by using the command **“cat path/to/parameters/file.ini | glycogenius”** in the preferred terminal to start the execution of the data analysis using the specified settings.

The user can also use the GUI to set the desired parameters and save the parameters file to run it on the CLI afterwards, as described on Section 3.3.12.

5. Outputted Files

During the execution of GlycoGenius' workflow, a number of files will be created at the working directory. Here we will divide the files into two groups and address every one of them.

5.1 Library Files

When you build a library, three files will be created:

- ".ggl" file;
- ".xlsx" file;
- ".csv" file.

5.1.1 The ".ggl" File

The ".ggl" file (with default name "[date]_[time]_glycans_lib.ggl") is the library file format used by GlycoGenius. It consists of two python dictionaries: one containing all the glycans created and the other containing all the settings used to build the library.

This file can be used in GlycoGenius, by pressing the "Import Library" button on the GUI. If you import a library, you don't need to worry about any parameters set on the left column of the "Set Parameters" window, as they will all be provided by the ".ggl" file.

It's recommended to name the ".ggl" file to something that makes it easy for you to understand how the library was built. If you want to check that information one more time, you can also import the library in the GUI and click on "Check Info" (**Fig.1, c.3**).

5.1.2 The .xlsx File

GlycoGenius will also output a human-readable library as an ".xlsx" file with default name "[date]_[time]_ glycans_lib.xlsx". On this file you can check if the library you've built is correct (ie.: the m/z matches the ones you expected, for example).

This file is purely intended for the user to be able to consult/check the library.

5.1.3 The .csv File

The final file created by GlycoGenius library generation is a ".csv" file with default name "[date]_[time]_ glycans_lib_skyline_transitions.csv". This file can be used as a Transitions List in the Skyline software (https://skyline.ms/project/home/software/Skyline/begin.view, Adams KJ et al., 2020).

5.2 Results Files

When you run an analysis and save the results with GlycoGenius, you will receive from 3 to 9 different files:

- A ".gg" file named "[date]_[time]_Analysis.gg";
- From 2 to 6+ ".xlsx" files:
  - A file named "[date]_[time]_Results[qc parameters].xlsx";
  - A file named "[date]_[time]_Found_Glycans_EICs.xlsx";
  - Two optional files named "[date]_[time]_raw_EIC_Plot_Data.xlsx" and "[date]_[time]_processed_EIC_Plot_Data.xlsx";
  - An optional file named "[date]_[time]_curve_fitting_Plot_Data .xlsx";
  - One or more optional files named "[date]_[time]_Isotopic_Fits_ Sample_[sample_number].xlsx".

- From 0 to 4 ".csv" files:
  - An optional file named "[date]_[time]_metaboanalyst_data.csv";
  - An optional file named "[date]_[time]_metaboanalyst_data_ compositions.csv";
  - An optional file named "[date]_[time]_metaboanalyst_data_ normalized.csv";
  - An optional file named "[date]_[time]_metaboanalyst_data_ compositions_normalized.csv".

These optional files are metaboanalyst output files (explained on Section 3.4), and the amount created will vary depending if the user has set to output composition/separated data, in addition to peak-separated data; and if the user has inserted a "*Internal Standard*" in the "*Library Building*" section of the "*Set Parameters*" window (normalized files, see Section 3.3).

5.2.1 The ".gg" File

Whenever you finish running an analysis successfully, a ".gg" Analysis file will be created. This file contains all the information regarding the analysis, including identified glycans, sample information, EICs traced, scores calculated, peaks picked, etc.

This file can be loaded into the GUI by going to "Select Files" and then "Reanalyze .gg File" or by simply double-clicking the file on the Explorer (Windows only, GlycoGenius will ask to create the file association the first time it's run). You can check the conditions under which the analysis was run by clicking the "File Information" button. Once loaded in the GUI, you'll have access to all the analysis information and can

review the data however you feel like, as well as edit the quality control thresholds for saving new output files.

You can also reanalyze ".gg" files in the CLI, but you can only change the thresholds to save new results files.

5.2.2 The ".xlsx" Files

By default, GlycoGenius saves two ".xlsx" files:

- **Results file**, containing all the samples analyzed, glycans identified, their QC scores and annotated fragments, if MS2 was performed;
- **EIC data file**, containing the datapoints for all the EICs of identified glycans. This data can be used for plotting the EICs in other programs, if preferred.

Among the optional ".xlsx" files, you can output files containing the EICs of **all** the glycans in your library, regardless of whether GlycoGenius found them in your sample. These files include raw EICs (traced solely from monoisotopic peaks, without any sort of filtering) as well as the processed ones, which you usually see in the GUI.

The other set of optional .xlsx files includes curve fitting and the isotopic fitting data:

- **Curve fitting file** contains the datapoints for the ideal gaussian traced for every peak, as well as the datapoints that were actually found in the sample, as well as the score for the fitting between both;
- **Isotopic fitting files**, one file per sample, and can be VERY heavy. It contains information on ALL the selected isotopic envelopes peaks used in the analysis, for ALL the peaks of ALL the glycans.

These optional files are meant purely for diagnostic, especially if you are using the CLI and can't check the isotopic fittings, at the very list, directly from the ".gg". Beware that outputting them can be very resource intense and might take some time.

5.2.3 The ".csv" Files

The ".csv files" outputted by GlycoGenius are all optional and are meant to be used as input to MetaboAnalyst (https://www.metaboanalyst.ca/, Pang Z et al., 2024). There can be up to four files, depending if you are using an Internal Standard (that can be set at "Set Parameters" window) and whether or not you choose to output compositions in addition to peak-separated data.

In MetaboAnalyst you can analyze this ".csv" file by selecting "Statistical Analysis [one factor]", on the "Plain text file (.txt or .csv)" field you select "Peak Intensities" for Data

Type and “Samples in Columns” in Format. Load your “.csv” file as a Data File and run the analysis.

Don’t forget to set the groups for your samples (on the “Set Groups” button on the “Save Results” window). Each group requires at least 3 samples for MetaboAnalyst to work properly.

6. The Minimum Amount of Information You Need to Run GlycoGenius Properly

To run a good analysis on GlycoGenius, you need at least the following files/information:

- The LC/CE-MS data file (in ".mzML" or ".mzXML" format, centroided preferably);
- Whether or not your glycans received a reducing end tag, and which reducing end tag is it or what is it's added mass/chemical formula or peptide sequence, if it's a peptide;
- Whether your sialic acids were derivatized or not (at the moment only Amidation/Ethyl-Esterification is supported);
- A rough estimate of the mass accuracy of the equipment used to acquire the data, to set the accuracy values (i.e.: QTOF equipment can be analyzed with 20 PPM tolerance, while Orbitrap equipment can be analyzed with much stricter tolerance, if you wish to, but it's not really required. Iontrap equipment will require a much looser tolerance (~200 PPM or more), for example).

With this, you can set the other parameters (search as the monosaccharides for your search space and retention times from 0 to 999 minutes) to broader values and end up with good data, but there is other information that might make the analysis MUCH faster:

- Know the retention/migration time interval at which your glycans left the column or capillary, so that you can limit the range of the analysis and reduce the analysis time considerably;
- Use centroided spectra files, which can be analyzed much faster than profile spectra. While GlycoGenius can analyze profile spectra, it doesn't benefit from it;
- Have an idea of your target search space, based on the biology of the sample, as a broader search space will result in more compositions, which means more time to analyze.

This extra information can significantly reduce your analysis time by avoiding looking at retention times and for glycans that you might know are not in your sample.

7. GlycoGenius License

GlycoGenius is currently licensed under GNU General Public License v3+ license, available at https://www.gnu.org/licenses/.

GlycoGenius empowers researchers in glycomics with an intuitive and automated workflow for analyzing complex glycan data. This guide provides detailed information of every function the program can achieve and practical tips to help you make the most of its advanced features.

Need help?

Visit the GitHub page to open issues and access the latest updates or e-mail the developer, Hector Franco Loponte, at hectorfloponte@gmail.com, if you have any questions.