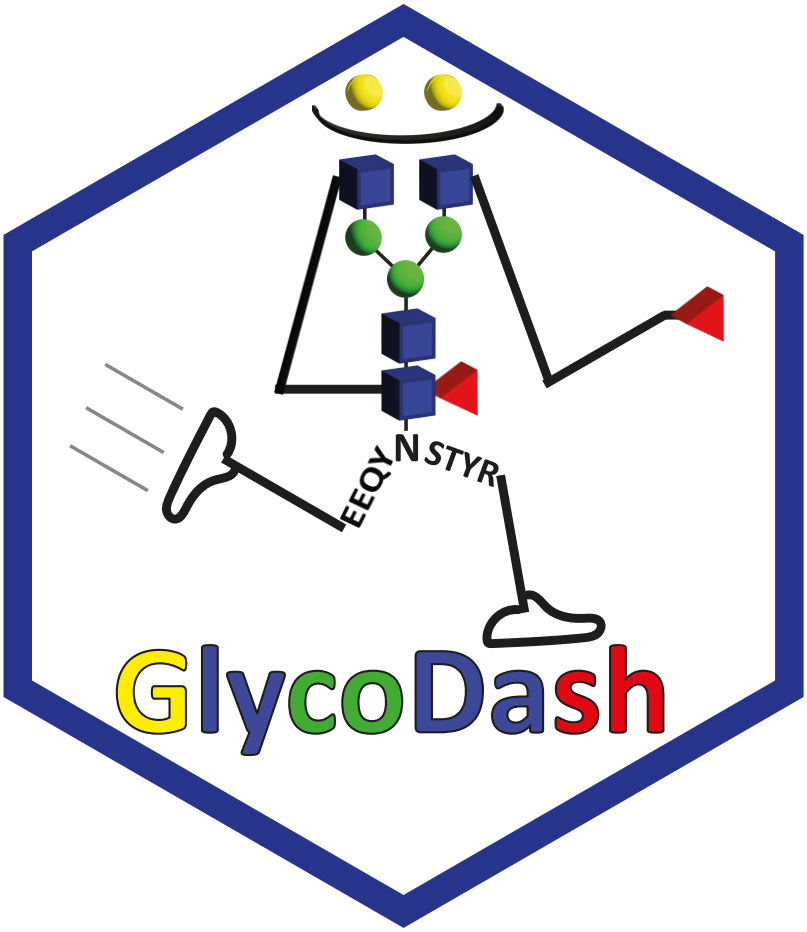
GlycoDash

User Guide



Version 1.6.0

Contents

Installation and requirements

Note: when connected to the LUMC network within Leiden University Medical Center, GlycoDash can be accessed via http://cpmtools:8888/ in your browser. This is a beta version where new features can be tested before they are implemented in an official release.

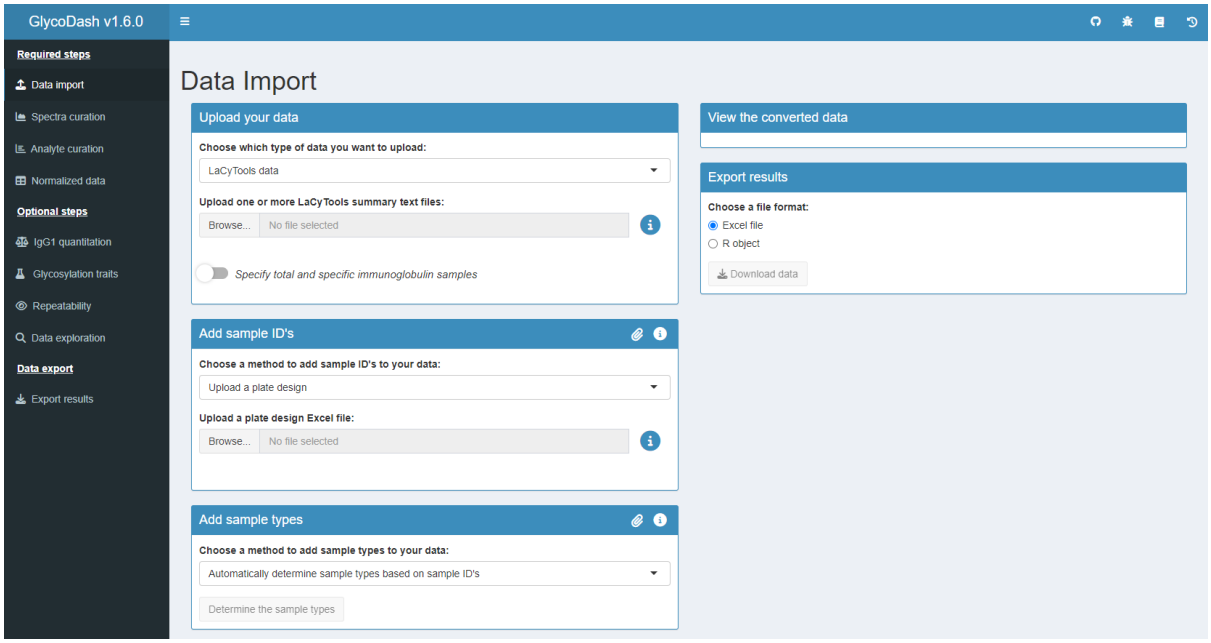
To install the latest official release of GlycoDash on your computer, follow the steps listed in the README section on GitHub: https://github.com/Center-for-Proteomics-and-Metabolomics/GlycoDash

We recommend using Docker to install and run GlycoDash, as this will automatically take care of package and version dependencies [1]. Alternatively, you can run GlycoDash within RStudio [2]. For the development of GlycoDash v1.6.0, R version 4.3.3 is used [3]. The required packages are listed in Appendix 1.

When running GlycoDash in RStudio, having different R and package versions installed may cause GlycoDash to malfunction.

User interface overview

When launching GlycoDash, you will be presented with the page shown below.



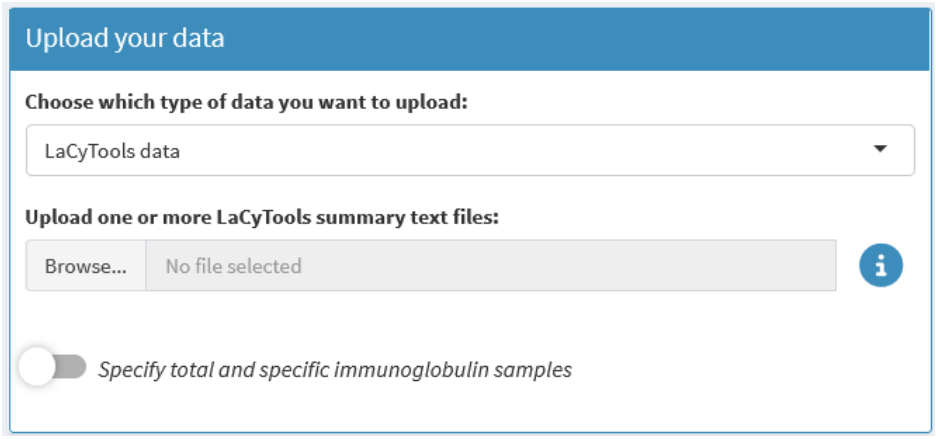
The top-left corner displays the version number of GlycoDash. The left column contains various tabs. The first four tabs (Data Import, Spectra Curation, Analyte Curation and Normalized Data) should be completed in the listed order, as detailed in subsequent sections. Later tabs are optional and can be skipped. The remaining tabs are optional and can be skipped.

The button in the top-right can be clicked to download a changelog of GlycoDash in HTML format.

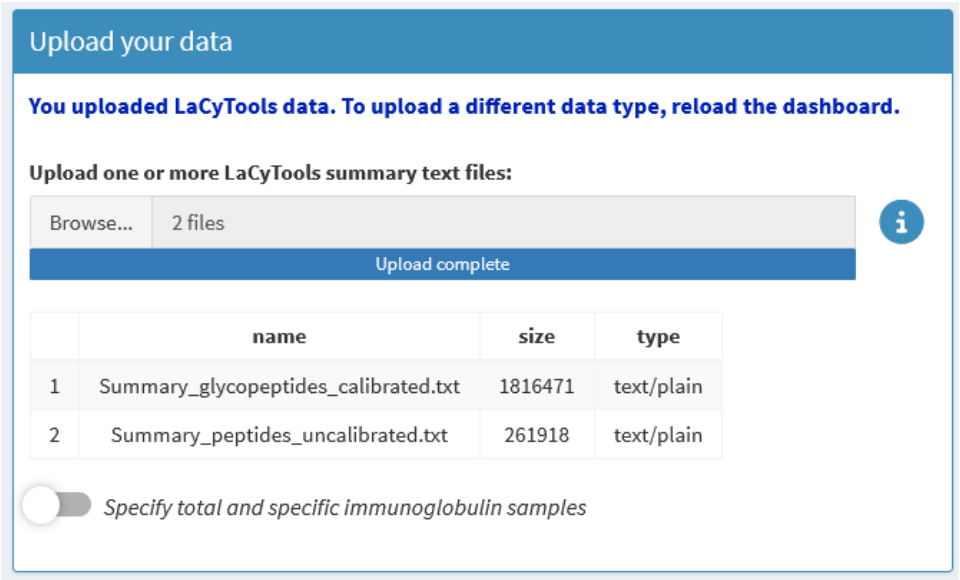
Data import

Raw data input

1. In the "Upload your data" box, choose the type of raw data that you want to upload: LaCyTools data or Skyline data.



2. Upload one or more files of the chosen data type. Hovering over the info box next to the upload box will show information about the formatting requirements. After uploading, the names and sizes of the uploaded files are listed.



- LaCyTools summary text files must contain at least the following outputs:
  - Background subtracted absolute intensity.
  - Mass accuracy (ppm)
  - Isotopic Pattern Quality (IPQ)
  - Signal-to-noise (S/N)

- The required formatting of Skyline CSV requires special attention. Skyline CSV files should contain at least the following columns:
  - *Protein name* – The entries in this column should consist only of letters (do NOT use numbers or spaces). They must specify the peptide to which the glycan is attached (see below). **Different**

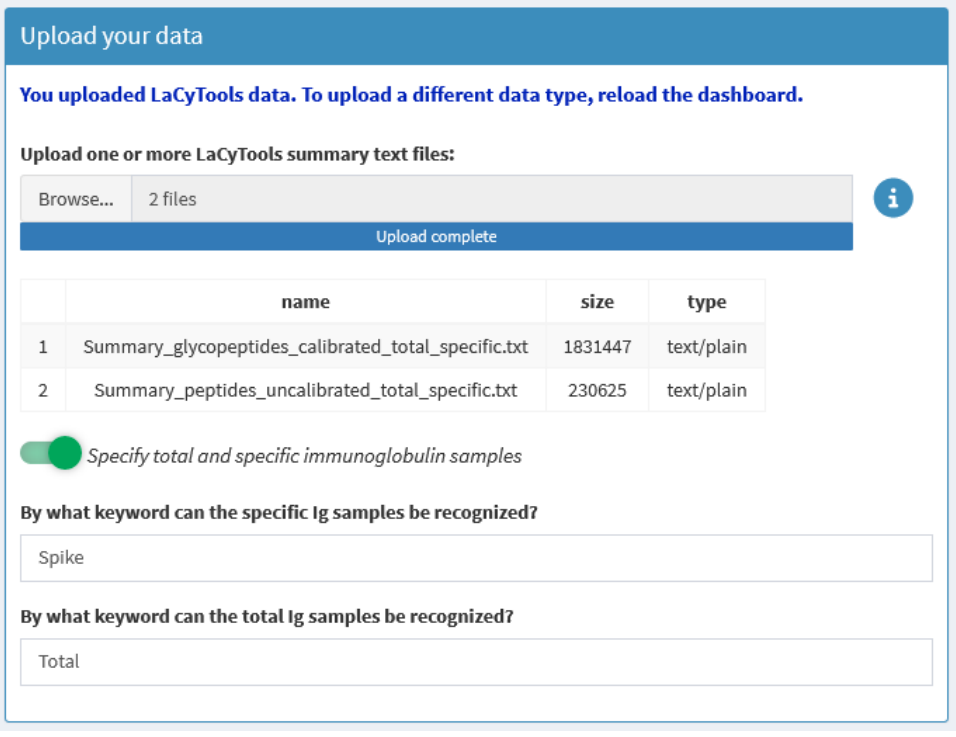
peptides coming from the same protein should have different names in this column.

- *Peptide* – This column should contain the **glycan compositions** of the glycans that are attached to the peptides specified under *Protein Name*. The glycan compositions should have similar formatting to that in LaCyTools, e.g. "H5N2" and "H4N4F1S1".
- *Precursor Charge* – This column should contain numbers that specify the charge state of the glycopeptides.
- For each sample name, the following columns should be present: *Total Area MS1, Isotope Dot Product,* and *Average Mass Error PPM*.

The uploaded data is converted into a clean data frame, with a column for each variable in the data. The converted data is displayed in the top-right box "View the converted data" in the dashboard.

For Skyline data, *Protein name* and *Peptide* are combined into one variable called "analyte" (e.g. "PepI" and "H3N4F1" are combined into analyte name "PepI1H3N4F1"). Additionally, analytes with isomeric glycan compositions are automatically detected and renamed after uploading Skyline data. For example, when "PepIH5N2" is present twice per sample (in each charge state), they are renamed to "PepIH5N2_a" and "PepIH5N2_b".

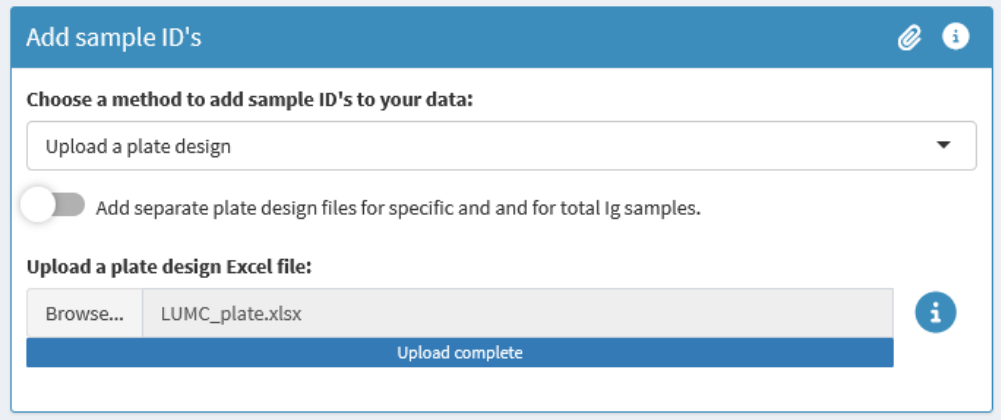
3. **Optional:** when applicable, specify keywords by which total and specific immunoglobulin samples can be recognized in the sample names. The table with converted data will be given an additional column called "group" which specifies whether a sample contains total or specific glycosylation data. Total and specific samples will be separately curated in later steps.



Adding sample ID's

1. Choose a method to add sample ID's:
    - *Upload a plate design* – This method requires each sample name to include a plate number and well position separated by an underscore in the correct format. Examples of correctly formatted samples names are:
        - "MS_**PL01_B09**_01_2020.raw" → *plate 1, well B09*
        - "IM1_**PL02_A11**_03_measurement.raw" → *plate 2, well A11*

      The plate design should be an Excel file, with different plates separated by an empty line. The plates in the plate design are automatically numbered starting at 1. Each cell in a plate layout should contain a sample ID for the corresponding sample. If a plate well was not measured, you can leave the corresponding cell empty. Duplicate sample ID's are allowed.
      An example plate design can be downloaded in GlycoDash by clicking the paperclip button.



      Optional: when applicable, you can upload separate plate designs for total and specific immunoglobulin samples.

    - *Upload a sample list* – Use this method when your sample names do not follow the previously mentioned format. The sample list should be an Excel file with two columns:
        - *sample_name* – This column should contain all the sample names that are present in your data (matching the names in the first column of the converted data).
        - *sample_id* – This column should contain the corresponding sample ID's corresponding to each sample name (e.g. "blank", "pool", "plasma").

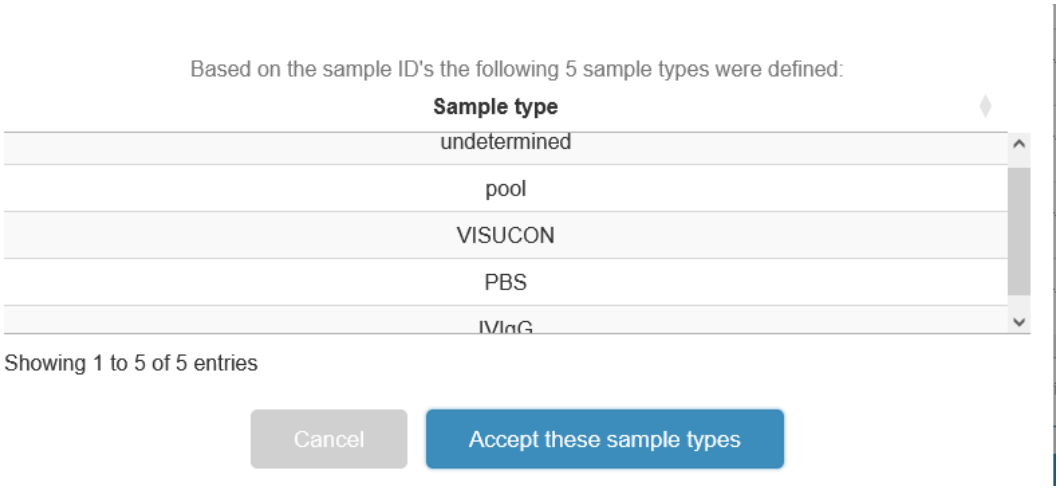
      An example sample list can be downloaded by clicking the paperclip button.

2. Upload the required Excel file. A "sample_id" column is automatically created and shown in the box with the converted data.

Adding sample types

1. Choose a method to add sample types.
    - *Automatically* – Sample types will be detected based on shared substrings of letters in sample ID's. For example, sample ID's "standard_1" and "standard_2" would both be assigned the sample type "standard".
    - *Upload a sample type list* – The list should be an Excel file with two columns:
        - *sample_id* – This column should contain each sample ID once (even if a sample ID appears multiple times in your data).
        - *sample_type* – This column should specify the sample type corresponding to each sample ID. For instance, sample ID "PBS" could be assigned the sample type "Blank". Sample ID's "standard_1", "standard_2" and "standard_3" could all be assigned the sample type "standard".
          An example sample type list can be downloaded by clicking the paperclip button.'

2. Depending on your choice in the previous step:
    - *Automatically* – Click the "Determine the sample types" button. A popup is shown with the detected sample types. To use these detected sample types, click the blue "Accept these sample types" button. If you change your mind and want to use the other method instead, click "Cancel".

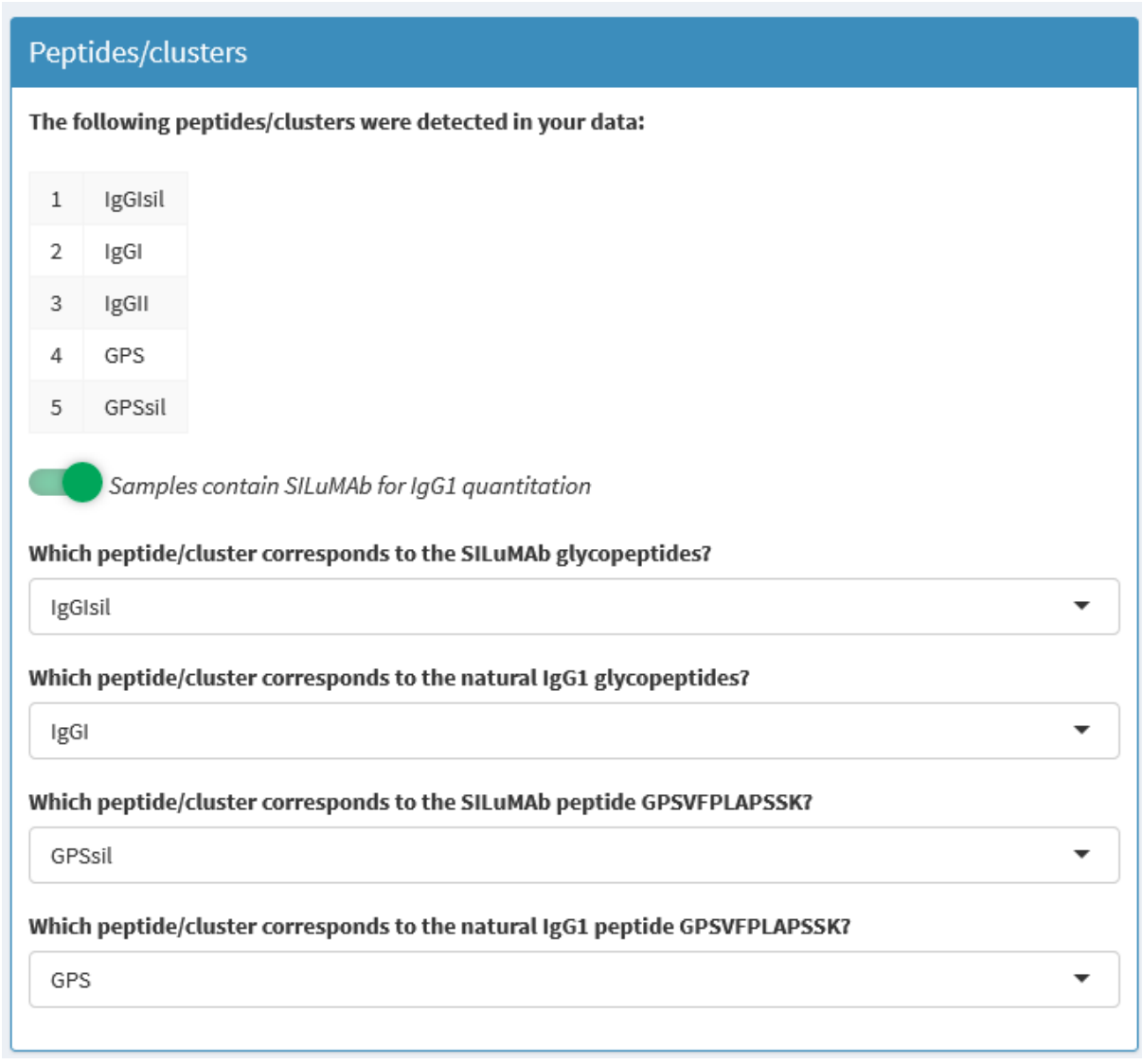


    - *Sample type list* – Upload your Excel file. The sample types are added to your data automatically.

Detection of glycosylation sites

Glycosylation sites in your data are automatically detected based on the analyte names. Each analyte name consists of two parts: the peptide and glycan compositions, separated by a "1". For example, the analyte "IgGI1H3N4F1" represents the glycan H3N4F1 attached to the peptide IgGI. The detected glycosylation sites are listed in a table.

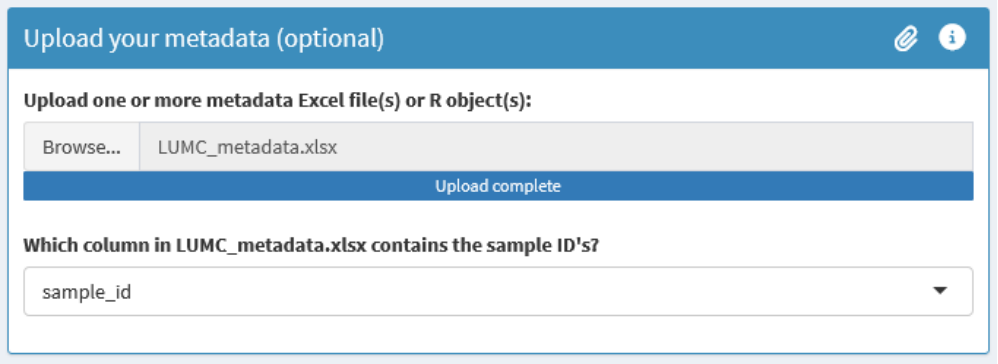
Optional: if your data contains SILuMAb (Stable-Isotope Labeled Universal Monoclonal Antibody) for IgG1 quantitation, select which of the detected glycosylation sites should be used for quantitation. The SIL glycopeptides, as well the SIL and natural versions of the peptide GPSVFPLAPSSK, will be excluded from subsequent curation steps. Natural IgG1 glycopeptides will continue to be utilized for data curation.



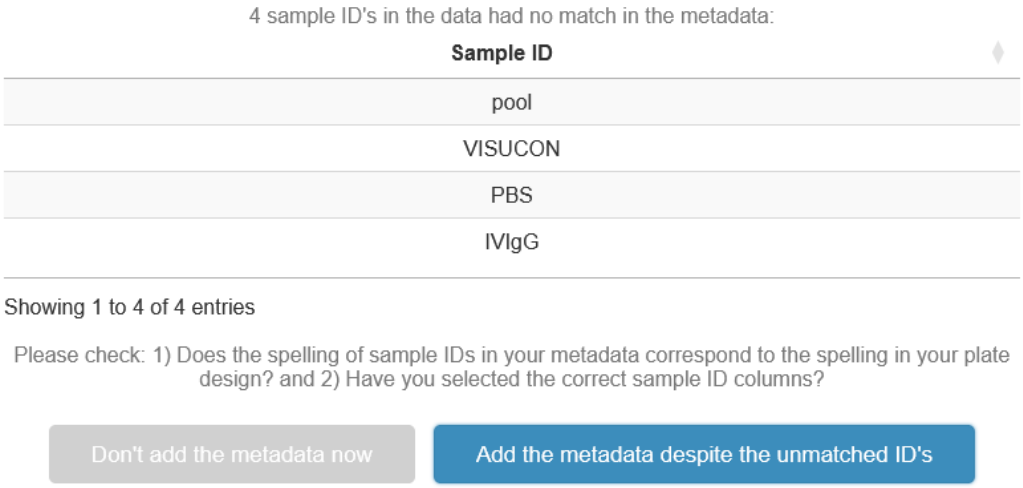
Adding metadata (optional)

You can upload an Excel file containing metadata (e.g. age, sex, disease status), which will be merged with the glycosylation data. This integration ensures that the final output data is ready for immediate further data analysis in a software program of your choice. The metadata can later also be used to perform analyte curation per biological group.

1. Prepare an Excel file with metadata. The file should contain a column with sample ID's (the name of this column does not matter in this case), and additional columns with metadata.
2. Upload the metadata file. In the dropdown box, select the name of the column that contains the sample ID's.



3. One of two things will happen:
    - If all sample ID's in your data have a match in the metadata, then the metadata will immediately be combined with your data and displayed in the box with converted data.
    - If there are sample ID's in your data that do not have a match in the metadata, a warning popup will appear. You can then decide whether you want to add the metadata or not.
      For example, in the example screenshot below, the sample ID's "pool", "VISUCON", "PBS" and "IVIgG" do not have a match in the metadata. This is expected these sample ID's belong to standards. In this case it would be appropriate to add the metadata despite the unmatched ID's.

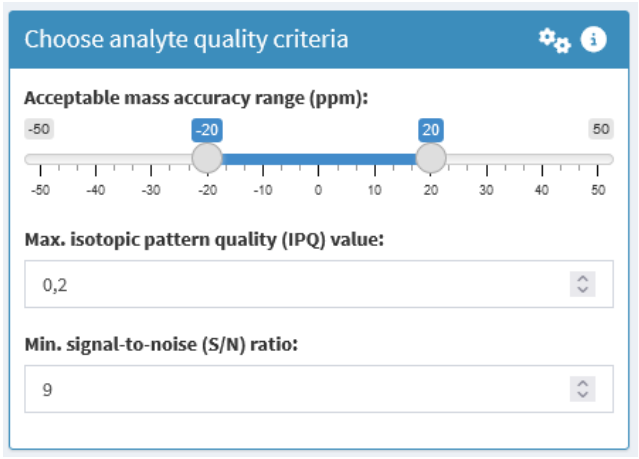


Spectra curation

Choosing analyte quality criteria

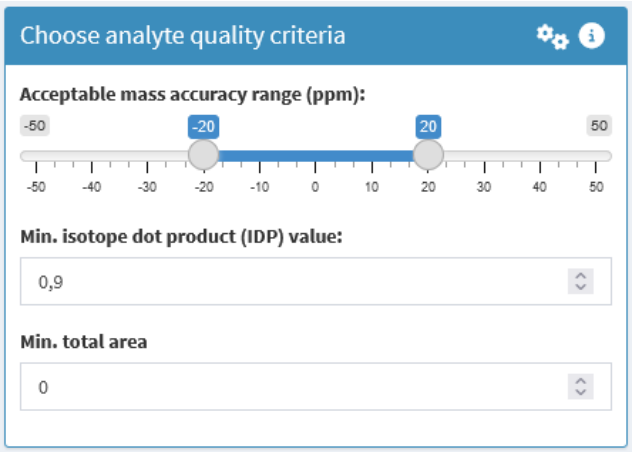
The analyte quality criteria are used to judge whether the signal of an analyte is of sufficient quality in a given sample to be reliably quantified. It is important to note that different charge states for the same glycopeptide are treated as distinct analytes. This information is used for spectra curation and for the subsequent analyte curation step. Depending on the type of data uploaded, there are three distinct quality criteria for which boundaries can be configured. An analyte is deemed to be of sufficient quality when it meets all three criteria. Only in very specific cases, one or two of the listed quality criteria may be used, by clicking the gears icon and deselecting the criteria that you wish to ignore.

LaCyTools data:



- **Acceptable mass accuracy range (ppm)** – Sets the acceptable mass accuracy in ppm. The maximum range is −50 to +50 ppm, with the default being −20 to +20 ppm.
- **Maximum isotopic pattern quality (IPQ) value** – Sets the maximum acceptable IPQ value for an analyte. The better the observed isotopic pattern of an analyte matches its theoretical isotopic pattern, the lower the IPQ will be. The default value is 0.2.
- **Minimum signal-to-noise (S/N) ratio** – The minimum S/N ratio that an analyte should have to be of sufficient quality. The default value is set to 9.

Skyline data:



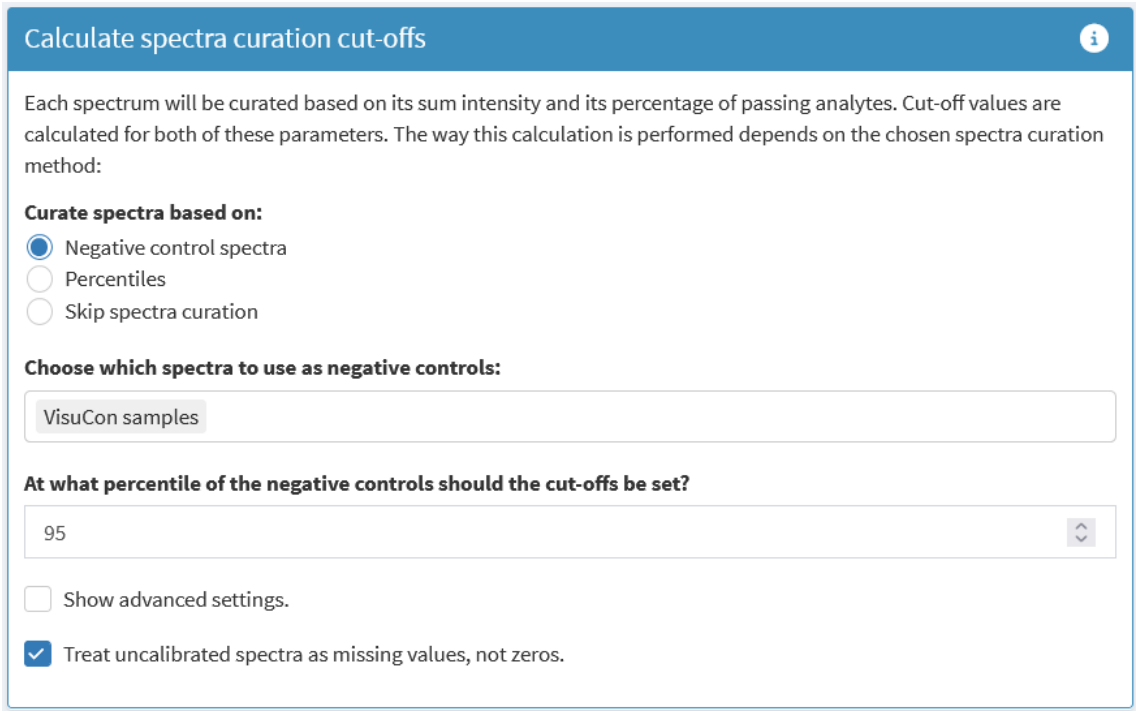
- **Acceptable mass accuracy range (ppm)** – Same as for LaCyTools data.
- **Minimum isotope dot product (IDP) value** – The minimum IDP value that an analyte should have. A higher IDP value means that the observed isotopic pattern fits the theoretical isotopic pattern better, with a maximum value of 1. The default value is set to 0.9.
- **Minimum total area** – As Skyline does not output S/N, the total area can be used as an alternative. By default this value is set to 0 (i.e. it is ignored).

Spectra curation cut-offs

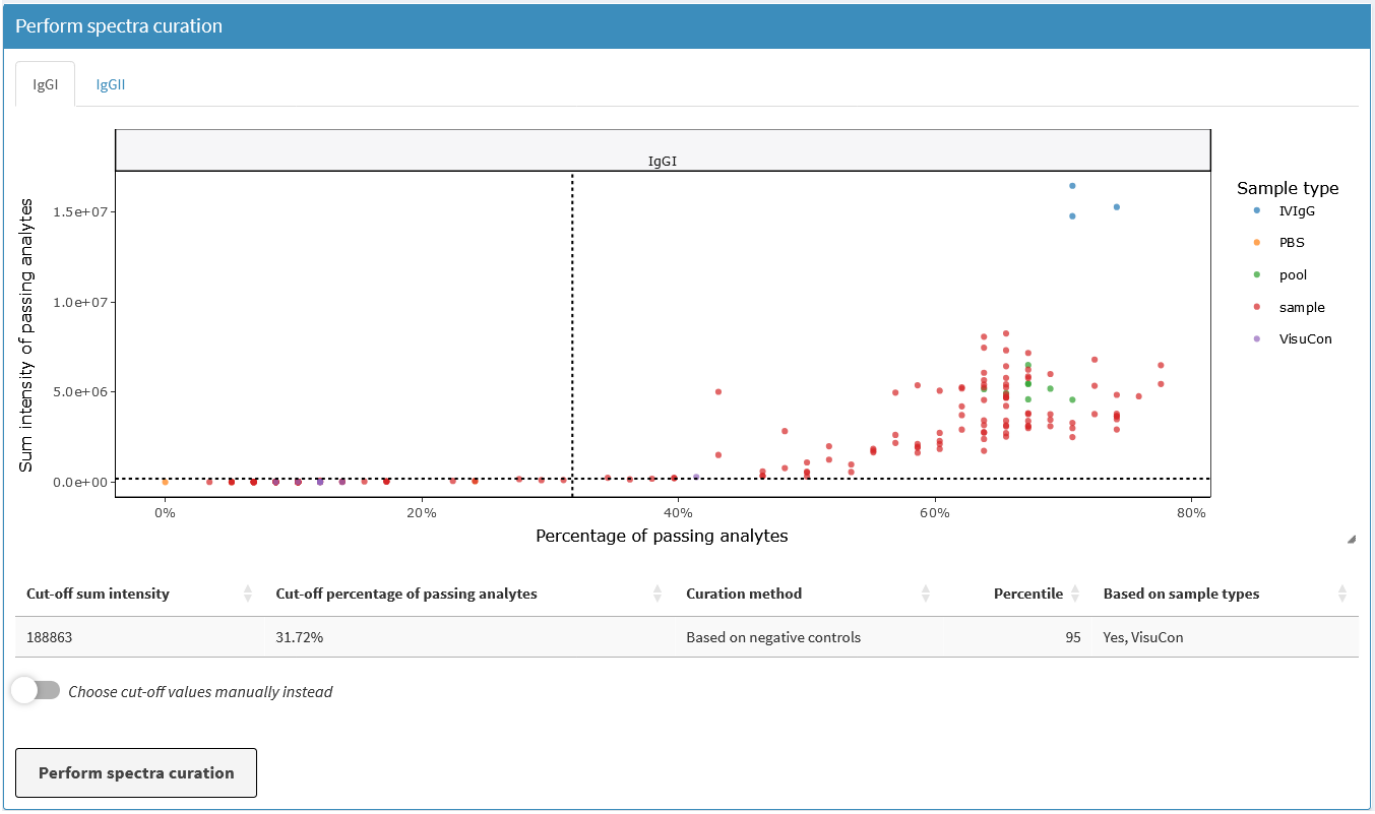
For each sample and glycosylation site in your data, the intensities of all passing analytes are summed, and the percentage of passing analytes out of all targeted analytes is calculated. Each spectrum will be curated based on these values. The two available methods for curating the spectra are described below. Alternatively, you have the option to skip spectra curation, in which case all spectra will be utilized in subsequent processing steps.

Curate spectra based on negative controls:

1. Choose which sample types should be used as negative controls.
2. Set a percentile of negative control measurements for calculating cut-off values. The default percentile is 95. GlycoDash will automatically compute cut-off values for both the percentage and sum intensity of passing analytes. These cut-offs are calculated separately for each glycosylation site.
    - By default, uncalibrated spectra are treated as missing values and are excluded from the calculations. You have the option to include them with a value of zero instead by unchecking the checkbox "Treat uncalibrated spectra as missing values, not zeros".



During curation, spectra whose percentage of passing analytes and sum intensity exceed the corresponding cut-off values will pass curation, and will be used for further data processing. In the "Perform spectra curation box", all spectra are displayed per glycosylation site in an interactive plot, color-coded by sample type. The cut-offs are indicated by dashed bars, and the exact cut-off values are displayed below the plot for reference.

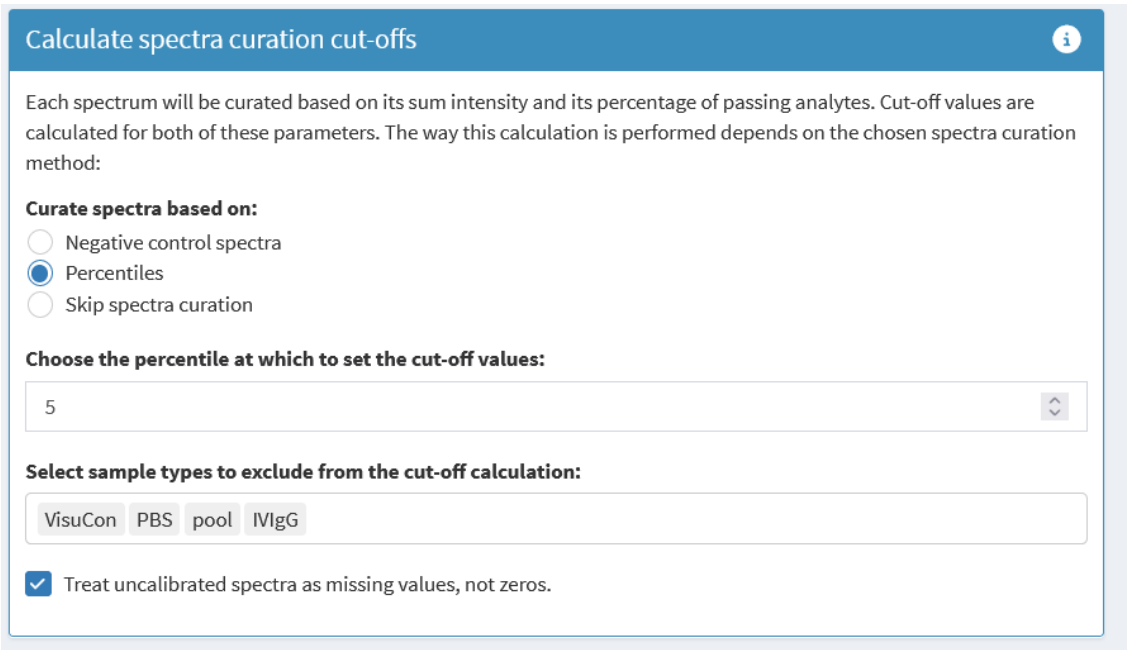


3. If you are satisfied with the cut-offs, click the “Perform spectra curation” button.

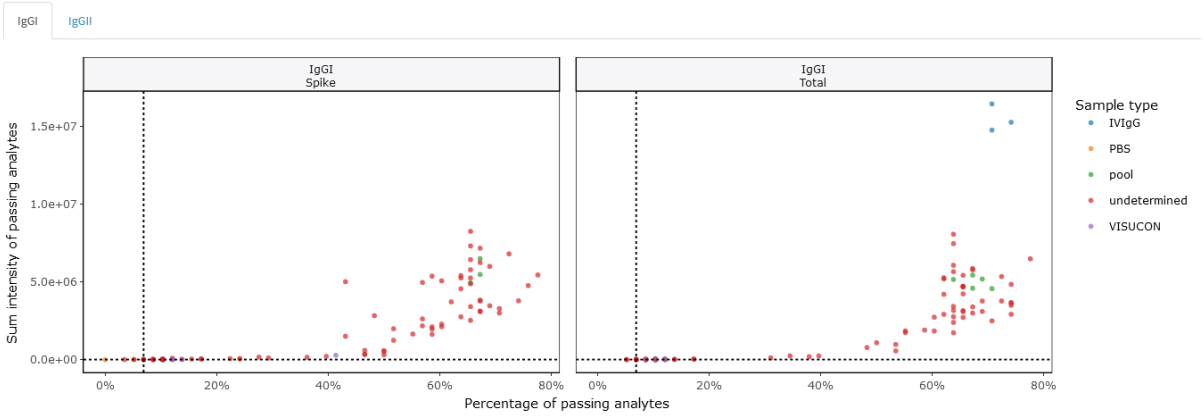
Curate spectra based on percentiles

In this scenario, cut-off values are determined at a specific percentile based on all measurements associated with a glycosylation site, with the option to exclude certain sample types from the calculation. This means a fixed percentile of all data, that with the lowest data quality, is excluded from further processing.

1. Choose the percentile at which to set the cut-off values. The default value is 5.
2. Optional: exclude sample types from the calculations. We recommend to exclude all types of controls. Excluded types will be curated based on the calculated cut-off.
3. Review the cut-offs in the “Perform spectra curation” box.
4. Once satisfied with the cut-offs, click the “Perform spectra curation” button.



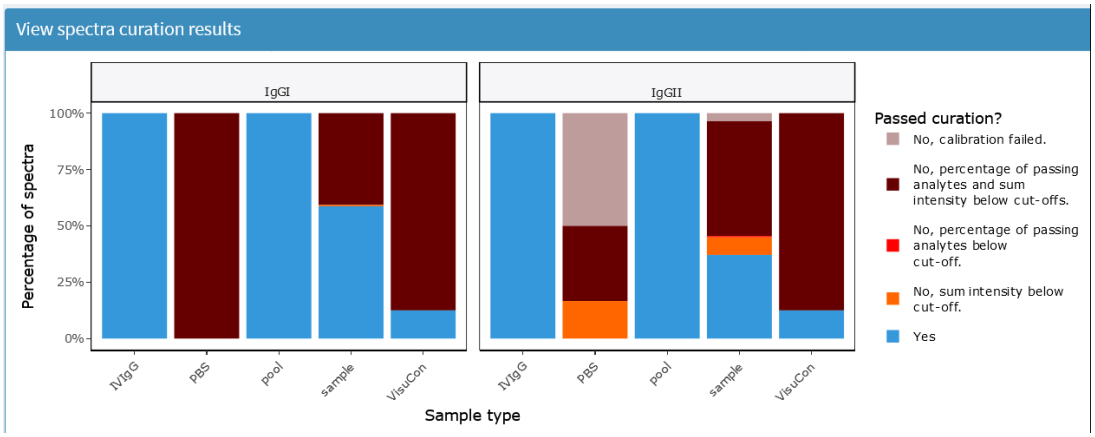
Note: in cases where your data includes both total and specific immunoglobulin samples, cut-off values will be set independently for total and specific spectra. When curating based on negative control spectra, which sample types to use as negative controls must be selected separately for total and specific spectra.



Group	Cut-off sum intensity	Cut-off percentage of passing analytes	Curation method	Percentile	Based on sample types
Spike	5272	6.90%	Based on percentiles	5	Yes, undetermined, VISUCON, PBS and pool
Total	6161	6.90%	Based on percentiles	5	Yes, undetermined, pool, VISUCON, IVIgG and PBS

Spectra curation results

After performing spectra curation, an interactive plot displaying the results is generated. The plot shows the percentage of passing and failing spectra for each sample type per glycosylation site. The reasons for failing curation are indicated by different colors. For LaCyTools input, the percentage of uncalibrated spectra is also displayed (first pass spectra curation).



Below the plot, three tables are shown with:

- Details of passing spectra per analyte.
- Overview of failed spectra.
- Details of failed spectra per analyte.

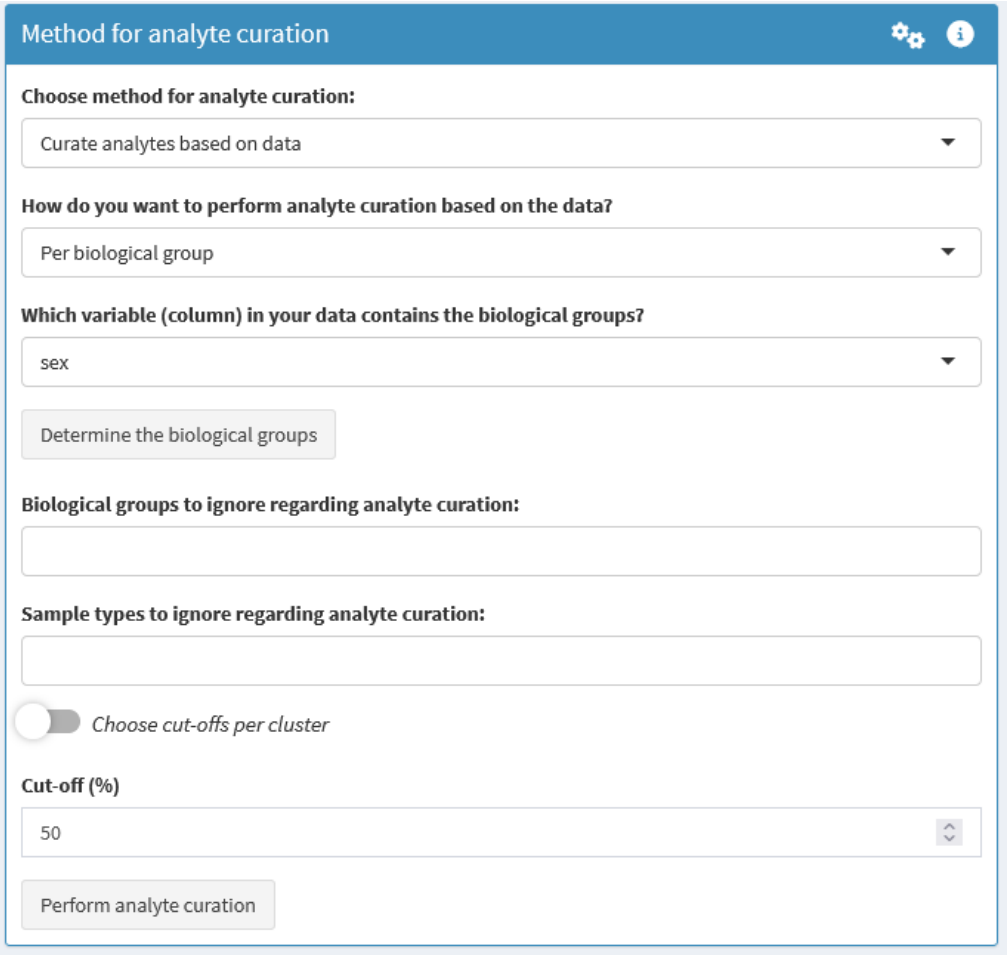
There is an option to download each of these tables as an Excel file at the bottom of the page, in the “Export results” box.

Analyte curation

Analyte curation methods

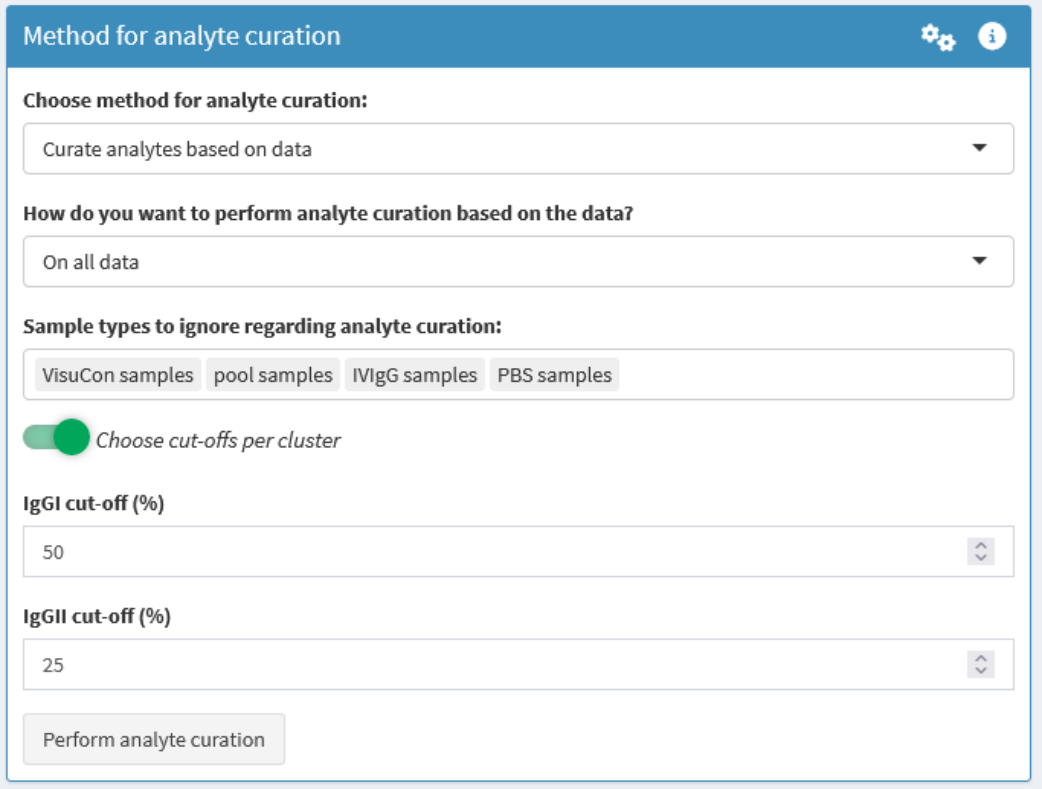
There are two main methods for analyte curation. The preferred method is to curate the analytes based on the data.* There are three different ways to do so:

1. *Per biological group* – An analyte passes curation if it meets the quality criteria in a percentage of spectra exceeding the chosen cut-off in at least one biological group (e.g. disease status or treatment). The analyte is then used for total area normalization in all samples, irrespective of the biological group.
   - Choose a variable (column) in your data that contains the biological groups.
   - Click the "Determine the biological groups" button.
   - A popup, displaying detected biological groups, appears for verification.
   - Optional: select biological groups to ignore (e.g. when dealing with a small number of samples in a specific group).
   - Optional: select sample types to ignore. Sample types without a biological group assigned (e.g. blanks) are automatically excluded, even if not selected here.
   - Choose a cut-off or different cut-offs per cluster.
   - Click the "Perform analyte curation" button.



* Alternatively, you can choose to supply an analyte list (as an Excel file or R object) that contains the names of all the analytes that should be kept. All charge states of these analytes will pass curation. This method should only be used if you performed analyte curation based on your data outside of GlycoDash.

2. *On all data* – When an analyte in a given charge state fulfills the quality criteria (as defined in the spectra curation step) in a percentage of spectra exceeding a chosen cut-off percentage, then it passes curation and is used for total area normalization during the next step in all samples.
    - There is an option to ignore certain sample types in the assessment. For instance, we recommend to exclude all controls.
    - Choose a cut-off, the default is 50%, which is applied to all glycosylation sites. To choose different cut-off percentages per cluster, toggle the *"Choose cut-offs per cluster"* switch.

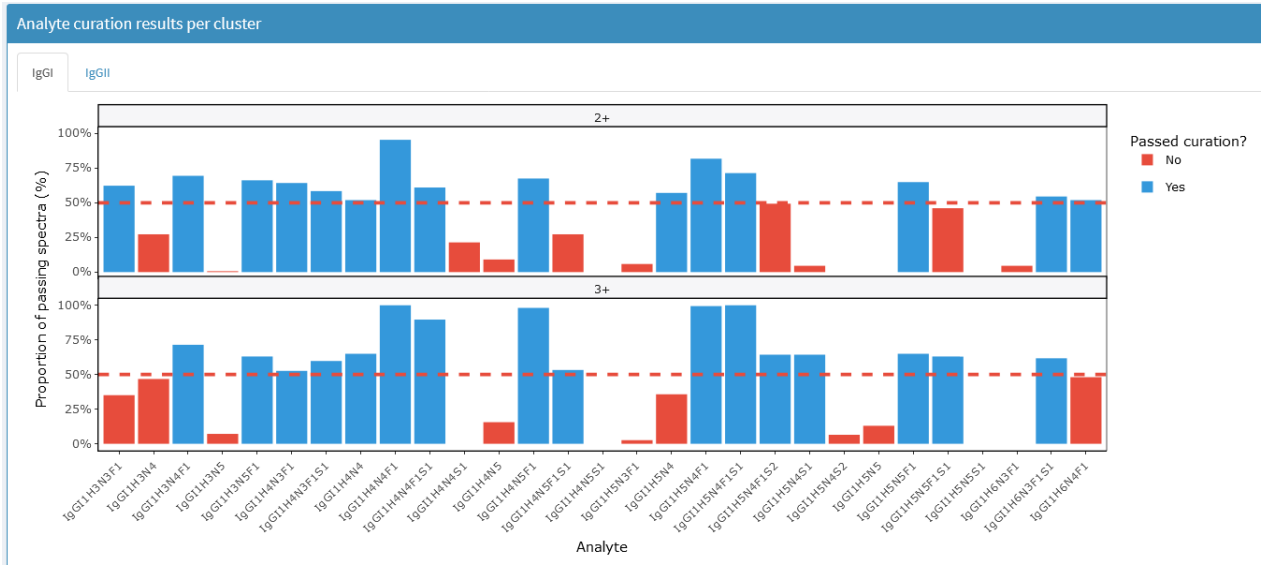


3. *Per sample* – The data is curated per sample. If an analyte meets the quality criteria in one sample, it is used for total area normalization in that sample. However, if the analyte does not meet the quality criteria in another sample, it will not be used for normalization in that particular sample.

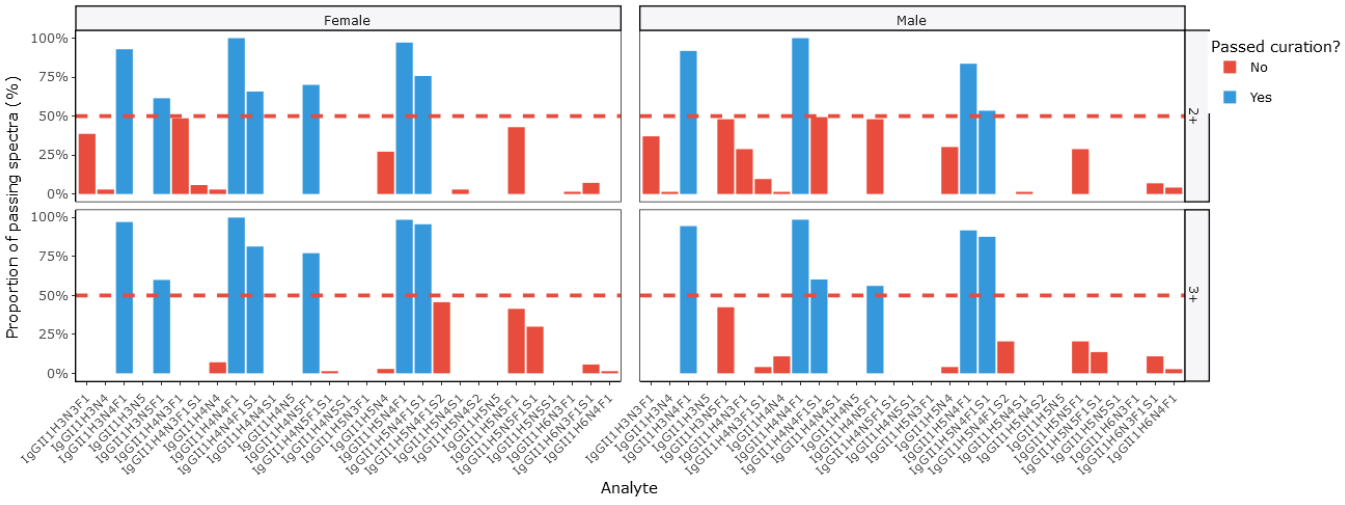
Analyte curation results

When performing analyte curation per sample, no plots or results will be displayed.

For analytes that were curated based on all data, an interactive plot is generated for each glycosylation site. Each analyte and charge state is represented by a bar in the plot, color-coded to indicate whether it passed curation.

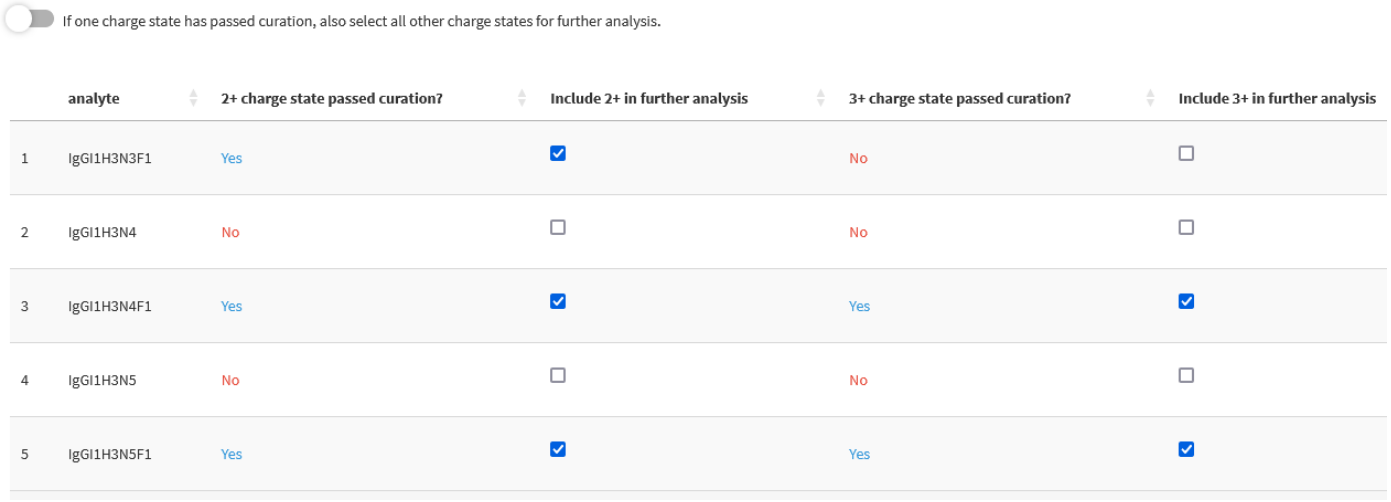


If analyte curation is performed per biological group, similar plots are generated for each biological group.



Following the curation of analytes based on all data or per biological group, a table is generated below the plots. This table lists all analyte names and includes two columns for each charge state: one indicating whether the analyte passed curation ("Yes" or "No"), and another with a checkbox to include the analyte in that charge state for further analysis.

Charge states of analytes that passed curation are pre-selected. Via the checkboxes, any analyte and charge state can de excluded or included for further analysis, if external information warrants this. Additionally, there is an option to automatically select all charge states of an analyte when it passes in at least one charge state. Strong interferences should be excluded in this case, e.g. by manual inspection of raw data or LaCyTools/Skyline output. Then, this likely increases accuracy at the expense of precision.



Note: if your data contains total and specific immunoglobulins, you have the option to exclude either one as a sample type during analyte curation. It is advisable to exclude total samples and base the curation solely on specific samples.

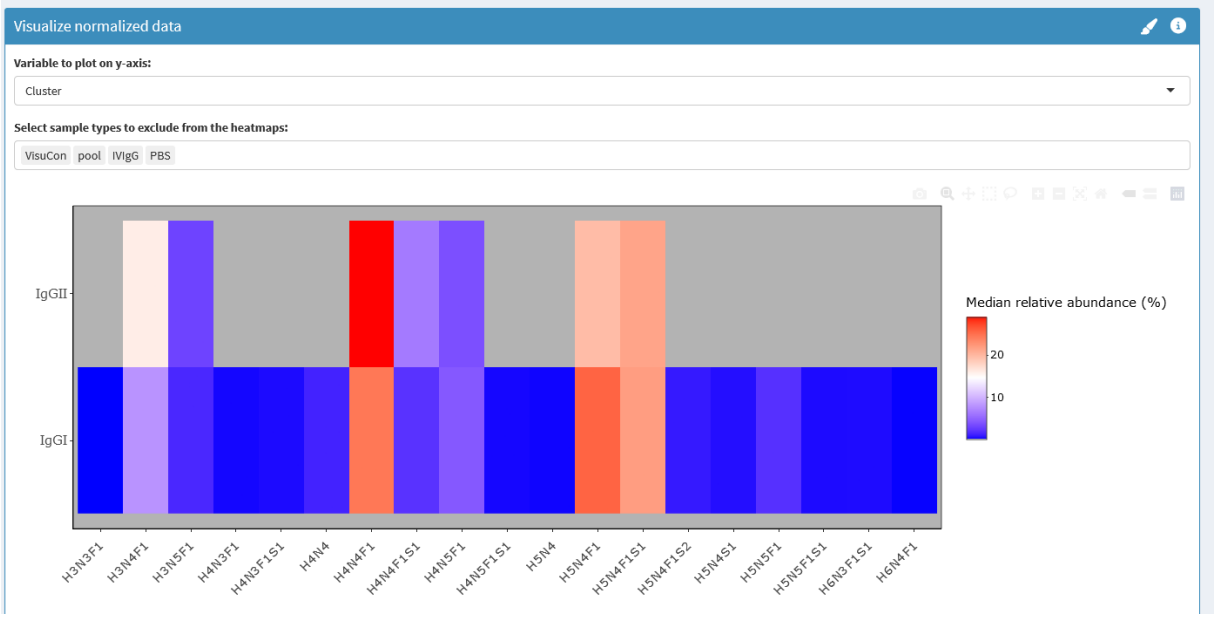
Normalized data

Following analyte curation, total area normalization is automatically conducted on the data per glycosylation site. A table displaying the normalized data is generated, including the following information:

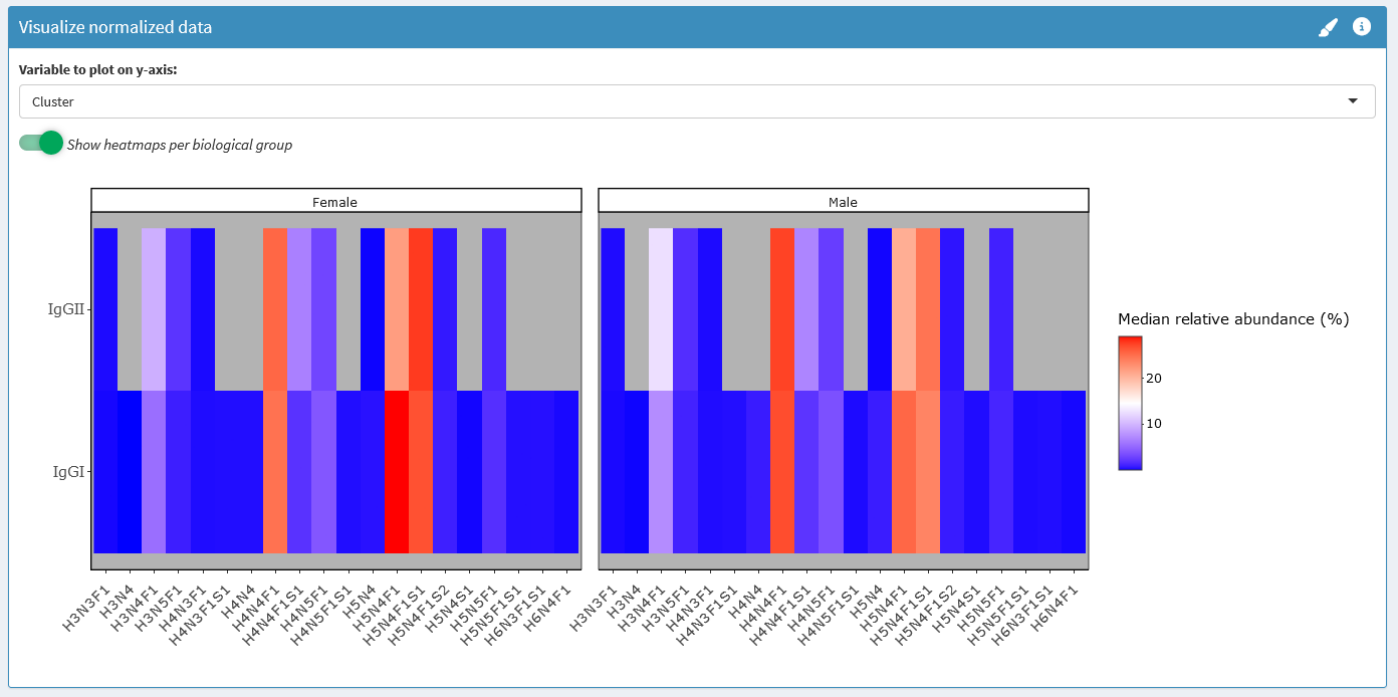
- Sample names
- Sample types
- Sample ID's
- Plate well positions, if applicable.
- The number of replicates for each sample ID (e.g. if there are ten pools in your data all labeled with the same sample ID "pool", then these will all have a value of 10 in the "number_of_replicates" column.
- Metadata, if applicable.
- Per glycosylation site: a column showing sum intensities.
- For each analyte: a column displaying relative abundances in percentages. The analytes associated with a specific glycosylation site will collectively sum to 100%.

Above the table, there is an option to normalize the charge states separately instead of combining them. At the bottom of the page, there is an option to download the normalized data as an Excel file or R object.

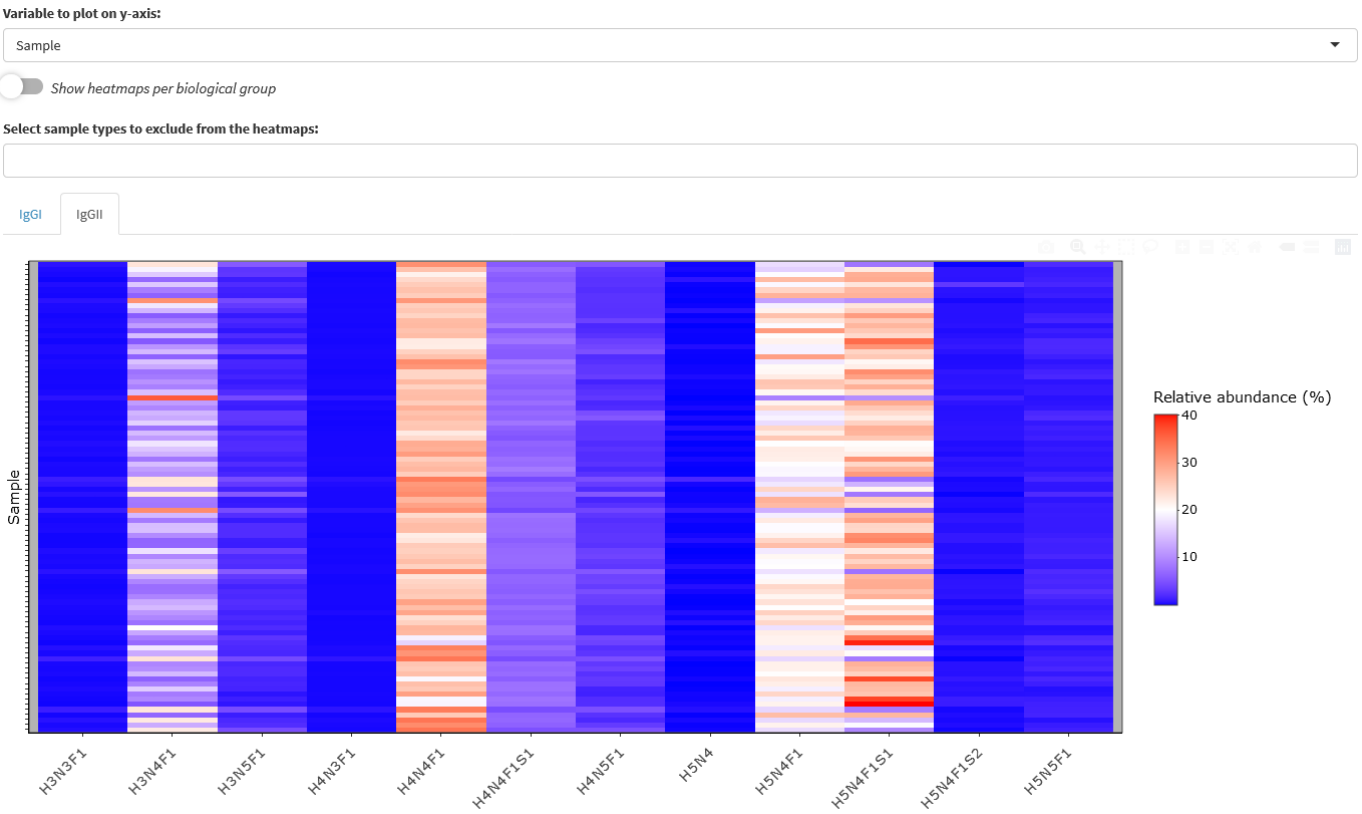
The normalized data is visually represented in one or more heat maps, with glycan compositions on the x-axis and glycosylation sites on the y-axis by default. Each color in the heat map corresponds to the median relative abundance for a specific peptide/glycan combination, calculated across all samples. It is possible to exclude certain sample types from the calculations.



If analyte curation was performed per biological group, there will be an option to separate heat maps per biological group (this option is enabled by default).



Alternatively, the individual samples can be plotted on the y-axis. In this case, a heat map is generated for each glycosylation site, with the ability to exclude sample types.



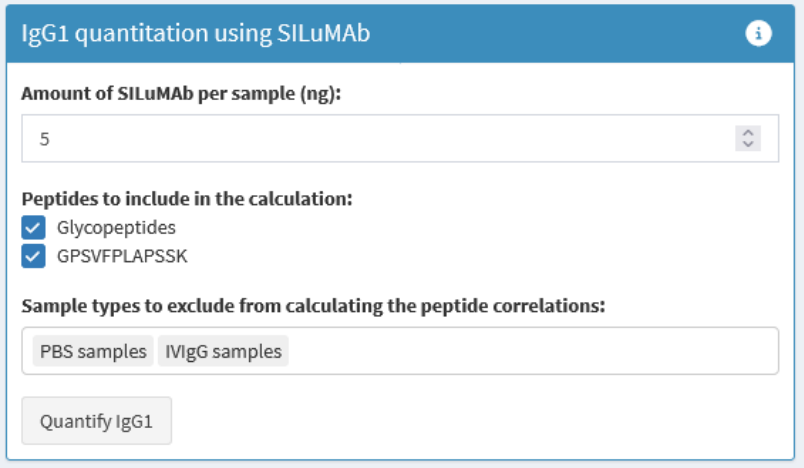
When plotting samples on the y-axis, there is also an option to create separate heat maps per biological group, if applicable.

The colors in the heat maps can be customized by clicking the paintbrush icon in the top-right corner of the box.

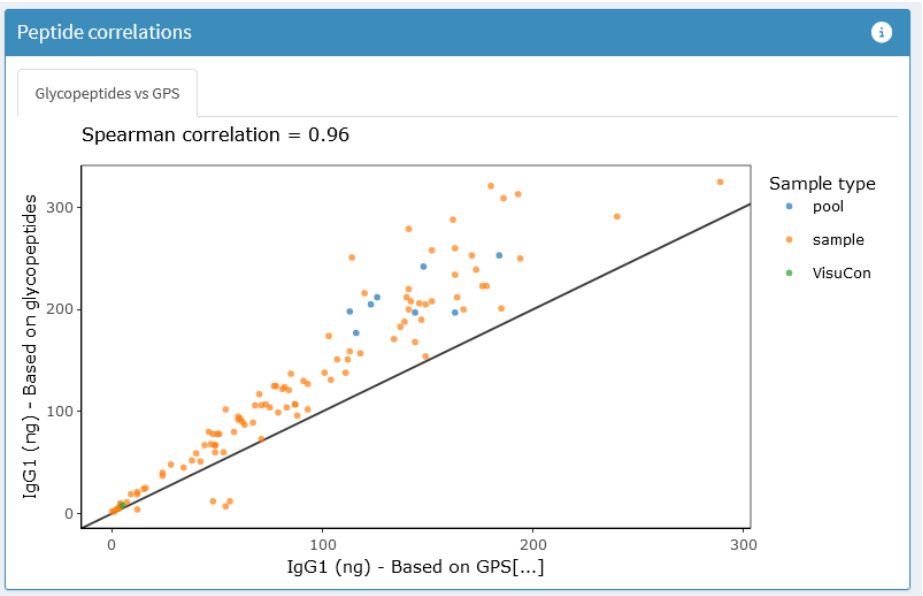
IgG1 quantitation (optional)

This option can only be used when your samples contain SILuMAb for IgG1 quantitation, and when you have specified which (glyco-)peptides should be used for quantitation during data import. For a detailed explanation of how IgG1 can be quantified using SILuMAb, please refer to previous reports. A brief summary of the quantitation method is also provided in the info boxes.

1. Specify the amount of SILuMAb (in ng) present in each sample.
2. Select the peptides to use for IgG1 quantitation. By default, both the glycopeptides and the GPSVFPLAPSSSK peptide are used for quantitation.
3. Optionally, choose sample types for which IgG1 should not be quantified.
4. Click the "Quantify IgG1" button.



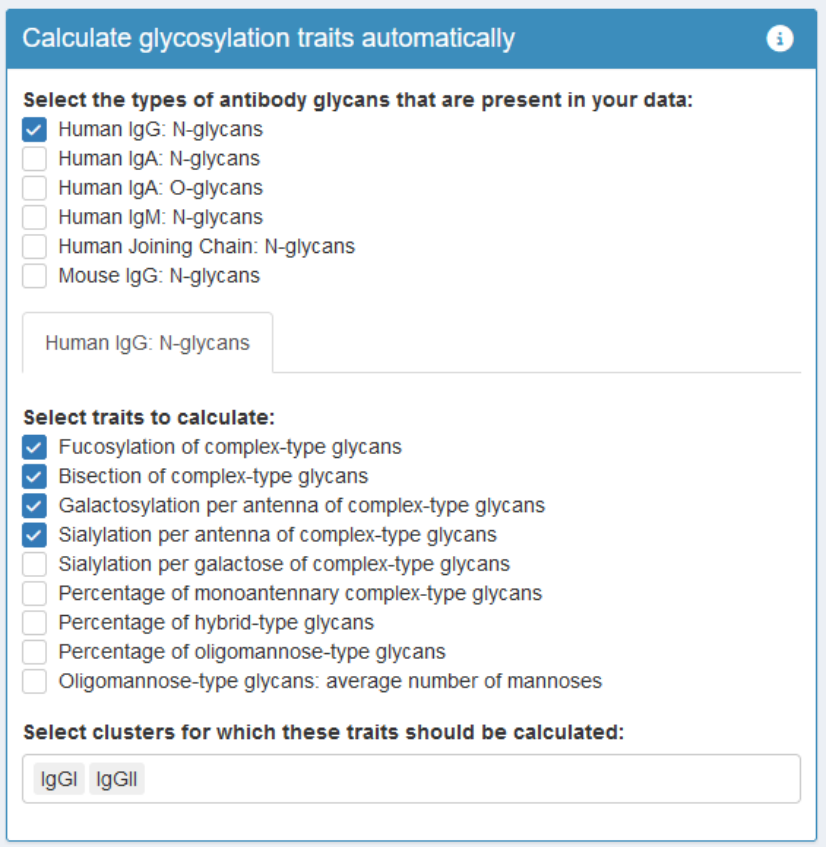
If you quantified IgG1 using both the glycopeptides and the GPSVFPLAPSSSK peptide, a plot is generated to visualize the correlation between the IgG1 quantities calculated based on the two different peptides (the reported IgG1 quantities are determined as the median of the two). The diagonal line represents the line of equality ($y = x$). The closer data points are to this line, the better the agreement between the quantities calculated based on both peptides.



Glycosylation traits (optional)

You can automatically calculate glycosylation traits for human IgG, IgA and IgM (including Joining Chain), and for mouse IgG. These calculations rely on a reference list containing *N*-glycan compositions with known structures, listed in Appendix 2. If your data includes glycan compositions not listed there, a warning message will be displayed when using this option.

1. Select the types of glycans that are present in your data.
2. Choose the traits you wish to calculate.
3. Select the glycosylation sites in your data for which you want to calculate the traits. The calculations are performed automatically.



A table displaying the formulas used for calculating the traits will be generated. This table can be downloaded as an Excel file, which can be easily modified and used for calculating custom glycosylation traits (described below). In some cases a message is shown and a trait is not reported, namely when:

- A trait equals 100 for all samples;
- A trait equals 0 for all samples;
- A trait would be calculated using only one glycan.

In the example screenshot below, the trait “IgGII_fucosylation” is not reported because it is 100 for all samples.

Formulas used to automatically calculate the glycosylation traits

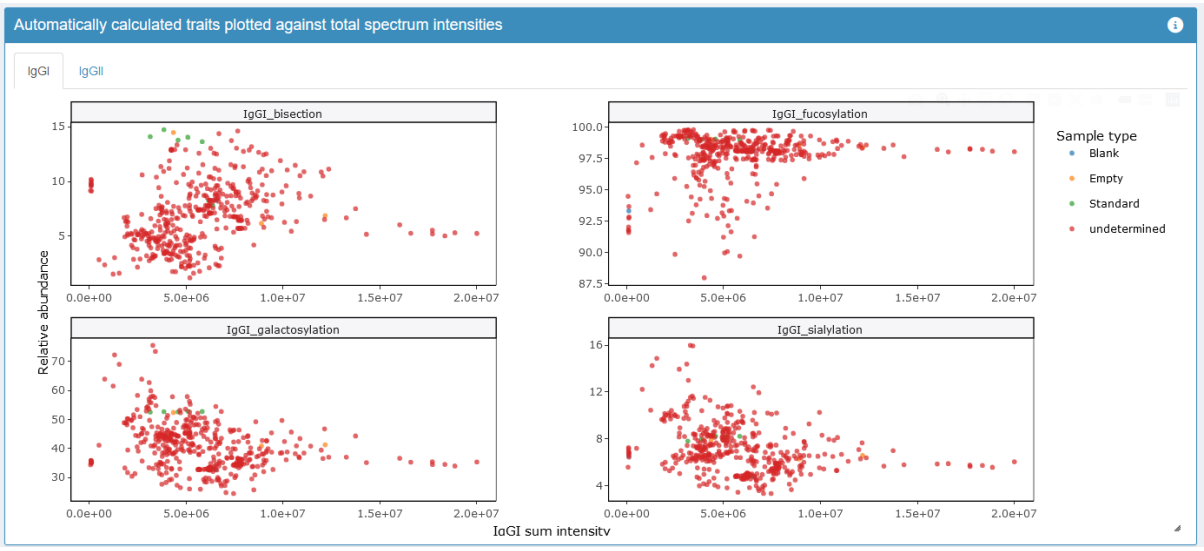
Download as Excel file

Show 10 entries Search:

trait	formula
All	All
IgGI_fucosylation	1 * (IgGI1H3N4F1 + IgGI1H4N4F1 + IgGI1H5N4F1 + IgGI1H4N4F1S1 + IgGI1H5N4F1S1 + IgGI1H5N4F1S2 + IgGI1H3N5F1 + IgGI1H4N5F1 + IgGI1H4N5F1S1)
IgGI_bisection	1 * (IgGI1H3N5F1 + IgGI1H4N5F1 + IgGI1H4N5F1S1)
IgGI_galactosylation	0.5 * (IgGI1H4N4F1 + IgGI1H4N4F1S1 + IgGI1H4N5F1 + IgGI1H4N5F1S1) + 1 * (IgGI1H5N4F1 + IgGI1H5N4F1S1 + IgGI1H5N4F1S2)
IgGI_sialylation	0.5 * (IgGI1H4N4F1S1 + IgGI1H5N4F1S1 + IgGI1H4N5F1S1) + 1 * IgGI1H5N4F1S2
IgGII_fucosylation	Not reported: 100 for all samples
IgGII_bisection	1 * (IgGII1H3N5F1 + IgGII1H4N5F1 + IgGII1H5N5F1)
IgGII_galactosylation	0.5 * (IgGII1H4N4F1 + IgGII1H4N4F1S1 + IgGII1H4N5F1) + 1 * (IgGII1H5N4F1 + IgGII1H5N4F1S1 + IgGII1H5N5F1)
IgGII_sialylation	0.5 * (IgGII1H4N4F1S1 + IgGII1H5N4F1S1)

Showing 1 to 8 of 8 entries Previous 1 Next

For each glycosylation site, the automatically calculated traits are plotted against total spectrum intensities as a sanity check. Correlations between traits and total spectrum intensities for standards (e.g. Pools or VisuCon) indicate that differences in traits between samples are (at least partly) a technical artefact, rather than a biological effect.



If your data is not suitable for automatic glycosylation trait calculation, there is an option to calculate them manually.

1. Create an Excel file with formulas specifying how traits should be calculated. An example Excel file can be download by clicking the paperclip icon. The excel should consist of two columns:
   - *"trait"* – Containing the names of all the glycosylation traits you wish to calculate. Trait names should not include spaces.
   - *"formula"* – Containing the formulas for the glycosylation traits. Analyte names in the formulas should include both the peptide and glycan composition, e.g. "IgGI1H4N4F1".

2. Upload the Excel file in the "Calculate custom glycosylation traits" box.

The traits will be calculated, and a table listing the trait names and formulas will be provided as a sanity check.

Formulas used to calculate the custom glycosylation traits

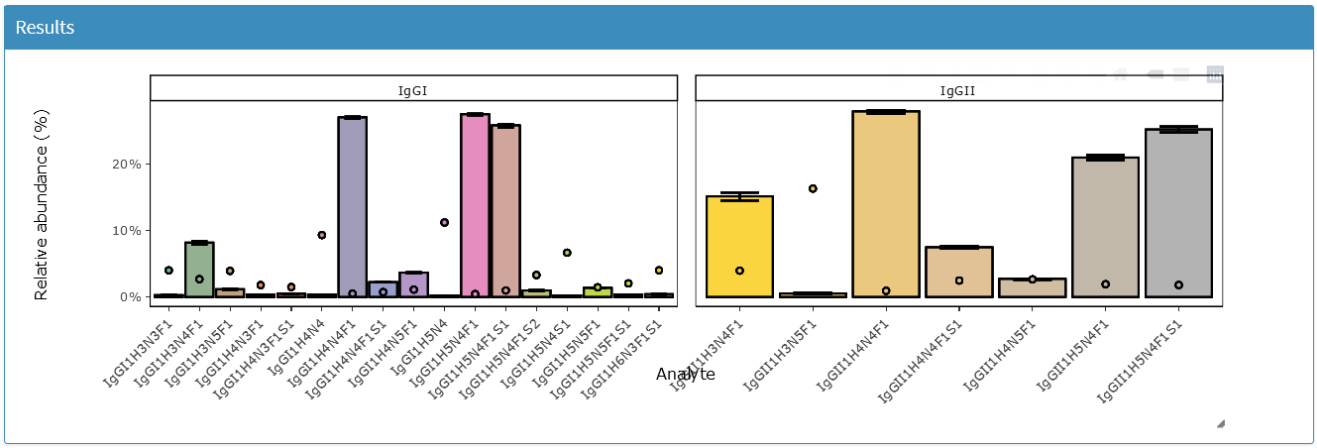
trait	formula
GLT_fucosylation	GLT1H4N5F1S1 + GLT1H5N4F1S1 + GLT1H5N4F1S2 + GLT1H5N5F1S1 + GLT1H5N5F1S2
GLT_bisection	GLT1H5N5F1S1 + GLT1H5N5F1S2
GLT_sialylation	0.5 * (GLT1H4N5F1S1 + GLT1H5N4F1S1 + GLT1H5N4S1 + GLT1H5N5F1S1) + GLT1H5N4F1S2 + GLT1H5N4S2 + GLT1H5N5F1S2

Showing 1 to 3 of 3 entries

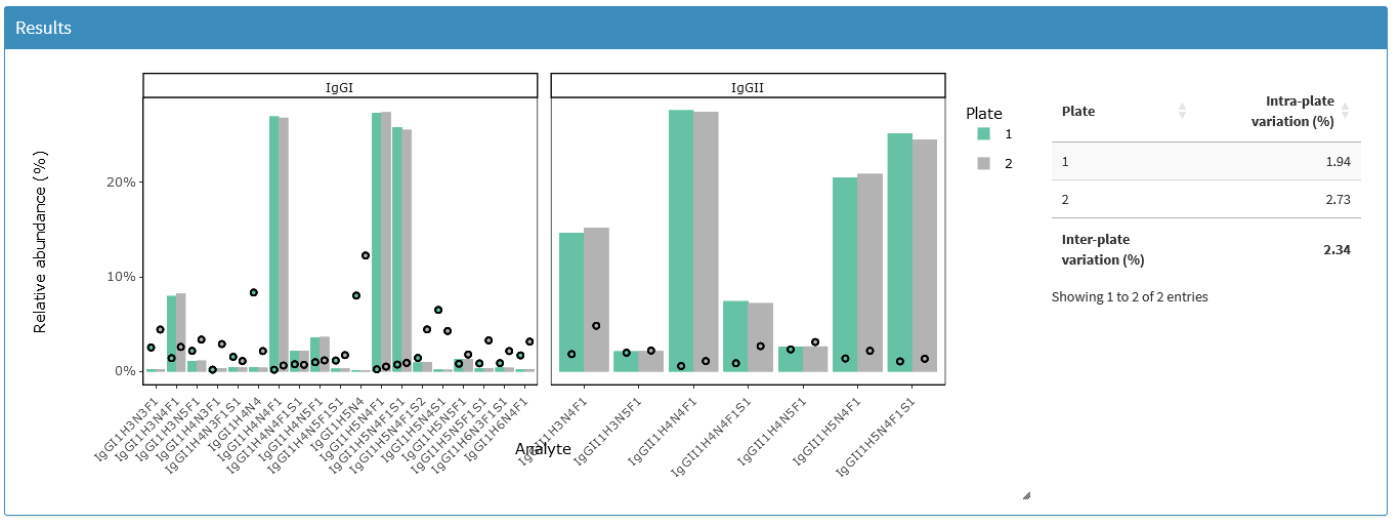
Repeatability (optional)

For sample ID's that appear more than once in the dataset (e.g. "pool"), the repeatability of analyte relative abundances can be visualized. Repeatability of multiple sample ID's you can be assessed by creating separate tabs using the "Add a tab +" button in the top-right corner of the box.

1. Select the sample ID for which repeatability should be visualized.
2. Optional: toggle the switch to assess repeatability per plate.
3. Click "Assess repeatability". An interactive bar plot is generated displaying the relative abundances of all analytes. Error bars on each bar represent the (mean ± standard deviation) for the corresponding analyte. A dot over each bar indicates the relative standard deviation (RSD = SD/mean × 100%).



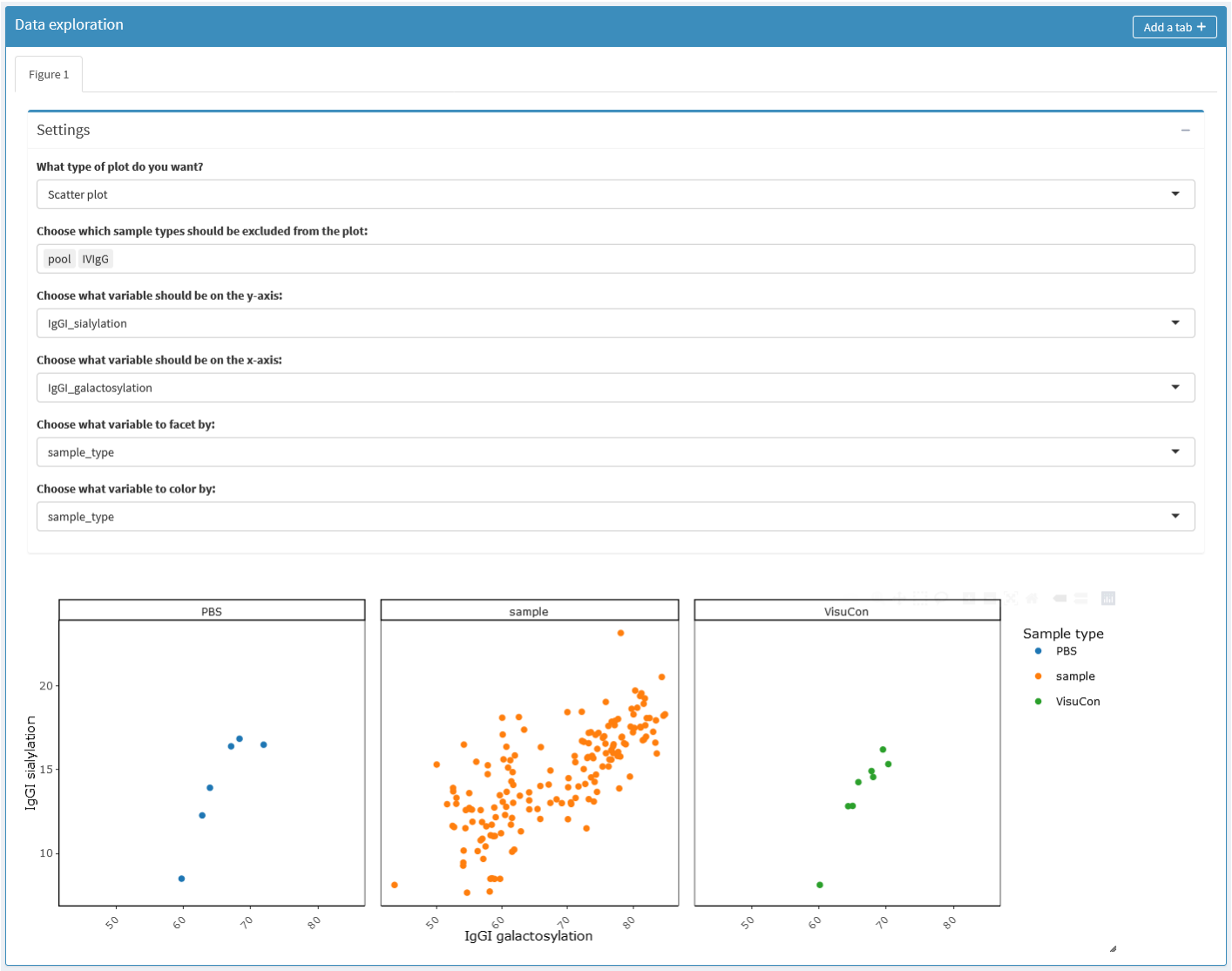
- When assessing repeatability per plate, a table is presented, showing median intra-plate variations calculated as the median of the analyte RSDs on a plate. At the bottom of the table, the inter-plate variation is displayed, calculated as the mean of the intra-plate variations.



Data exploration (optional)

The data exploration tab offers various options for simple data visualization. To create multiple plots, click the “Add a tab +” button in the top-right corner of the box.

1. Choose the type of plot to create: boxplot, scatter plot, or histogram.
2. Optional: select sample types to exclude from the plot.
3. Select the variable to be plotted on the x-axis.
4. Select the variable to be plotted on the y-axis (not applicable for histograms).
5. Optional: choose a variable to facet by.
6. Optional: choose a variable to color by.



Data export

In the data export tab you can download the processed data as an Excel file or as an R object. The processed data includes:

- Sample names, sample types and sample ID's.
- Plate wells, if applicable.
- Metadata, if applicable.
- Sum intensities of all glycosylation sites.
- Relative abundances of all analytes.
- Calculated glycosylation traits, if applicable.
- IgG1 quantities, if applicable.

Additionally, you can download a data processing report in HTML format that can be opened in any browser. This report contains the GlycoDash version that was used, the names of all uploaded files, all choices that were made during each processing step in GlycoDash, and all data visualizations. The generated plots retain their interactivity in this HTML document.

Appendix 1: R packages

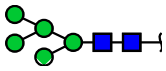
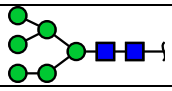
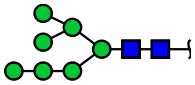
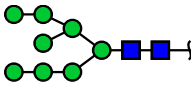
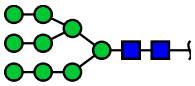
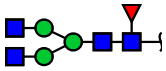
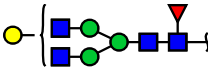
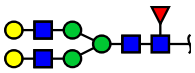
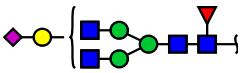
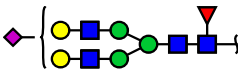
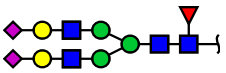
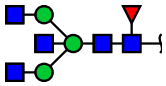
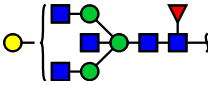
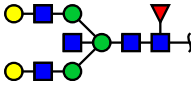
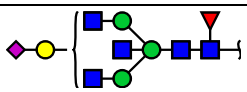
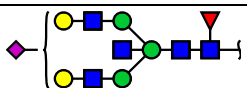
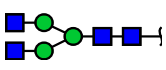
The following R packages are currently used in GlycoDash:

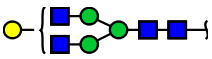
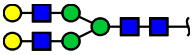
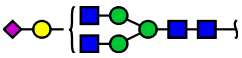
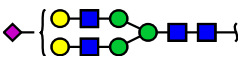
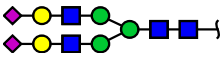
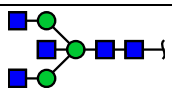
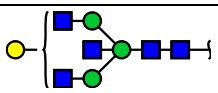
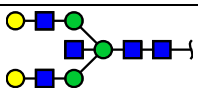
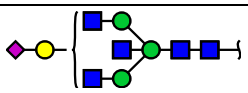
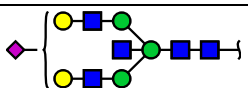
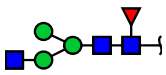
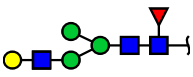
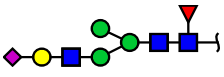
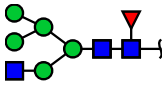
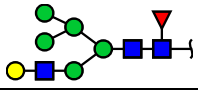
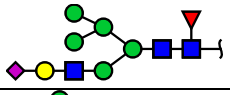
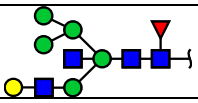
- bsplus (version 0.1.4)
- config (version 0.3.1)
- dashboardthemes (version 1.1.6)
- DT (version 0.27)
- dplyr (version 1.1.0)
- ggplot2 (version 3.4.1)
- gluer (version 1.6.2)
- golem (version 0.4.0)
- htmltools (version 0.5.2)
- kableExtra (version 1.3.4)
- knitr (version 1.42)
- magrittr (version 2.0.3)
- plater (version 1.0.4)
- plotly (version 4.10.1)
- processx (version 3.8.0)
- purrr (version 1.0.1)
- readxl (version 1.4.2)
- rlang (version 1.1.0)
- RLumShiny (version 0.2.3)
- rmarkdown (version 2.20)
- RColorBrewer (version 1.1-3)
- shiny (version 1.7.4)
- shinyalert (version 3.0.0)
- shinyBS (version 0.61.1)
- shinybusy (version 0.3.1)
- shinycssloaders (version 1.0.0)
- shinydashboard (version 0.7.2)
- shinydashboardPlus (version 2.0.3)
- shinyFeedback (version 0.4.0)
- shinyjqui (version 0.4.1)
- shinyjs (version 2.1.0)
- shinyWidgets (version 0.7.6)
- snakecase (version 0.11.0)
- spelling (version 2.2)
- testthat (version 3.1.7)
- tidyr (version 1.3.0)
- tidyselect (version 1.2.0)
- writexl (version 1.4.2)

Appendix 2: Glycan compositions for automatically calculating glycosylation traits.

Below are lists of glycan compositions, and their assumed structures, that can be used for automatically calculating glycosylation traits.

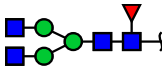
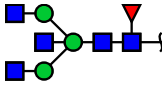
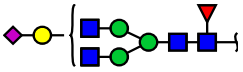
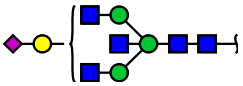
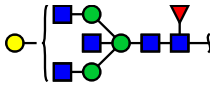
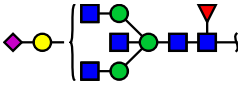
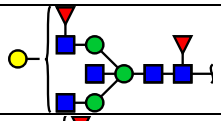
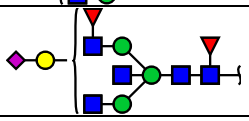
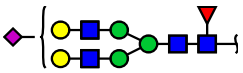
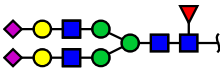
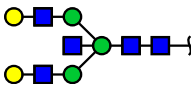
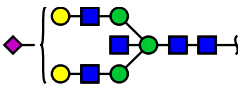
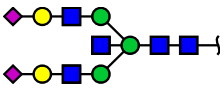
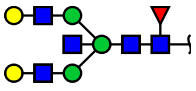
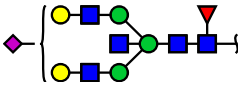
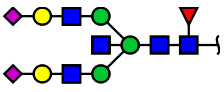
Human IgG

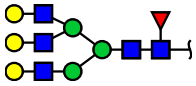
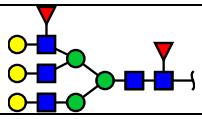
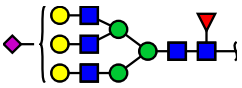
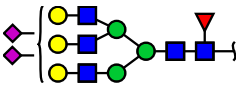
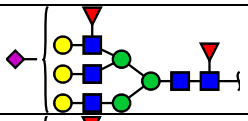
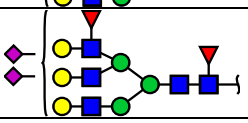
H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H3N4F1	
H4N4F1	
H5N4F1	
H4N4F1S1	
H5N4F1S1	
H5N4F1S2	
H3N5F1	
H4N5F1	
H5N5F1	
H4N5F1S1	
H5N5F1S1	
H3N4	

H4N4	
H5N4	
H4N4S1	
H5N4S1	
H5N4S2	
H3N5	
H4N5	
H5N5	
H4N5S1	
H5N5S1	
H3N3F1	
H4N3F1	
H4N3F1S1	
H5N3F1	
H6N3F1	
H6N3F1S1	
H6N4F1	

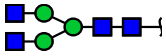
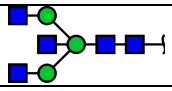
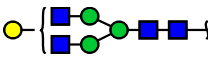
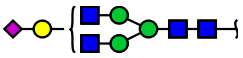
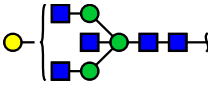
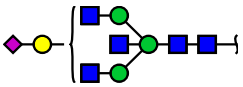
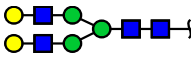
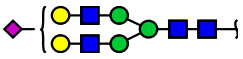
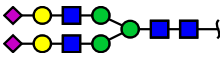
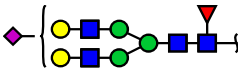
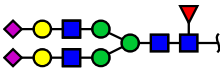
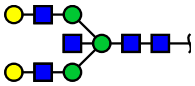
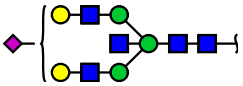
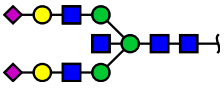
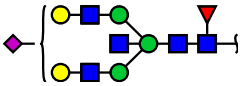
Human IgA

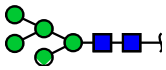
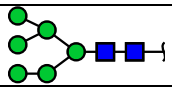
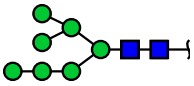
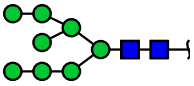
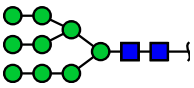
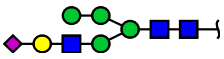
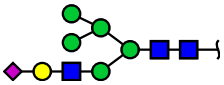
IgA2 N47

H3N4F1	
H3N5F1	
H4N4F1S1	
H4N5S1	
H4N5F1	
H4N5F1S1	
H4N5F2	
H4N5F2S1	
H5N4F1S1	
H5N4F1S2	
H5N5	
H5N5S1	
H5N5S2	
H5N5F1	
H5N5F1S1	
H5N5F1S2	

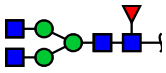
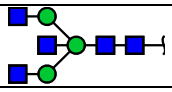
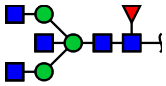
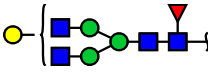
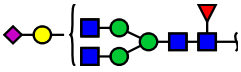
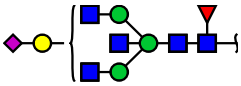
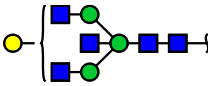
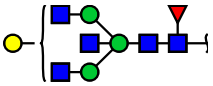
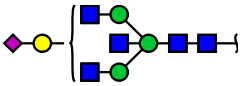
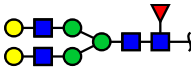
H5N6F1	
H5N6F2	
H5N6F1S1	
H5N6F1S2	
H5N6F2S1	
H5N6F2S2	

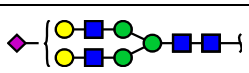
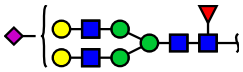
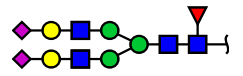
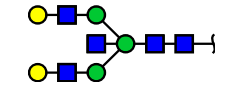
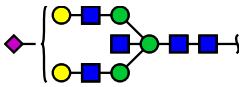
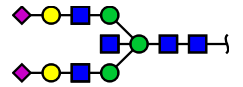
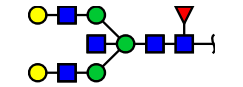
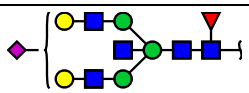
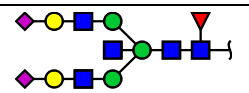
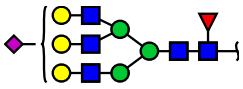
IgA1/2 N144/131

H3N4	
H3N5	
H4N4	
H4N4S1	
H4N5	
H4N5S1	
H5N4	
H5N4S1	
H5N4S2	
H5N4F1S1	
H5N4F1S2	
H5N5	
H5N5S1	
H5N5S2	
H5N5F1S1	

H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H5N3S1	
H6N3S1	
H4N3S1	

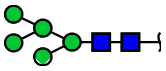
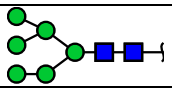
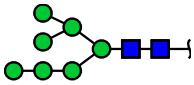
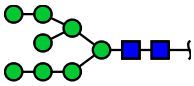
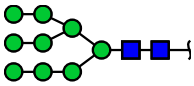
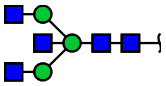
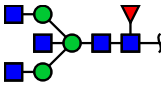
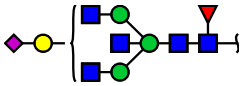
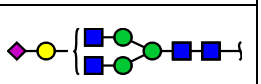
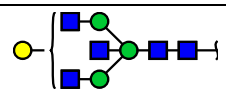
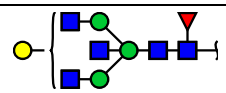
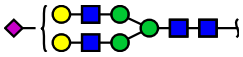
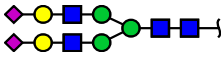
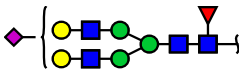
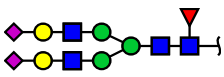
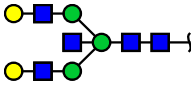
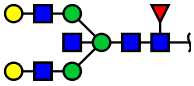
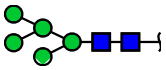
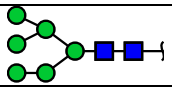
IgA2 N205

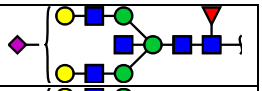
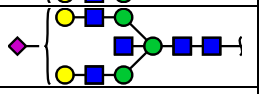
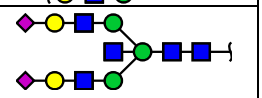
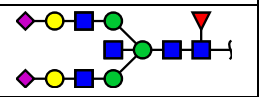
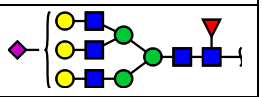
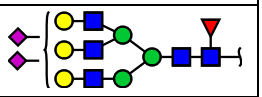
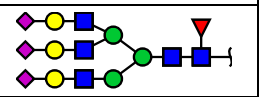
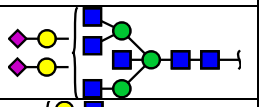
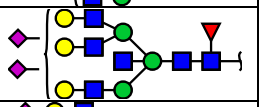
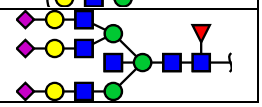
H3N4F1	
H3N5	
H3N5F1	
H4N4F1	
H4N4F1S1	
H4N5F1S1	
H4N5	
H4N5F1	
H4N5S1	
H5N4F1	

H5N4S1	
H5N4F1S1	
H5N4F1S2	
H5N5	
H5N5S1	
H5N5S2	
H5N5F1	
H5N5F1S1	
H5N5F1S2	
H6N5F1S1	

IgA1/2 N340/327

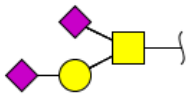
The glycans below may carry one sulphate group ("Su1" at the end of the composition).

H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H3N5	
H3N5F1	
H4N5F1S1	
H4N4S1	
H4N5	
H4N5F1	
H5N4S1	
H5N4S2	
H5N4F1S1	
H5N4F1S2	
H5N5	
H5N5F1	
H5N2	
H6N2	

H5N5F1S1	
H5N5S1	
H5N5S2	
H5N5F1S2	
H6N5F1S1	
H6N5F1S2	
H6N5F1S3	
H5N6S2	
H6N6F1S2	
H6N6F1S3	

IgA1 *O*-glycans

This IgA1 glycopeptide can carry up to six *O*-glycans. Each *O*-glycan has a core GalNAc, which can carry one galactose. In the compositions below, each "N" is therefore a separate *O*-glycan. Sialic acids are assumed to be bonded to a galactose. The exception to this are compositions where there are more sialic acids than galactoses. In that case, any extra sialic acids are assumed to be bonded to the core GalNAc.

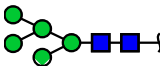
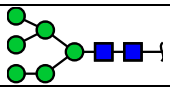
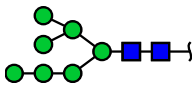
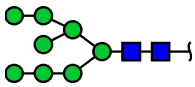
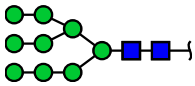
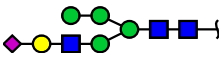
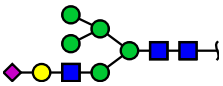
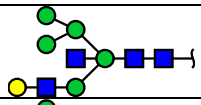
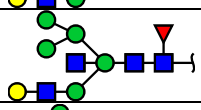
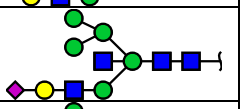
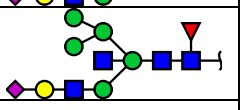
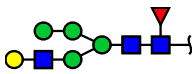
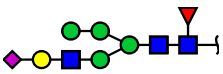
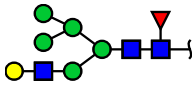
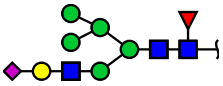
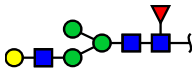
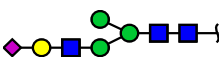


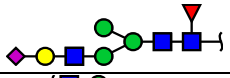
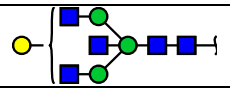
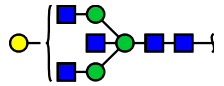
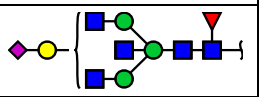
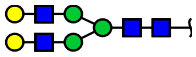
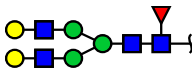
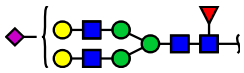
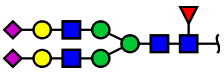
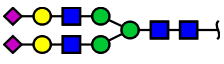
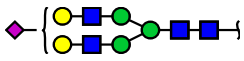
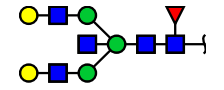
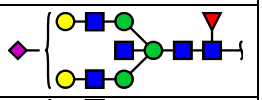
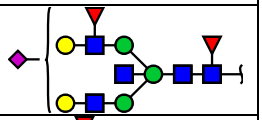
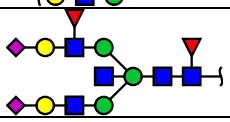
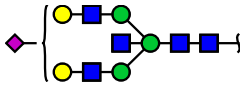
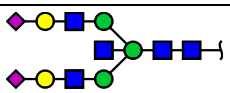
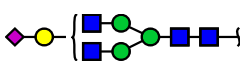
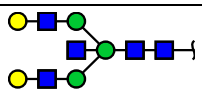
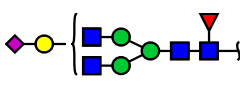
Supported compositions for automatic traits:

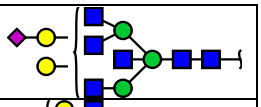
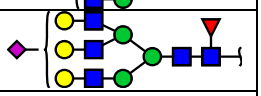
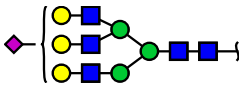
H2N3S2	H4N4S5
H2N4S2	H4N5
H2N5S1	H4N5S1
H3N3S1	H4N5S2
H3N3S2	H4N5S3
H3N3S3	H4N5S4
H3N3S4	H4N6S2
H3N4	H4N6S3
H3N4S1	H4N6S4
H3N4S2	H5N5
H3N4S3	H5N5S1
H3N4S4	H5N5S2
H3N5	H5N5S3
H3N5S1	H5N5S4
H3N5S2	H5N6S3
H3N5S3	H5N6S4
H3N6	H4N6
H3N6S1	H2N5S2
H3N6S2	H5N5S5
H3N6S3	H3N5S4
H4N4	H4N5S5
H4N4S1	H5N6
H4N4S2	H5N6S2
H4N4S3	H4N4S5
H4N4S4	H4N5

Human IgM

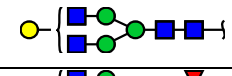
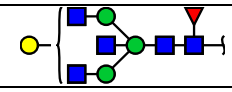
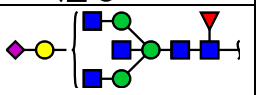
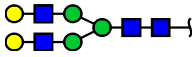
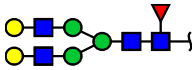
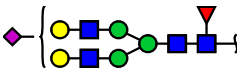
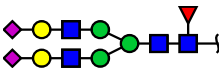
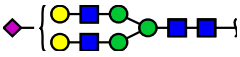
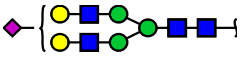
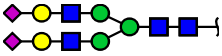
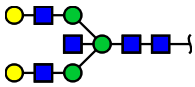
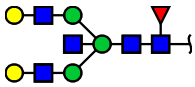
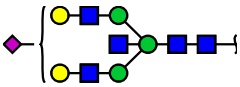
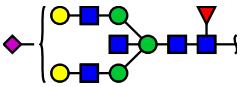
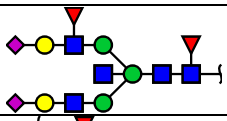
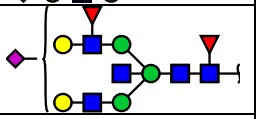
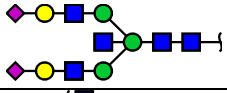
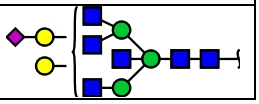
N46

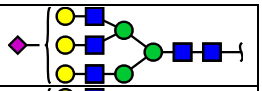
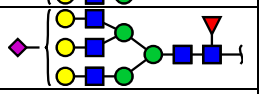
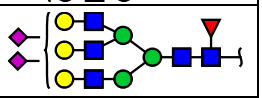
H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H5N3S1	
H6N3S1	
H6N4	
H6N4F1	
H6N4S1	
H6N4F1S1	
H6N4S2	Bisected hybrid
H7N4S2	Bisected hybrid
H5N3F1	
H5N3F1S1	
H6N3F1	
H6N3F1S1	
H4N3F1	
H4N3S1	

H4N3F1S1	
H4N5	
H4N5F1	
H4N5F1S1	
H5N4	
H5N4F1	
H5N4F1S1	
H5N4F1S2	
H5N4S2	
H5N4S1	
H5N5F1	
H5N5F1S1	
H5N5F2S1	
H5N5F1S2	
H5N5S1	
H5N5S2	
H4N4S1	
H5N5	
H4N4F1S1	

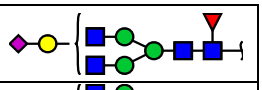
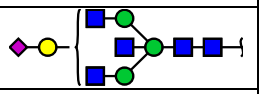
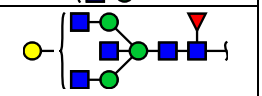
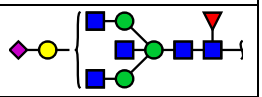
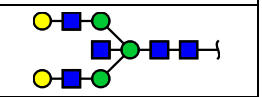
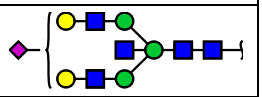
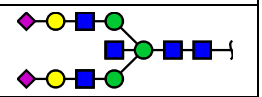
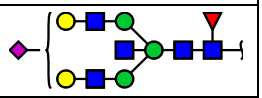
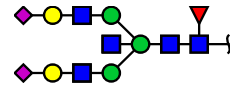
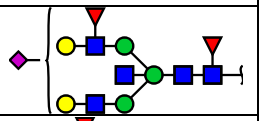
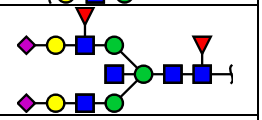
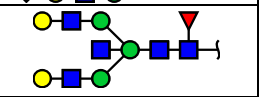
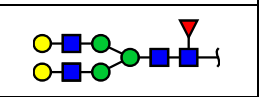
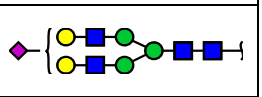
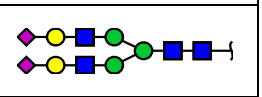
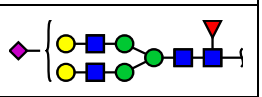
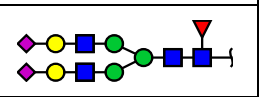
H5N6S1	
H6N5F1S1	
H6N5S1	

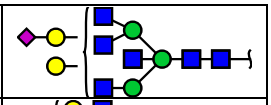
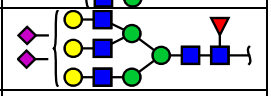
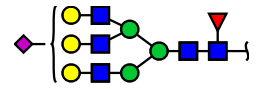
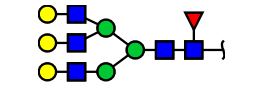
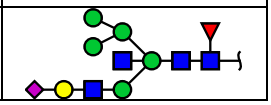
N209

H4N4	
H4N5F1	
H4N5F1S1	
H5N4	
H5N4F1	
H5N4F1S1	
H5N4F1S2	
H5N4F2S2	
H5N4S1	
H5N4S2	
H5N5	
H5N5F1	
H5N5S1	
H5N5F1S1	
H5N5F1S2	
H5N5F2S2	
H5N5S2	
H5N6S1	

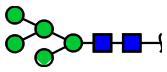
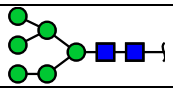
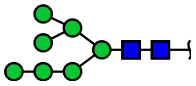
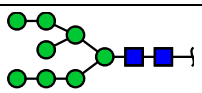
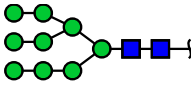
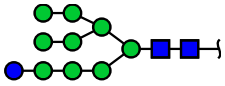
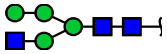
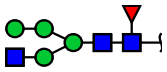
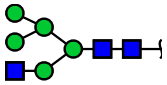
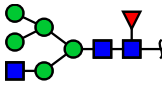
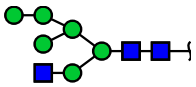
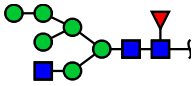
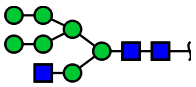
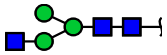
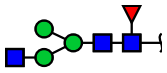
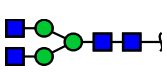
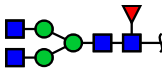
H6N5S1	
H6N5F1S1	
H6N5F1S2	

N272

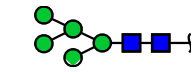
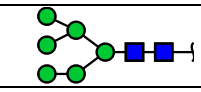
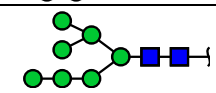
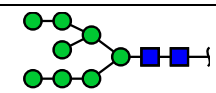
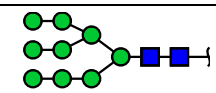
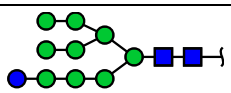
H4N4F1S1	
H4N5S1	
H4N5F1	
H4N5F1S1	
H5N5	
H5N5S1	
H5N5S2	
H5N5F1S1	
H5N5F1S2	
H5N5F2S1	
H5N5F2S2	
H5N5F1	
H5N4F1	
H5N4S1	
H5N4S2	
H5N4F1S1	
H5N4F1S2	

H5N6S1	
H6N5F1S2	
H6N5F1S1	
H6N5F1	
H6N4F1S1	

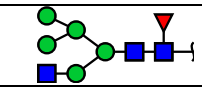
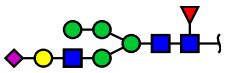
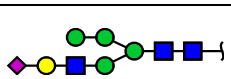
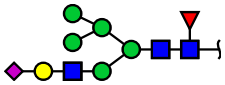
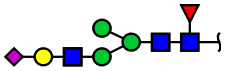
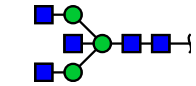
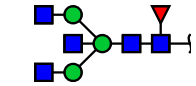
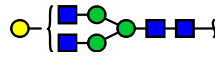
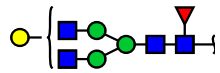
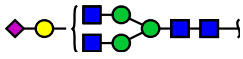
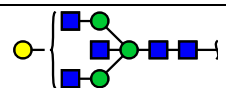
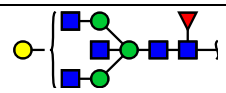
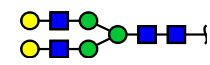
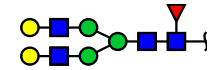
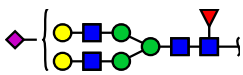
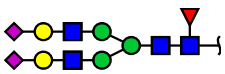
N279

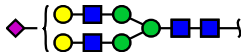
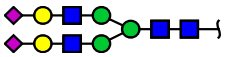
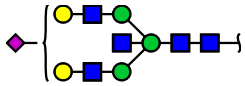
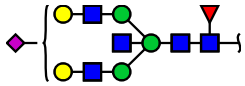
H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H10N2	
H4N3	
H4N3F1	
H5N3	
H5N3F1	
H6N3	
H6N3F1	
H7N3	
H3N3	
H3N3F1	
H3N4	
H3N4F1	

N440

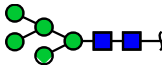
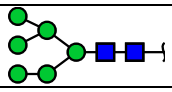
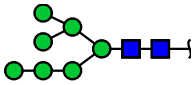
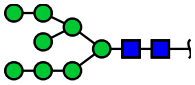
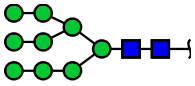
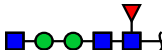
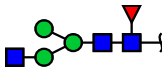
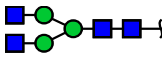
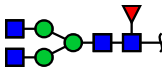
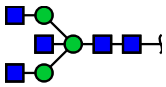
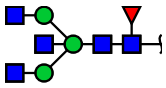
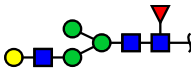
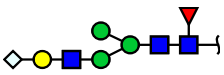
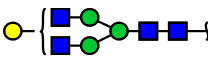
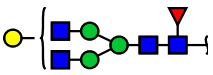
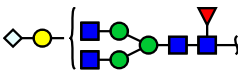
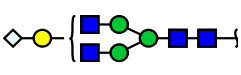
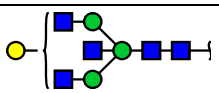
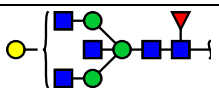
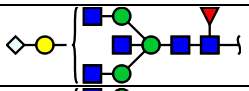
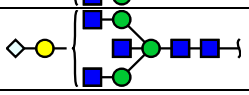
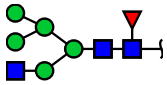
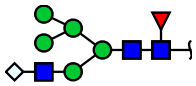
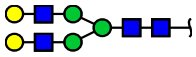
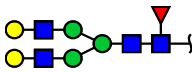
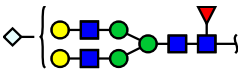
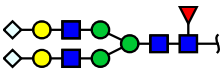
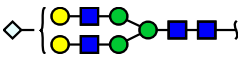
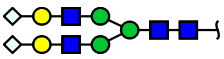
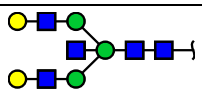
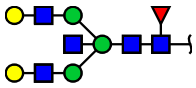
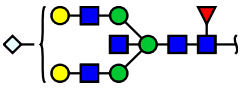
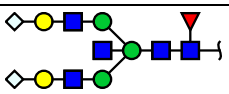
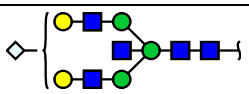
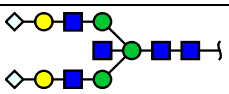
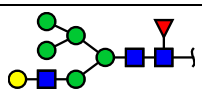
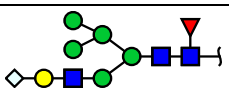
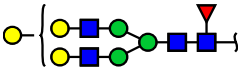
H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H10N2	

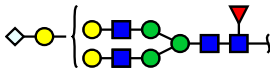
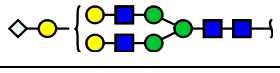
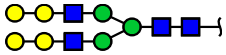
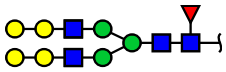
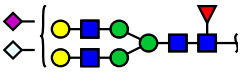
Human Joining Chain (JC)

H5N3F1	
H5N3F1S1	
H5N3S1	
H6N3F1S1	
H4N3S1	
H4N3F1S1	
H3N5	
H3N5F1	
H4N4	
H4N4F1	
H4N4S1	
H4N5	
H4N5F1	
H5N4	
H5N4F1	
H5N4F1S1	
H5N4F1S2	

H5N4S1	
H5N4S2	
H5N5S1	
H5N5F1S1	

Mouse IgG

H5N2	
H6N2	
H7N2	
H8N2	
H9N2	
H2N3F1	
H3N3F1	
H3N4	
H3N4F1	
H3N5	
H3N5F1	
H4N3F1	
H4N3F1G1	
H4N4	
H4N4F1	
H4N4F1G1	
H4N4G1	
H4N5	
H4N5F1	
H4N5F1G1	
H4N5G1	
H5N3F1	
H5N3F1G1	
H5N4	
H5N4F1	
H5N4F1G1	
H5N4F1G2	
H5N4G1	
H5N4G2	
H5N5	
H5N5F1	
H5N5F1G1	
H5N5F1G2	
H5N5G1	
H5N5G2	
H6N3F1	
H6N3F1G1	
H6N4F1	

H6N4F1G1	
H6N4G1	
H7N4	
H7N4F1	
H5N4F1G1S1	

References

[1] D. Merkel, "Docker: lightweight linux containers for consistent development and deployment," *Linux j,* vol. 239, no. 2, p. 2, 2014.

[2] J. Allaire, "RStudio: integrated development environment for R," *Boston, MA,* vol. 770, no. 394, pp. 165-171, 2012.

[3] B. D. Ripley, "The R project in statistical computing," *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network,* vol. 1, no. 1, pp. 23-25, 2001.