Aug 12, 2024

Heron Data Suite: biomedical quantitative analysis applications for Skyline results files

DOI
dx.doi.org/10.17504/protocols.io.8epv5xwj5g1b/v1

Mary Cunningham[1], Stephen Cunningham[2], Matthew Renfrow[1]

[1]Department of Biochemistry and Molecular Genetics, The University of Alabama at Birmingham;
[2]Independent researcher

Mary Cunningham: Graduate Student;
Stephen Cunningham: Software Developer
Matthew Renfrow: Professor

Mary Cunningham
The University of Alabama at Birmingham

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.8epv5xwj5g1b/v1

External link: https://www.herondata.app/

Protocol Citation: Mary Cunningham, Stephen Cunningham, Matthew Renfrow 2024. Heron Data Suite: biomedical quantitative analysis applications for Skyline results files. **protocols.io https://dx.doi.org/10.17504/protocols.io.8epv5xwj5g1b/v1**

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol and it's working.

Created: March 07, 2024

Last Modified: August 12, 2024

Protocol Integer ID: 96329

Keywords: Proteomics, Skyline, Skyline MS, mass spectrometry, biomedical, absolute quantification, AQUA peptides, targeted quantification, calibration curve, peptide ratio results, light:heavy ratio, Heron Data

Funders Acknowledgement:
Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship (Parent F31)
Grant ID: F31DK127833

Abstract

This protocol describes the use of the *Heron Data Suite* application (**https://www.herondata.app/**). The *Heron Data Suite* simplifies the analysis of biomedical proteomics results processed in Skyline. Post-data processing, *Heron Data* organizes results and goes the extra mile to account for technical replicates. *Heron Data* contains two apps currently: Replicate Absolute Quantification (*Heron Quant*) and Glycopeptide Area Under the Curve (*Heron Glyco*).
The purpose of *Heron Quant* is to further analyze 'Peptide Ratio Quantification' .csv exports from the open source proteomics tool *Skyline* (https://skyline.ms/project/home/software/Skyline/begin.view). *Heron Quant* improves the rigor and reproducibility of Skyline for biomedical proteomics purposes by averaging replicates of a single sample or standard and plots the averaged calibration curves and outputting concentration averages. *Heron Quant* will calculate the R-squared value and linear equations of the calibration curves, as well as the coefficient of variance and the standards deviation of replicates.
Heron Glyco is designed specifically for analysis of glycopeptide area under the curve CSV files from Skyline. *Heron Glyco* will organize your Skyline AUC data by glycopeptide into separate worksheets, calculate the percent relative abundance of glycopeptide for each replicate, and plot the percent relative abundance in pie charts of all grouped glycopeptides in each sample.
Heron Data Suite will receive future updates and new application additions.
Need support? Contact us using the form on our website: https://www.herondata.app/contact

Materials

- Heron *Quant*: 'Peptide Ratio Results' .csv Excel file exported from Skyline's Document Grid after absolute quantification analysis of mass spectrometry results (Step 1 illustrates it's export).
- Heron *Glyco*: 'AUC' .csv Excel file exported from Skyline's Document Grid after area under the curve data processing of mass spectrometry results (Step 1 illustrates it's export).

Safety warnings

Need support? Contact us using the form on our website: https://www.herondata.app/contact

Before start

Before using *Heron Data Suite*, all mass spec files should have been processed in *Skyline* by quantification using their absolute quantification features (*Heron Quant*) or for glycopeptide area under the curve (*Heron Glyco*). This will give you the 'Peptide Ratio Results' .csv export file you need to submit to *Heron Data* for further analysis. Step 1 of this protocol will instruct you where to find this document in Skyline. A helpful tutorial for absolute quantification of proteins in *Skyline* can be found on their website in this PDF:
https://skyline.ms/_webdav/home/software/Skyline/@files/tutorials/AbsoluteQuant-20_1.pdf
Visit this page on *Skyline's* website for more information and webinars on absolute quantification:
https://skyline.ms/wiki/home/software/Skyline/page.view?name=tutorial_absolute_quant .

ASSUMPTIONS

Assumptions using Heron Data: Replicate Absolute Quantification (*Heron Quant*)

1. 'Peptide Ratio Results' .csv export file comes from *Skyline* and has been quantified using their tutorial for Absolute Quantification
2. Mass spec file nomenclature should use the following template to allow proper parsing of replicates: date_samplename_numberofreplicate Ex. 20240312_sampleX_1 , 20240312_sampleX_2 , 20240312_sampleY_1 , 20240312_sampleY_2 , 20240312_STD1_1 , 20240312_STD1_2. If your nomenclature was not setup in this format, go into the .csv file and manually edit the sample names to fit this template (Excel find and replace feature will streamline file name editing). File name formatting is how *Heron Quant* groups your technical replicates for a single sample.
3. Technical replicates are ran sequentially as shown above

Assumptions using Heron Data: Glycopeptide AUC (*Heron Glyco*)

1. 'AUC' .csv export file comes from *Skyline.* Data has been completely processed (i.e. peak boundaries have been verified)
2. Mass spec file nomenclature should use the following template to allow proper parsing of replicates: date_samplename_numberofreplicate Ex. 20240312_sampleX_1 , 20240312_sampleX_2 , 20240312_sampleY_1 , 20240312_sampleY_2 , 20240312_STD1_1 , 20240312_STD1_2. If your nomenclature was not setup in this format, go into the .csv file and manually edit the sample names to fit this template (Excel find and replace feature will streamline file name editing). File name formatting is how *Heron Glyco* groups your technical replicates for a single sample.
3. When naming the glycopeptides in Skyline analysis, please use this nomenclature for "Protein Name": GalNAc#Gal#
4. If including sialic acid in analysis, use this nomenclature: SA#GalNAc#Gal#

1. Using *Heron Data: Replicate Absolute Quantification* with your Skyline 'Peptide Ratio Results" .csv export

1 If you're wanting replicate absolute quantification, stay here. If you want glycopeptide area under the curve, proceed to section 2.

2 Please read the *Replicate Absolute Quantification Assumptions* under "Guidelines" of this protocol before beginning to ensure your document's compatibility with *Heron Quant*.

3 To get the "Peptide Ratio Results" .csv Excel file needed for *Heron Quant*, open your analyzed and quantified Skyline document.

4 Click "View" and click "Document Grid"

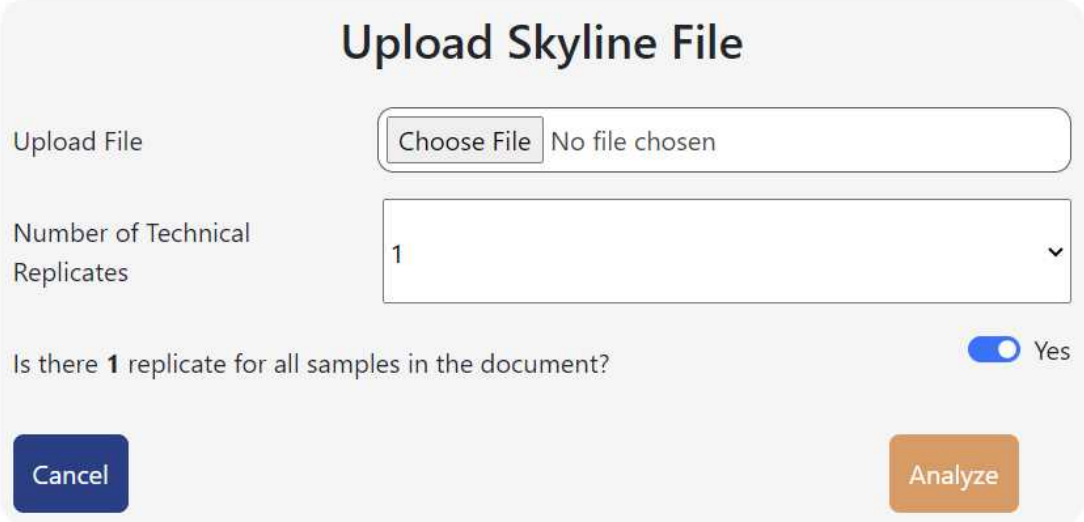
5 In the top left corner of the Document Grid pop-up, click the "Reports" drop-down menu. Select "Peptide Ratio Results"

6 In the middle of the Document Grid table, select "Export..". Name your file and where you would like to save it in the file browser.

7 Once you have your "Peptide Ratio Results" .csv Excel export file from Skyline, visit the app's website: **https://www.herondata.app/**

8 Click the "Heron Quant" button.

9 Click "Choose File" and upload your .csv Excel file.



10 Use the pop-up file browser to locate and select your Excel file. THe name of the file should appear in the window.

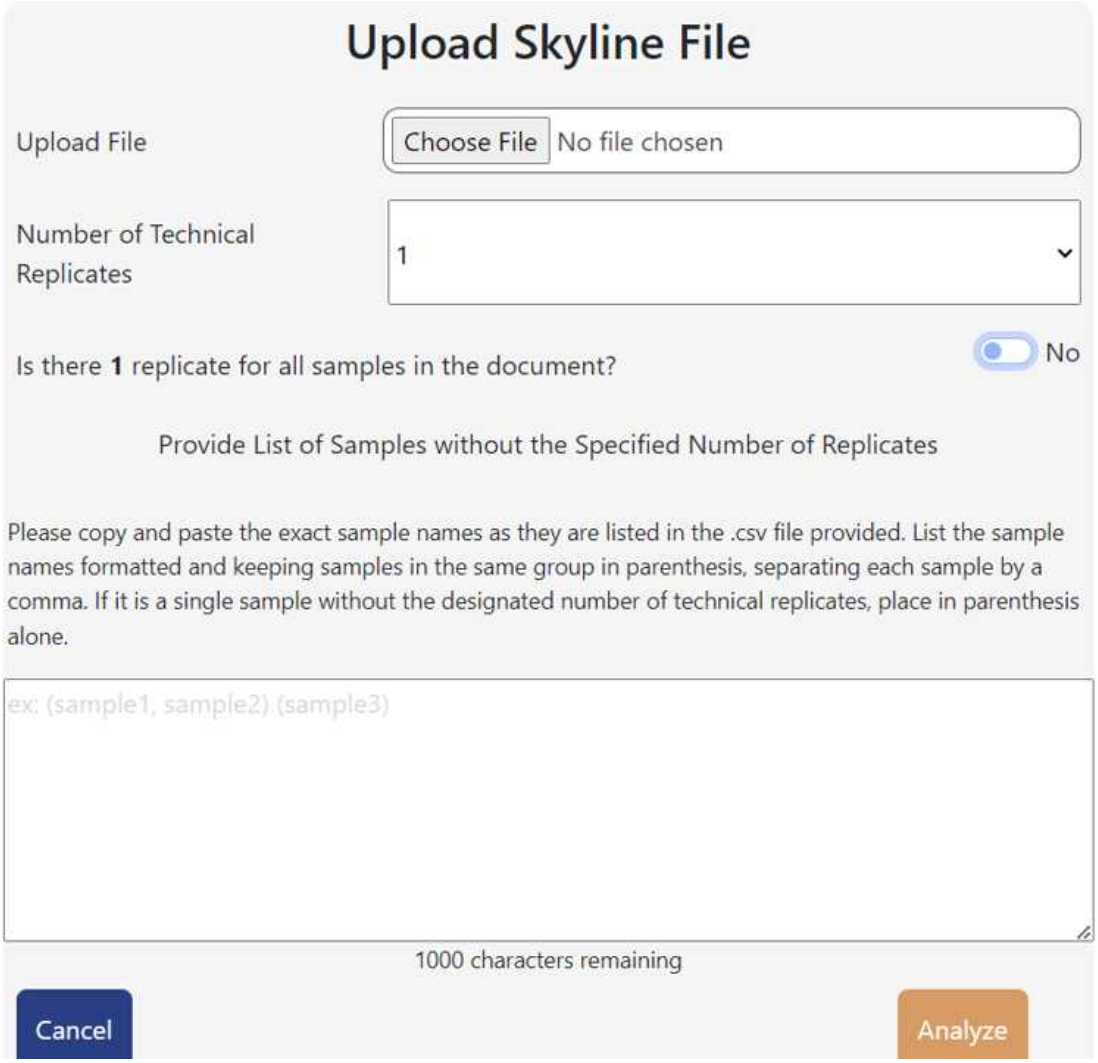
11 Below, use the drop-down menu to select the number of technical replicates each of your samples has.

12 It will then ask if if there are that number of technical replicates for each sample. This is an option in case one sample lost replicates during running either due to running error, bad samples, not wanting to use certain replicates, or not all of the samples in your run had the same amount of replicates.

12.1 If all samples have the same number of replicates you selected, select "Yes".

12.2 If not all samples in your data set have the same amount of technical replicates, select no. A dialogue box will appear. Provide a list of samples without the specified number of replicates.

- Please copy and paste the exact sample names as they are listed in the .csv file provided. This is how the technical replicates will be grouped together. List the sample names formatted keeping replicates in same group in parenthesis, separating each sample in the group by a comma. If it is a single replicate, place in parenthesis alone.
- Ex. (sample1, sample2, sample3, sample4) (sampleA) (sampleB) (sampleC, sampleD)



13 Click "Analyze"

14 The analyzed data page will appear. Table 1 is the data for averaging replicates to generate calibration curves. You can choose how many rows you would like to display per page and toggle through pages at the bottom of the table.

Table 1. Data Averages

Filter Columns...

Peptide Name	Replicate Name	Peptide Pea...	Peptide Retenti...	Protein	Quantification	Quantification Replicate Avg	Ratio To Standard	Peptide Result Ratio Avg
TPLTATLSK	20240205_Orbi_IG_STD1Mix_1	0.4	25.65	Peptide 1	86.4313 fmol		0.0513	
TPLTATLSK	20240205_Orbi_IG_STD1Mix_2	0.4	25.67	Peptide 1	92.0972 fmol		0.0532	
TPLTATLSK	20240205_Orbi_IG_STD1Mix_3	0.4	25.54	Peptide 1	86.0579 fmol	88.19546666666668	0.0512	0.0519

This table contains:

14.1 Peptide Name, Replicate Name, Peptide Peak Ratio, Peptide Retention Time, Protein, Quantification, and Ratio To Standard as were exactly listed in the .csv file analyzed

14.2 In addition, *Heron Quant* has now grouped the replicates for each sample and averaged both the quantification (blue highlighted column named "Quantification Replicate Avg) and the ratio to standard (blue highlighted column "Ratio Replicate Avg")

15 Scroll below Table 1 to see Table 2. Table 2 is statistical analysis of your averaged replicates. You can choose how many rows you would like to display per page and toggle through pages at the bottom of the table. You'll see the quantification average and peptide ratio averages from Table 1.

Table 2. Data Statistics

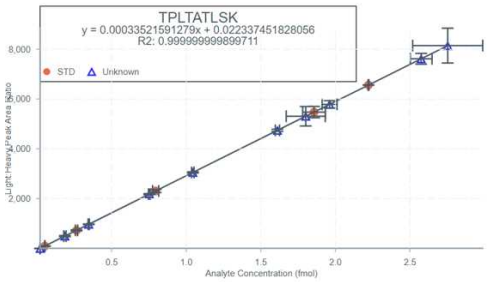
Peptide	Replicate Name	Quantification Average (X axis)	Quantification ST...	Quantification CV	Ratio Average (X axis)	Ratio STDEV	Ratio CV
TPLTATLSK	20240205_Orbi_IG_STD1Mix_3	88.19546666666668	3.3841541252332674	0.038371066599416304	0.0519	0.001126942766958...	0.0217137334674 07...
TPLTATLSK	20240205_Orbi_IG_STD2Mix_3	721.0948	22.163672013229235	0.030736141785004185	0.26406666666666667	0.007439310003846...	0.028172090395783...
TPLTATLSK	20240205_Orbi_IG_STD3Mix_3	2304.5940333333333	61.96351945010331	0.026886956467764574	0.7948666666666666	0.020778915595702...	0.026141385048690...

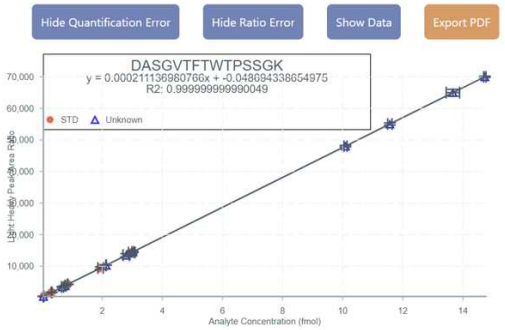
The values *Heron Quant* calculates in addition includes:

15.1
- The standard deviation (STDEV) and coefficient of variance (CV) of the quantification averages
- The STDEV and CV of the peptide ratio results averages

16 Export Table 2 as a CSV by clicking the "Export CSV" button.

17 Below Table 2 are all calibration curves for each peptide in the .csv with the averages plotted of each sample and standard. The X-axis is the Quantification Averages titled Analyte Concentration (units). The Y-axis is the Peptide Ratio Results average titled "Light : Heavy Peak Area Ratio". Standards (STD) are plotted a red dot and "unknown" samples are plotted a blue triangle.





17.1 The linear regression equation and R^2 values are provided at the top of each calibration curve.

17.2 You can click on each data point to see the sample name toggle on and click again to toggle it off.

17.3 Click the "Show Data" button to show the exact x and y coordinates over each data point. Click again to "Hide Data".

18 Click "Export PDF" to export the calibration curves as a PDF file

19 Note: Leaving the tab open will store your results. If you exit the tab or the browser the page will refresh to the home page.

2. Using *Heron Data: Glycopeptide AUC* with Skyline 'AUC' .csv export

20 If you're wanting glycopeptide area under the curve, stay here. If you want replicate absolute quantification, go to section 1.

21 Please read the *Heron Glyco Assumptions* under "Guidelines" of this protocol before beginning.

22 To get the "AUC" .csv Excel file needed for *Heron Glyco*, open your analyzed Skyline document.

23 Click "View" and click "Document Grid"

24 In the top left corner of the Document Grid pop-up, click the "Reports" drop-down menu. Select "AUC".

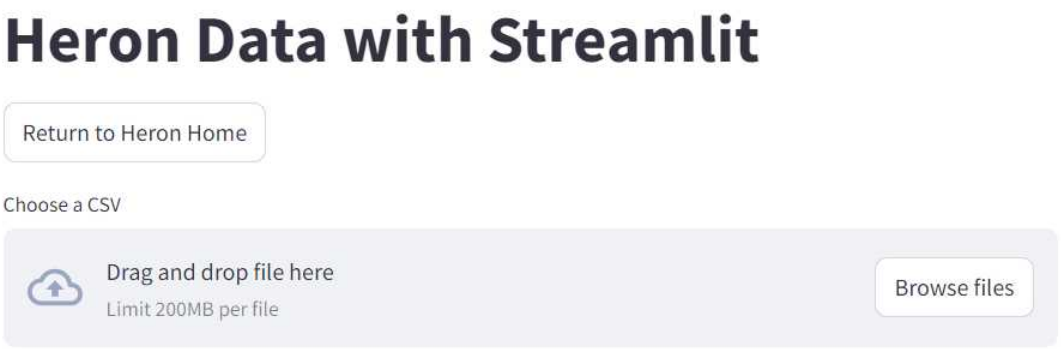
25 You can customize this report to include any additional data you would like. *Heron Glyco* will only consider the following data columns: protein name, replicate name, and total area. NOTE: If you have multiple charge states, *Heron Glyco* recognizes multiple total areas for the same protein (glycopeptide) and same replicate and will add the two charge states together. For more information on Skyline reports, visit Skyline's tutorial on custom reports that defines values in the report document. https://skyline.ms/wiki/home/software/Skyline/page.view?name=tutorial_custom_reports

26 In the middle of the Document Grid table, select "Export..". Name your file and where you would like to save it in the file browser.

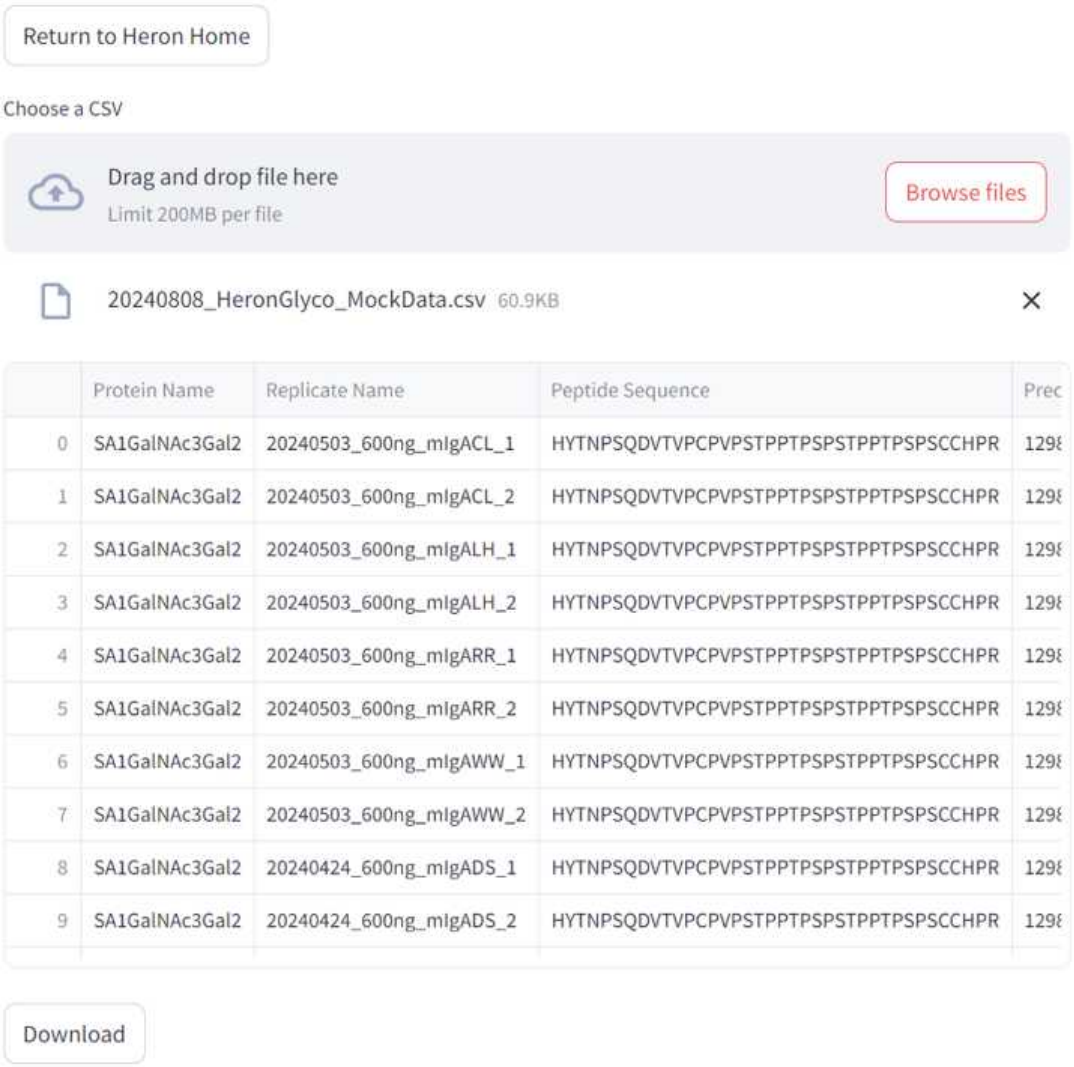
27 Once you have your "AUC" .csv Excel export file from Skyline, visit the app's website: **https://www.herondata.app/**

28 Click the *Glycopeptide Area Under the Curve* button.

29 Click *Browse Files* and select the .csv export from Skyline for your AUC data and click open. The page will then update to show in grid format the data in the file you have selected.



30 After visually confirming this is the correct file, scroll down and click the *download* button. Name the file in the file browser and click save. This will download your results file.



	Protein Name	Replicate Name	Peptide Sequence	Prec
0	SA1GalNAc3Gal2	20240503_600ng_mIgACL_1	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
1	SA1GalNAc3Gal2	20240503_600ng_mIgACL_2	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
2	SA1GalNAc3Gal2	20240503_600ng_mIgALH_1	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
3	SA1GalNAc3Gal2	20240503_600ng_mIgALH_2	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
4	SA1GalNAc3Gal2	20240503_600ng_mIgARR_1	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
5	SA1GalNAc3Gal2	20240503_600ng_mIgARR_2	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
6	SA1GalNAc3Gal2	20240503_600ng_mIgAWW_1	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
7	SA1GalNAc3Gal2	20240503_600ng_mIgAWW_2	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
8	SA1GalNAc3Gal2	20240424_600ng_mIgADS_1	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129
9	SA1GalNAc3Gal2	20240424_600ng_mIgADS_2	HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR	129

31 Open the download.
Note: *Heron Glyco* has now made the file type saved as an Excel Workbook file.

32 About the download: Your data will be organized by glycopeptide. Each glycopeptide will have it's own tabbed worksheet and will have the areas (and/or averaged areas if multiple charge states) listed horizontally by replicate.

Glycopeptides organized into respective Excel worksheets

	A	B	C
1	**Peptide**	**Replicate 1**	**Replicate 2**
2	**GalNAc5Gal3**	316727	460418
3	**SA1GalNAc5Gal3**	1075169	1278588
4	**SA2GalNAc5Gal3**	3277112	3508001
5	**SA3GalNAc5Gal3**	1841557	1868765
6	**SA4GalNAc5Gal3**	45798	68225
7	**Summed Total Area**	6556363	7183997
8			

33 Below the Total Areas, you'll see a section titled "Percent Relative Abundance". *Heron Glyco* has calculated the percent relative abundance by the area of the single glycopeptide divided by the total of the area of all glycopeptides in that worksheet multiplied by 100. This is done per replicate.

Percent Relative Abundance (area/total area for peaks considered * 100)

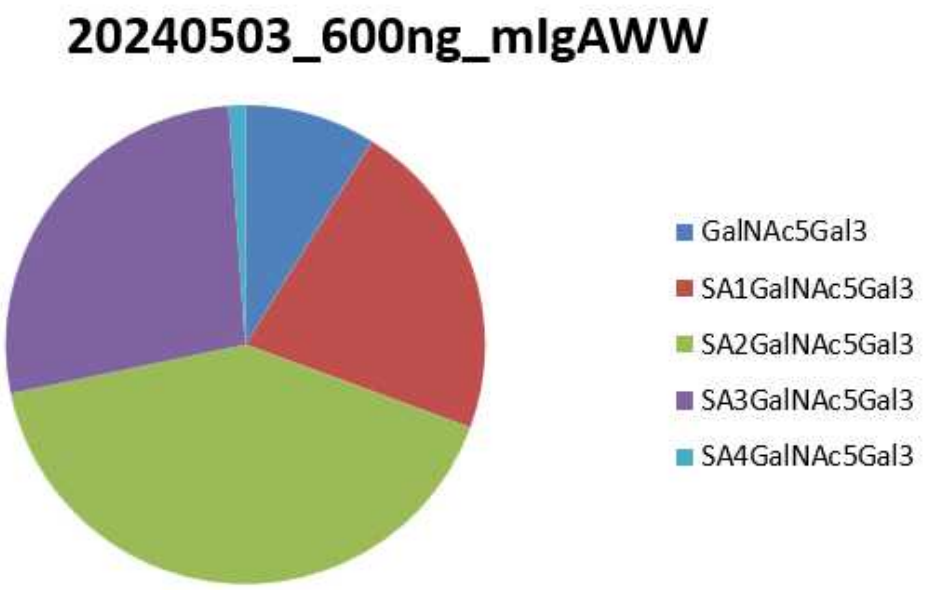
Peptide	**Replicate 1**	**Replicate 2**
GalNAc5Gal3	4.830833802	6.408939202
SA1GalNAc5Gal3	16.39886321	17.79772458
SA2GalNAc5Gal3	49.9836876	48.83076928
SA3GalNAc5Gal3	28.08808786	26.0128867
SA4GalNAc5Gal3	0.698527522	0.949680241

34 The next section below is the Replicate Average Percent Relative Abundance. Here, *Heron Glyco* has averaged any replicates' percent relative abundance together. Replicates are determined by nomenclature specified in the *Assumptions* under G*uidelines* of this protocol. Please see that for more information on setting up replicate names. If no replicates are denoted, it will use the percent relative abundance calculated above.

Replicate Average % Relative Abundance

Peptide	Averaged Replicate 1 and 2
GalNAc5Gal3	5.619886502
SA1GalNAc5Gal3	17.0982939
SA2GalNAc5Gal3	49.40722844
SA3GalNAc5Gal3	27.05048728
SA4GalNAc5Gal3	0.824103881

35 Each worksheet contains a pie chart, per sample using the value calculated in the Replicate Average Percent Relative Abundance. Each pie chart title is the name of the replicate it represents. The pie chart is showing what percentage of the whole sample that each glycopeptide within that worksheet represents.



Protocol references

Heron Data Suite GitHub page: https://github.com/Step-henC/heron-data-internet

Schilling, B., Rardin, M. J., MacLean, B. X., Zawadzka, A. M., Frewen, B. E., Cusack, M. P., Sorensen, D. J., Bereman, M. S., Jing, E., Wu, C. C., Verdin, E., Kahn, C. R., Maccoss, M. J., & Gibson, B. W. (2012). Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Molecular & cellular proteomics : MCP*, *11*(5), 202–214. https://doi.org/10.1074/mcp.M112.017707

Pino, L. K., Searle, B. C., Bollinger, J. G., Nunn, B., MacLean, B., & MacCoss, M. J. (2020). The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass spectrometry reviews*, *39*(3), 229–244. https://doi.org/10.1002/mas.21540

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics (Oxford, England)*, *26*(7), 966–968. https://doi.org/10.1093/bioinformatics/btq054

Absolute Quantification Skyline Tutorial PDF:
https://skyline.ms/_webdav/home/software/Skyline/@files/tutorials/AbsoluteQuant-20_1.pdf

Skyline's tutorial on Custom Reports
https://skyline.ms/wiki/home/software/Skyline/page.view?name=tutorial_custom_reports