



GlypNirO: An automated workflow for quantitative *N*- and *O*-linked glycoproteomic data analysis

Toan K. Phung^{‡1}, Cassandra L. Pegg^{‡1} and Benjamin L. Schulz^{*1,2,§}

Full Research Paper

Open Access

Address:

¹School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, 4072, Australia and ²ARC Training Centre for Biopharmaceutical Innovation, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, QLD 4072, Australia

Email:

Benjamin L. Schulz* - b.schulz@uq.edu.au

* Corresponding author ‡ Equal contributors

§ Telephone: +61 7 336 54875

Keywords:

glycoproteomics; mass spectrometry; *N*-glycosylation; *O*-glycosylation; Python

Beilstein J. Org. Chem. **2020**, *16*, 2127–2135.

<https://doi.org/10.3762/bjoc.16.180>

Received: 16 June 2020

Accepted: 20 August 2020

Published: 01 September 2020

This article is part of the thematic issue "GlycoBioinformatics".

Guest Editors: N. H. Packer and F. Lisacek

© 2020 Phung et al.; licensee Beilstein-Institut.

License and terms: see end of document.

Abstract

Mass spectrometry glycoproteomics is rapidly maturing, allowing unprecedented insights into the diversity and functions of protein glycosylation. However, quantitative glycoproteomics remains challenging. We developed GlypNirO, an automated software pipeline which integrates the complementary outputs of Byonic and Proteome Discoverer to allow high-throughput automated quantitative glycoproteomic data analysis. The output of GlypNirO is clearly structured, allowing manual interrogation, and is also appropriate for input into diverse statistical workflows. We used GlypNirO to analyse a published plasma glycoproteome dataset and identified changes in site-specific *N*- and *O*-glycosylation occupancy and structure associated with hepatocellular carcinoma as putative biomarkers of disease.

Introduction

Glycosylation is a key post-translational modification critical for protein folding and function in eukaryotes [1-3]. Diverse types of glycosylation are known, all involving modification of specific amino acid residues with complex carbohydrate structures. *N*-Linked glycosylation of asparagines and *O*-linked glycosylation of serines and threonines are the most widely encountered and well studied in eukaryotes. A key feature of

glycosylation critical to its biological functions and important for its analysis is its high degree of heterogeneity [4]. This heterogeneity can take the form of variable occupancy, also known as macroheterogeneity – the presence or absence of modification at a particular site in a protein, due to inefficient transfer of the initial glycan structure [5]. In addition, the non-template-driven synthesis of glycan structures means that there

can be multiple different glycan structures attached at the same site in a pool of mature glycoproteins [6]. This structural heterogeneity is also known as microheterogeneity. This heterogeneity in glycan structure and occupancy can be influenced by many genetic and environmental factors. As such, protein glycosylation is often regulated in response to physiological or pathological conditions [7]. Accurately profiling the site-specific occupancy and structural heterogeneity of glycosylation across glycoproteomes can therefore provide insight into the biology of healthy and diseased states [8].

The current state-of-the-art technology for the characterisation, identification, and quantification of the glycome or glycoproteome is liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) [9]. Popular and powerful glycoproteomic workflows typically involve standard proteomic sample preparation and protease digestion, coupled with depletion of abundant proteins or enrichment of glycopeptides to enable their measurement. There have also been several advances in glycopeptide quantification strategies including chemical labelling, label-free and data-independent acquisition methods [10]. Progress in MS technology in particular has enabled deep and sensitive measurement of highly complex glycoproteomes, generating large amounts of high quality data [11]. With that comes the need for robust and automated workflows for extracting meaningful results. Numerous software packages have been developed for analysis of outputs from MS technology to automate the process of transformation of raw MS data into ion intensities and matching them with appropriate glycan and peptide sequence databases for glycopeptide identification (reviewed in [12–16]). However, there are few efficient, robust, and automated workflows for glycopeptide quantification. There are several freely available software programs for quantitative label-free glycoproteomics using MS1 or data-dependent acquisition. These include LaCyTools [17], MassyTools [18], and GlycoSpectrumScan [19], which use a predefined list of analytes and masses to interrogate MS1 data, and I-GPA [20], GlycopeptideGraphMS [21], GlycoFragwork [22], and GlycReSoft [23], which integrate identification and abundance/intensity information for glycopeptides (a recent review is provided in [10]). Importantly, the complexity of glycan heterogeneity requires that downstream analysis often involves manual processing in addition to standard informatics workflows.

Here, we developed and used GlypNirO, an automated bioinformatic workflow for label-free quantitative *N*- and *O*-glycoproteomics that focuses on improving robustness and throughput. Our workflow uses a collection of scripts built on an in-house sequence string handling library and the scientific Python data handling package pandas [24], and integrates outputs of two commonly used software packages in glycoproteomic MS data

analysis, Proteome Discoverer (Thermo Fisher) and Byonic (Protein Metrics), to extract occupancy and glycoform abundance of all identified glycopeptides from LC–MS/MS datasets. We applied the workflow to a published dataset comparing the plasma glycoproteomes of liver cancer patients (hepatocellular carcinoma, HCC) and healthy controls [20]. Our analysis revealed differences in occupancy and glycan compositions of several proteins as potential HCC tumor biomarkers.

Results and Discussion

GlypNirO

Byonic is powerful software that allows identification of glycopeptides and peptides from complex glyco/proteomic LC–MS/MS datasets but does not perform quantification. Proteome Discoverer allows robust and facile measurement of peptide abundances using MS1 peptide area under the curve (AUC) information. We developed GlypNirO to integrate the outputs from Byonic and Proteome Discoverer to improve the efficiency, ease, and robustness of quantitative glycoproteomic data analysis. GlypNirO takes Byonic and Proteome Discoverer output files, and user-defined sample information and processing parameters, performs a series of automated data integration and computational steps, and provides informative and intuitive output files with site or peptide-specific glycoform abundance data. Glyco/peptide identifications from Byonic are linked with AUC data from Proteome Discoverer by matching the experimental scan number. The sites of glycosylation within each peptide assigned by Byonic are identified and clearly labelled. While identification of glycopeptides based on peptide sequence and glycan monosaccharide composition is comparatively reliable with modern LC–MS/MS and data analytics, it is much more difficult to unambiguously assign the precise site of modification within a glycopeptide. GlypNirO therefore provides two options for analysis: site-specific or peptide-specific. If the user trusts Byonic's site-specific assignment, then all peptide variants that contain that site are included in calculations of its occupancy and glycoform distribution. If the user prefers to perform a peptide-specific analysis, then each proteolytically unique peptide form is treated separately. GlypNirO then calculates the occupancy and proportion of each glycoform at each site/peptide, and provides output files with the protein name, site and/or peptide information, and occupancy and glycoform abundance. Full details are provided in the Experimental section.

To provide a proof-of-concept use of GlypNirO, we performed an exploratory reanalysis of a previously published dataset [20] obtained from the ProteomeXchange Consortium via the MassIVE repository (PXD003369, MSV000079426). This study performed glycoproteomic LC–MS/MS analysis of whole plasma or plasma depleted of six abundant proteins from liver

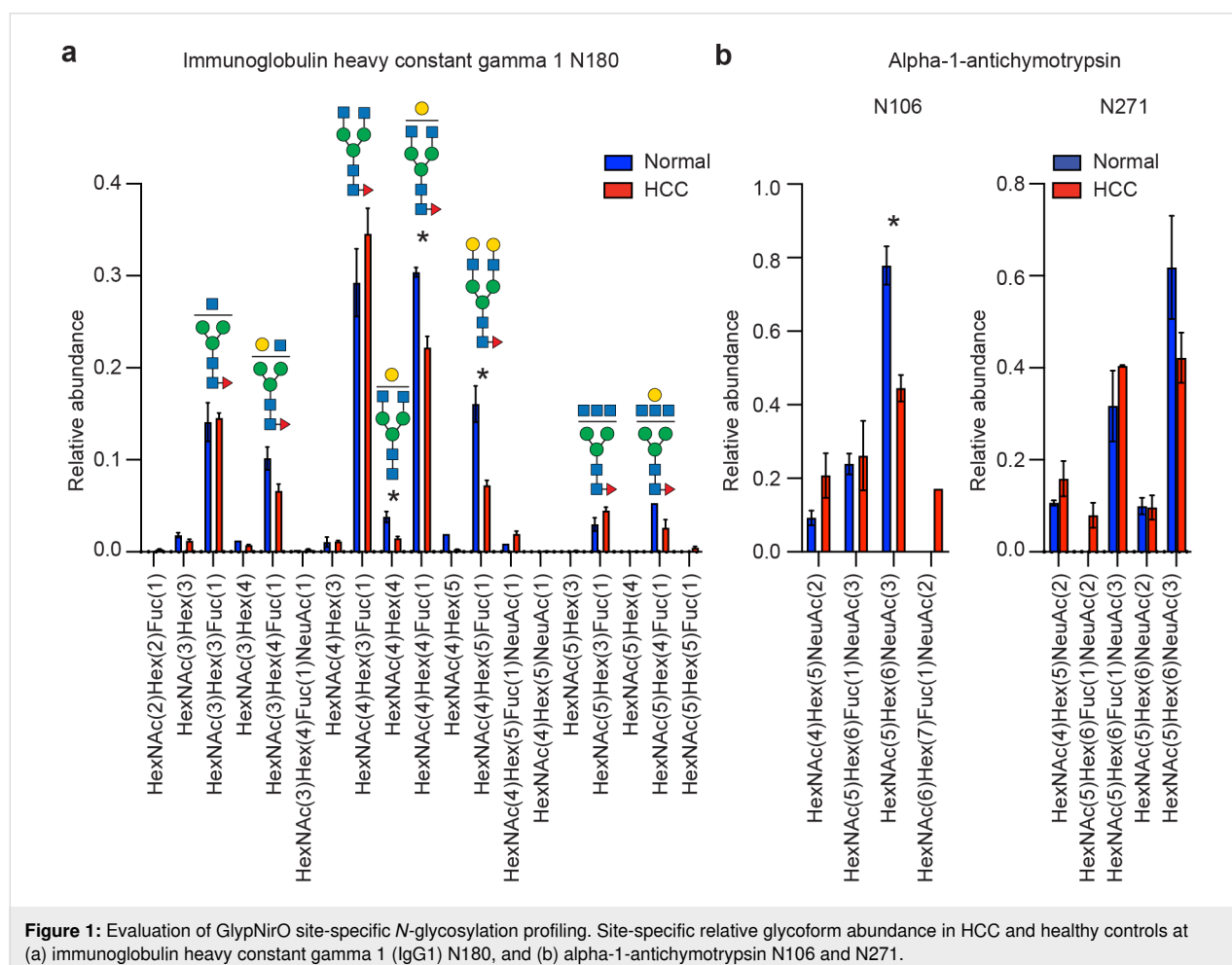
cancer (hepatocellular carcinoma (HCC)) patients and healthy controls. We identified glycopeptides and peptides in the datafiles from these samples using Byonic, searching separately for *O*- and *N*-glycopeptides (Supporting Information File 1, Tables S1–S24), and processed the files with Proteome Discoverer (Supporting Information File 1, Tables S25–S36). We then used GlypNirO to process these results files. This analysis was able to identify and measure 851 *N*-glycopeptides (site-specific) from 150 proteins and 301 *O*-glycopeptides (peptide level) from 89 proteins (Supporting Information File 1, Tables S37–S40).

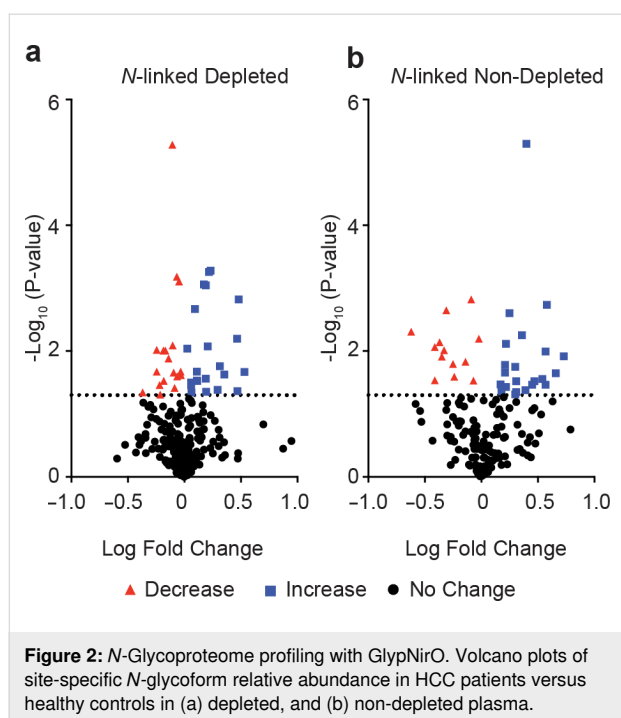
Several changes in site-specific glycosylation associated with HCC had been previously identified [20]. We benchmarked the performance of our workflow using GlypNirO with these previously reported changes. Consistent with previous analysis, we found that agalactosylated *N*-glycans on IgG were increased in abundance in HCC (Figure 1a), and the relative abundance of the HexNAc(5)Hex(6)NeuAc(3) composition at multiple sites on alpha-1-antichymotrypsin was decreased in HCC (Figure 1b).

N-Glycoproteome analysis

To extend our analysis, we next investigated the full suite of *N*-glycosylation sites that we were able to identify and measure with GlypNirO. Comparing the site-specific glycoform relative abundance and occupancy, we identified 111 unique glycopeptides with increased and 128 with decreased abundance in HCC compared with healthy controls in depleted plasma, and 93 increased and 67 decreased in HCC in non-depleted plasma ($P < 0.05$, Figure 2a and 2b). This analysis with GlypNirO of site-specific relative glycoform abundance confirmed that HCC was associated not only with changes in glycoprotein abundance in plasma, but with changes in the proportions of different glycan structures at specific sites in diverse glycoproteins.

Examining the data in more detail identified several sites with multiple glycoforms with statistically significant changes in abundance. Specifically, HCC patients had decreased abundance of disialylated *N*-glycans at alpha-1-antitrypsin N271 and haptoglobin N184 (Figure 3a and 3b), with increased abundance of non-sialylated *N*-glycans at fibrinogen N78





(Figure 3c), and decreased abundance of trisialylated *N*-glycans at alpha-2-HS-glycoprotein N176 (Figure 3d). Together, this suggests an overall decrease in sialylation of *N*-glycans across the plasma glycoproteome in HCC.

O-Glycoproteome analysis

The plasma *O*-glycoproteome is perhaps somewhat neglected [25], despite the importance of *O*-glycosylation to diverse aspects of fundamental biology, health, and disease. We therefore investigated all *O*-glycosylation sites that we were able to identify and measure with GlypNirO. Because there are often multiple potential sites of *O*-glycosylation within a tryptic peptide and site-specific assignment is challenging with CID or HCD fragmentation information, we used peptide-centric analysis of the plasma *O*-glycoproteome. Comparing peptide-specific glycoform relative abundance and occupancy, we identified 41 unique *O*-glycopeptides with increased and 27 with decreased abundance in HCC compared with healthy controls in depleted plasma, and 17 increased and 26 decreased in HCC in non-depleted plasma ($P < 0.05$, Figure 4a and 4b). As the dataset we analysed measured enriched glycopeptides, it is likely that unglycosylated peptide forms are underrepresented.

We could identify both changes in peptide-specific *O*-glycan compositions and in *O*-glycan occupancy. HCC patients had increased glycan occupancy and decreased abundance of monosialylated *O*-glycan on fibrinogen alpha chain G₂₇₂GSTSYGT-GSETESPR (Figure 5a). HCC patients showed a relative

decrease in disialylated and an increase in monosialylated *O*-glycan abundance on both plasma protease C1 inhibitor V₄₅AATVISK and histidine-rich glycoprotein S₂₇₁STTKPPFKPHGSR (Figure 5b and 5c). Together, and consistent with our *N*-glycoproteome analyses, this suggests that HCC is associated with an overall decrease in sialylation of *N*- and *O*-glycans across the plasma glycoproteome.

Conclusion

GlypNirO is an automated software pipeline that integrates glyco/peptide identification from Byonic and quantification from Proteome Discoverer, and provides output that is appropriate for both manual inspection and further statistical analyses. We note that all glycopeptide identification and quantification workflows will include false positive and negative results, and users should ensure data is appropriately searched and curated before processing with GlypNirO. Additionally, modern LC-MS/MS glycoproteomics cannot fully structurally characterise glycans and often struggles to confidently assign the precise sites of modification; ambiguities which may confound quantification workflows. Our proof-of-principle analysis of a plasma glycoproteome dataset demonstrated that GlypNirO can be used to detect changes in site-specific glycosylation occupancy and structure of *N*- and *O*-glycosylation in complex glycoproteomes. Specifically, we found that HCC was associated with decreased sialylation of both *N*- and *O*-glycans at specific sites on selected plasma glycoproteins. GlypNirO will be a useful tool for enabling robust high-throughput quantitative glycoproteomics.

Experimental

Byonic and Proteome Discoverer analysis

We identified glycopeptides and peptides using Byonic (Protein Metrics, v. 3.8.13) searching all DDA files ($n = 12$) downloaded from a previously published dataset [20] obtained from the ProteomeXchange Consortium via the MassIVE repository (PXD003369, MSV000079426). Two searches were conducted on each file, one *N*-linked and one *O*-linked. A human protein database was used (UniProt UP000005640, downloaded April 20, 2018 with 20,303 reviewed proteins) [26]. Cleavage specificity was set as C-terminal to Arg/Lys with a maximum of one missed cleavage. The precursor mass tolerance was 10 ppm and fragment ion mass tolerances for CID and HCD were 0.5 Da and 20 ppm, respectively. Carbamidomethylation of cysteines was set as a fixed modification, and dynamic modifications included deamidation of asparagine, monooxidised methionine, and the formation of pyroglutamate at *N*-terminal glutamic acid and glutamine. All variable modifications were set as “Common 1” allowing each modification to be present once on a peptide. For *N*-linked searches (N-X-S/T) a database of 164 *N*-glycans was used (Supporting Information File 1, Table S41) and for the

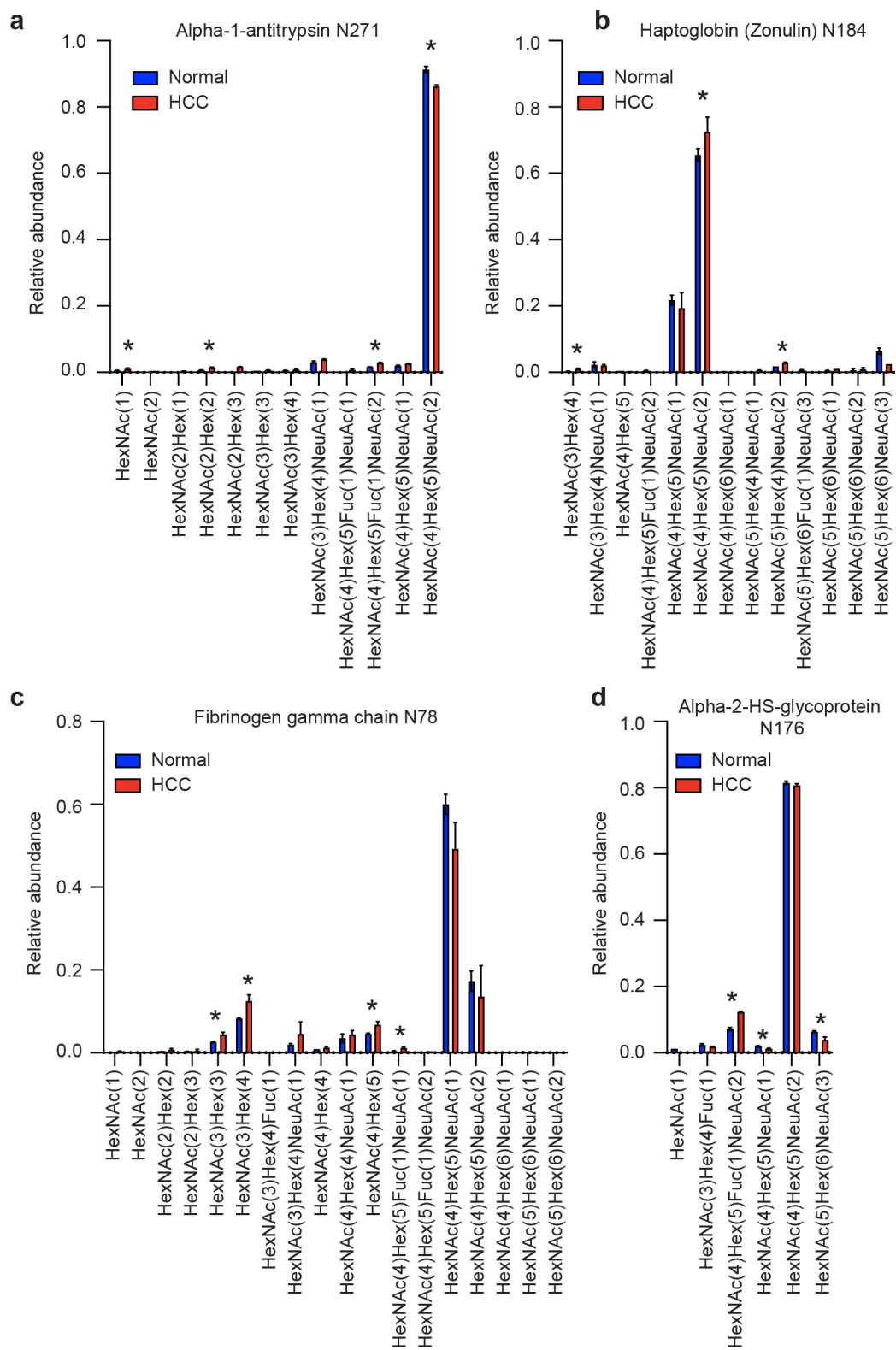
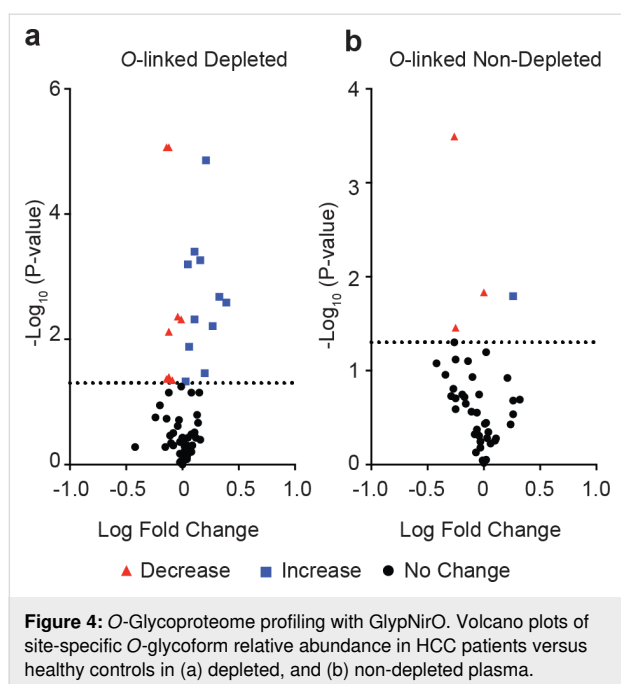


Figure 3: Site-specific N-glycopeptide profiling with GlypNirO. Site-specific relative glycoform abundance in HCC patients and healthy controls at (a) alpha-1-antitrypsin N271, (b) haptoglobin N184, (c) fibrinogen gamma chain N78, and (d) alpha-2-HS-glycoprotein N176. $N = 3$; values show mean; error bars show standard error of the mean; *, $P < 0.05$.



O-linked searches (at any S/T) a database of 49 O-glycans (Supporting Information File 1, Table S42) was used. All glycan modifications were set as “Rare 1” allowing each modification to be present once on a peptide. A maximum of two common modifications and one rare modification were allowed per peptide. A protein false discovery rate cut-off of 1% was applied along with the peptide automatic score cutoff [27]. Precursor peak areas were calculated using the Precursor Ions Area Detector node in Proteome Discoverer (v. 2.0.0.802 Thermo Fisher Scientific). Text output files from Proteome Discoverer and Byonic were then used in GlypNirO (<https://github.com/bschulzlab/glypnirO> and Supporting Information File 3).

Output combination and preprocessing

GlypNirO was built and used in Python 3.8.3 with backward compatibility tested up to Python 3.6. Each Byonic output file was first iteratively prepared for linking with AUC information from the Proteome Discoverer output. Using a regular expression pattern provided by UniProtKB, the UniProtKB accession ID of each protein from the *Protein Name* column of the Byonic output was parsed and saved into a new temporary *master id* column. If a UniProtKB accession ID could not be matched, the entire protein name was saved into the *master id* column. *Reverse* (decoy) sequences and *Common contaminant proteins* were filtered and removed from the dataset.

To combine data from different isoforms of the same protein, the Byonic output was grouped by accession ID in the *master id* column. From the *Scan number* column, the numeric scan number associated with a PSM was extracted into a temporary *Scan*

number column. Area Under the Curve (AUC) information from the *First Scan* column from the Proteome Discoverer output text file was assigned to Byonic data at each corresponding scan number, in the *Area* column. Entries with no AUC value and those not meeting a user-defined Byonic score cutoff (200 here) were removed from the data set.

Using the *Glycans NHFAGNa* and *Modification Type(s)* column, the script obtained the monosaccharide composition of the attached glycan. In the standard Byonic output, only the Δ mass of the modification is directly indicated on the modified peptide sequence, with no direct indication of the identity of the corresponding modification. The script therefore calculated the theoretical mass of the glycan from the *Glycans NHFAGNa* column, and matched this to the corresponding amino acid in the peptide. This allowed the unambiguous assignment of each site of glycosylation from the Byonic output. Options were provided to either include Byonic assignments of site-specificity, or not, in calculation for the final output.

Unique PSM selection and glycoform AUC calculation

The compiled dataset as a whole was sorted based on two levels in descending order, first by *Area* and then by *Score*. Two options were available for glyco/peptide grouping: site-specific analysis, or peptide-specific analysis. For site-specific analysis, the site-specificity of glycosylation assigned by Byonic was trusted, and all peptide variants that contained that site were included in calculations of its occupancy and glycoform distribution. PSMs with identical unmodified peptide sequence, glycan monosaccharide composition, calculated *m/z*, and site of glycosylation were grouped. For each group, the PSM precursor *m/z* value with the highest associated *Area* was selected as the unique PSM. The *Area* associated with each unique PSM was used for the calculation of the total AUC of each glycoform at each identified glycosylation site.

For peptide-specific analysis, the precise site of glycosylation within a peptide as assigned by Byonic was ignored, and each proteolytically unique peptide form was treated separately. PSMs with identical unmodified peptide sequence, glycan monosaccharide composition, and calculated *m/z* were grouped. As with site-specific analysis, for each group, the PSM with the highest *Area* was selected as its unique PSM. The *Area* of each unique PSM was used for the calculation of the total AUC of each glycoform for each unique proteolytic peptide.

Proportional data analysis and final output

In order to allow comparisons of site-specific glycoform abundance and occupancy between different samples, the proportion of each glycoform was calculated with and without inclusion of

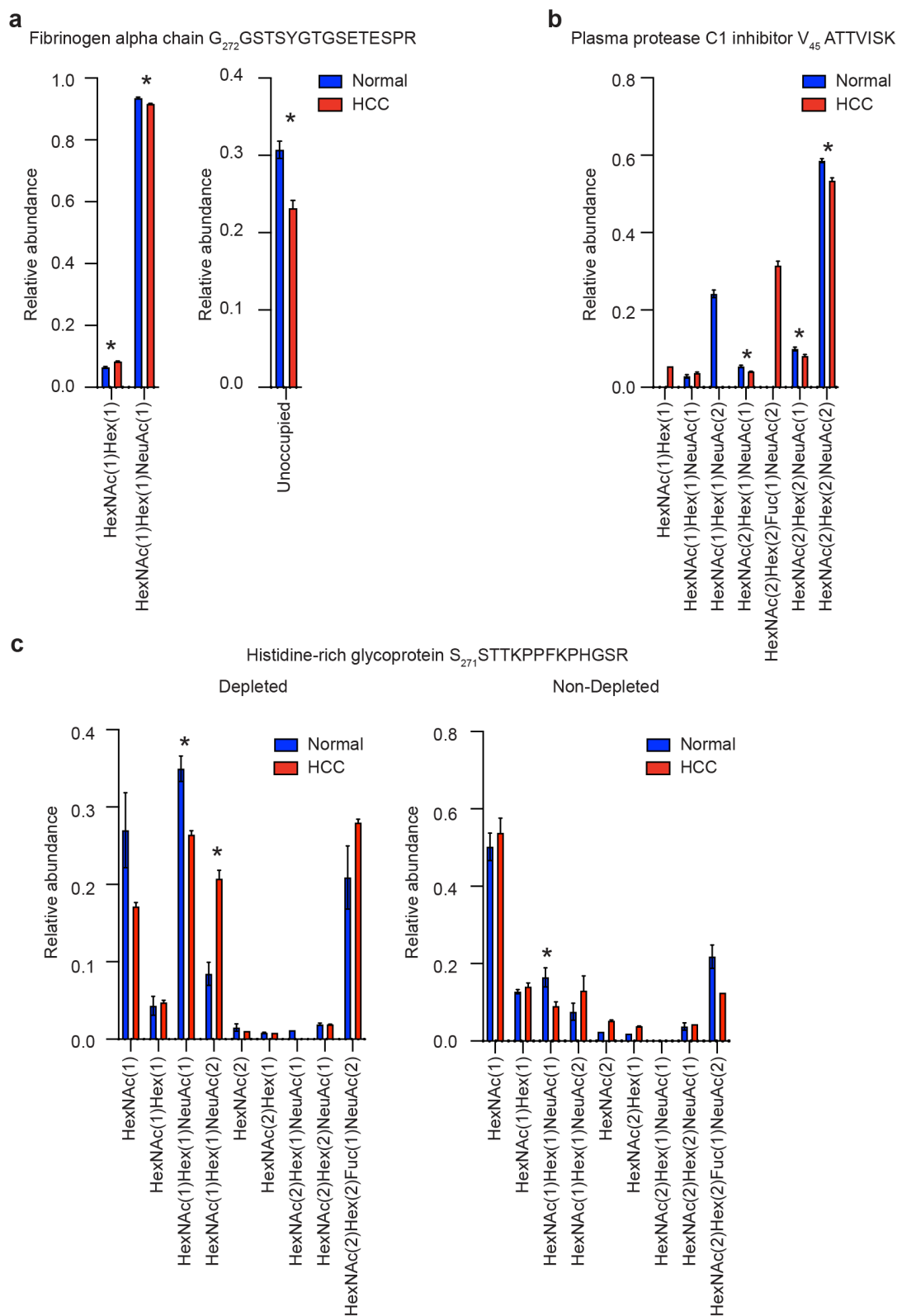


Figure 5: Peptide-specific *O*-glycosylation profiling with GlypNirO. Peptide-specific relative glycoform abundance in HCC patients and healthy controls on (a) fibrinogen alpha chain G₂₇₂GSTSYGTGSETESPR, (b) plasma protease C1 inhibitor V₄₅AATVISK, and (c) histidine-rich glycoprotein S₂₇₁STTKPPFKPHGSR. *N* = 3; values show mean; error bars show standard error of the mean; *, *P* < 0.05.

unglycosylated peptides. For calculation of proportion, glycosylation status was assumed to not quantitatively affect detection. These results were concatenated into the final output file, where columns are the different samples and rows are the different peptide and glycoforms that have been analysed. The protein name of each glycosylated protein detected in the analysis was also included, parsed from the online UniProtKB database using an inhouse Python library.

Statistical analyses

Significant differences in glycoform abundances between healthy and diseased samples were evaluated using an unpaired two-tailed t-test without corrections for multiple comparisons. Missing values were not imputed. Spectra were manually validated for glycoforms of interest.

Supporting Information

Supporting Information File 1

Supplementary Tables S1–S42.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-16-180-S1.zip>]

Supporting Information File 2

GlypNirO workflow overview.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-16-180-S2.pdf>]

Supporting Information File 3

Glypniro-master; automated script for processing and combining Byonic and PD standard output.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-16-180-S3.zip>]

Funding

This work was funded by an Australian Research Council Discovery Project DP160102766 to BLS, an Australian Research Council Industrial Transformation Training Centre IC160100027 to BLS, and a National Health and Medical Research Council Ideas Grant APP1186699 to BLS and CLP.

ORCID® iDs

Toan K. Phung - <https://orcid.org/0000-0002-2964-6070>

Cassandra L. Pegg - <https://orcid.org/0000-0002-6080-7047>

Benjamin L. Schulz - <https://orcid.org/0000-0002-4823-7758>

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: <https://doi.org/10.1101/2020.06.15.153528>

References

- Aebi, M. *Biochim. Biophys. Acta, Mol. Cell Res.* **2013**, *1833*, 2430–2437. doi:10.1016/j.bbamcr.2013.04.001
- Joshi, H. J.; Narimatsu, Y.; Schjoldager, K. T.; Tytgat, H. L. P.; Aebi, M.; Clausen, H.; Halim, A. *Cell* **2018**, *172*, 632–632.e2. doi:10.1016/j.cell.2018.01.016
- Shental-Bechor, D.; Levy, Y. *Curr. Opin. Struct. Biol.* **2009**, *19*, 524–533. doi:10.1016/j.sbi.2009.07.002
- Riley, N. M.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. *Nat. Commun.* **2019**, *10*, 1311. doi:10.1038/s41467-019-09222-w
- Zacchi, L. F.; Schulz, B. L. *Glycoconjugate J.* **2016**, *33*, 359–376. doi:10.1007/s10719-015-9641-3
- Bertozzi, C. R.; Rabuka, D. Structural basis of glycan diversity. In *Essentials of Glycobiology*; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2009.
- Cobb, B. A. *Glycobiology* **2020**, *30*, 202–213. doi:10.1093/glycob/cwz065
- Rabinovich, G. A.; van Kooyk, Y.; Cobb, B. A. *Ann. N. Y. Acad. Sci.* **2012**, *1253*, 1–15. doi:10.1111/j.1749-6632.2012.06492.x
- Wuhrer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2007**, *849*, 115–128. doi:10.1016/j.jchromb.2006.09.041
- Delafield, D. G.; Li, L. *Mol. Cell. Proteomics* **2020**, mcp.R120.002095. doi:10.1074/mcp.r120.002095
- Thaysen-Andersen, M.; Packer, N. H.; Schulz, B. L. *Mol. Cell. Proteomics* **2016**, *15*, 1773–1790. doi:10.1074/mcp.o115.057638
- Hu, H.; Khatri, K.; Klein, J.; Leymarie, N.; Zaia, J. *Glycoconjugate J.* **2016**, *33*, 285–296. doi:10.1007/s10719-015-9633-3
- Cao, W.; Liu, M.; Kong, S.; Wu, M.; Zhang, Y.; Yang, P. *Mol. Cell. Proteomics* **2020**, mcp.R120.002090. doi:10.1074/mcp.r120.002090
- Chen, Z.; Huang, J.; Li, L. *TrAC, Trends Anal. Chem.* **2019**, *118*, 880–892. doi:10.1016/j.trac.2018.10.009
- Abrahams, J. L.; Taherzadeh, G.; Jarvas, G.; Guttman, A.; Zhou, Y.; Campbell, M. P. *Curr. Opin. Struct. Biol.* **2020**, *62*, 56–69. doi:10.1016/j.sbi.2019.11.009
- Hu, H.; Khatri, K.; Zaia, J. *Mass Spectrom. Rev.* **2017**, *36*, 475–498. doi:10.1002/mas.21487
- Jansen, B. C.; Falck, D.; de Haan, N.; Hipgrave Ederveen, A. L.; Razdorov, G.; Lauc, G.; Wuhrer, M. *J. Proteome Res.* **2016**, *15*, 2198–2210. doi:10.1021/acs.jproteome.6b00171
- Jansen, B. C.; Reiding, K. R.; Bondt, A.; Hipgrave Ederveen, A. L.; Palmblad, M.; Falck, D.; Wuhrer, M. *J. Proteome Res.* **2015**, *14*, 5088–5098. doi:10.1021/acs.jproteome.5b00658
- Deshpande, N.; Jensen, P. H.; Packer, N. H.; Kolarich, D. *J. Proteome Res.* **2010**, *9*, 1063–1075. doi:10.1021/pr900956x
- Park, G. W.; Kim, J. Y.; Hwang, H.; Lee, J. Y.; Ahn, Y. H.; Lee, H. K.; Ji, E. S.; Kim, K. H.; Jeong, H. K.; Yun, K. N.; Kim, Y.-S.; Ko, J.-H.; An, H. J.; Kim, J. H.; Paik, Y.-K.; Yoo, J. S. *Sci. Rep.* **2016**, *6*, 21175. doi:10.1038/srep21175
- Choo, M. S.; Wan, C.; Rudd, P. M.; Nguyen-Khuong, T. *Anal. Chem. (Washington, DC, U. S.)* **2019**, *91*, 7236–7244. doi:10.1021/acs.analchem.9b00594
- Mayampurath, A.; Yu, C.-Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H. *Anal. Chem. (Washington, DC, U. S.)* **2014**, *86*, 453–463. doi:10.1021/ac402338u

23. Klein, J.; Carvalho, L.; Zaia, J. *Bioinformatics* **2018**, *34*, 3511–3518.
doi:10.1093/bioinformatics/bty397
24. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, Austin, Texas, USA, June 28–July 3, 2010; 2010; pp 56–61.
doi:10.25080/majora-92bf1922-00a
25. Darula, Z.; Medzihradzky, K. F. *Mol. Cell. Proteomics* **2018**, *17*, 2–17.
doi:10.1074/mcp.mr117.000126
26. The UniProt Consortium. *Nucleic Acids Res.* **2019**, *47*, D506–D515.
doi:10.1093/nar/gky1049
27. Bern, M. W.; Kil, Y. J. *J. Proteome Res.* **2011**, *10*, 5296–5301.
doi:10.1021/pr200780j

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>). Please note that the reuse, redistribution and reproduction in particular requires that the authors and source are credited.

The license is subject to the *Beilstein Journal of Organic Chemistry* terms and conditions: (<https://www.beilstein-journals.org/bjoc>)

The definitive version of this article is the electronic one which can be found at:
<https://doi.org/10.3762/bjoc.16.180>