

Glyco-Informatics

Understanding the glycome: an interactive view of glycosylation from glyco compositions to glycoepitopes

Daive Alocci^{2,3}, Marie Ghraichy^{2,4}, Elena Barletta^{2,3},
Alessandra Gastaldello^{2,3}, Julien Mariethoz^{2,3}, and Frederique Lisacek^{2,3,5,1}

²Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, 7 Route de Drize, 1227 Geneva, Switzerland, ³Computer Science Department CUI, University of Geneva, 7 Route de Drize, 1227 Geneva, Switzerland, ⁴Division of Immunology, University Children's Hospital Zurich, Zurich, Switzerland, and ⁵Section of Biology, University of Geneva, Geneva, Switzerland

¹To whom correspondence should be addressed: Tel: +41-22-379-0195, Fax: +41-22-379-0250;
e-mail: frederique.lisacek@sib.swiss

Received 25 January 2018; Revised 16 February 2018; Editorial decision 28 February 2018; Accepted 2 March 2018

Abstract

Nowadays, due to the advance of experimental techniques in glycomics, large collections of glycan profiles are regularly published. The rapid growth of available glycan data accentuates the lack of innovative tools for visualizing and exploring large amount of information. Scientists resort to using general-purpose spreadsheet applications to create *ad hoc* data visualization. Thus, results end up being encoded in publication images and text, while valuable curated data is stored in files as supplementary information. To tackle this problem, we have built an interactive pipeline composed with three tools: Glynsight, EpitopeXtractor and Glydin'. Glycan profile data can be imported in Glynsight, which generates a custom interactive glycan profile. Several profiles can be compared and glycan composition is integrated with structural data stored in databases. Glycan structures of interest can then be sent to EpitopeXtractor to perform a glycoepitope extraction. EpitopeXtractor results can be superimposed on the Glydin' glycoepitope network. The network visualization allows fast detection of clusters of glycoepitopes and discovery of potential new targets. Each of these tools is standalone or can be used in conjunction with the others, depending on the data and the specific interest of the user. All the tools composing this pipeline are part of the Glycomics@ExpASY initiative and are available at <https://www.expasy.org/glycomics>.

Key words: glycan profile, glycoepitope/web tool, glycoinformatics, visualization

Introduction

In the last decade, high-throughput experimental techniques for glycan identification and quantification have been improved (Ruhaak et al. 2009; Ruhaak et al. 2010; Kemna et al. 2017; Planinc et al. 2017; Adamczyk et al. 2018). Furthermore, several studies have been published with large collections of samples generating considerable amount of glycome profiling data (Knezević et al. 2009; Pucić

et al. 2011). In fact, glycomics is starting to play an important role in biomarker discovery (Adamczyk et al. 2012; Ruhaak et al. 2013; Taniguchi and Kizuka 2015) especially related to cancer. Aberrant glycosylation patterns have been observed in several types of cancer cells (Jin et al. 2017), in diseases like diabetes (Keser et al. 2017) or in processes like inflammation (Schnaar 2016). Changes in glycan structural properties take place not only in reaction to a pathological

state but also according to aging, body index mass and other factors like smoking (Knezevic et al. 2010; Krištić et al. 2014; Gudelj et al. 2015). Each of these high-throughput studies has generated a large quantity of the so-called glycan profiles defined by a list of glycan compositions associated with their quantification in a specific sample.

Although glycoinformaticians have created software solutions to support some aspects of glycan identification and quantification (Ceroni et al. 2008; Jansen et al. 2015, 2016), most of the analyses are still carried out manually. This is both time-consuming and labor-intensive. The lack of tools for comparing and visualizing glycan profiles confines scientists to using general purpose spreadsheet applications to create *ad hoc* visualizations. Thus, results end up being encoded in publication images and text, while valuable curated data are stored in tables and published as Supplementary material. In this scenario, GlycoViewer (Joshi et al. 2010) was the first most innovative and pioneering visualization to summarize and compare sets of glycan structures. Although it did not account for the quantification data at the time, this software could summarize up to hundreds of structures in a single figure allowing the generation of high-level views of glycans from one protein, or a cell, or a tissue or a whole organism. Then, GlycomeAtlas (Konishi and Aoki-Kinoshita 2012) can be considered as the first successful visualization tool for quantitative glycomic data. This software has been specifically designed for exploring glycan profiles produced by Consortium for Functional Glycomics (CFG) (Raman et al. 2005, 2006; Ismail et al. 2011; North et al. 2012). Although GlycomeAtlas can be used to easily navigate in CFG data, it cannot be extended nor run with other datasets. GlycoPattern (Agravat et al. 2014), together with GlycanMotifMiner (Chollet et al. 2012), represent another effort to visualize and mine CFG data. The implementation of dendrograms and heatmaps supports the detection of similar motifs in glycan binding proteins (GBP) and the comparison of several glycan array experiments. Both software have been designed and developed within CFG to define a common way of storing, visualizing and analyzing data for the consortium.

To tackle the scarcity of visualization tools that do not depend on the origin of the dataset, and to speed up the analysis of glycan profiles and link the latter to functional information, we introduce a series of three tools that can be run independently as well as chained to one another to promptly visualize and link information at different levels of granularity and highlight functional properties of glycans. Glynsight is the first tool of the pipeline. It has been designed to visualize and compare glycan profiles and it processes glycan compositions. Glynsight is using an innovative visualization to bring out hidden patterns inside or among glycan profiles. This web-based tool is provided with an upload function for spreadsheets and allows the user to export each profile as an image or the whole collection as a JSON file. Glynsight is also connected to GlyConnect (<https://glyconnect.expasy.org>, manuscript in preparation), our glycomics dashboard, to suggest potential glycan structures related to specific compositions. Secondly, EpitopeExtractor is a unique tool for extracting glycan determinants from a list of glycan structures. It is based on our substructure search engine, GlyS3 (Alocci et al. 2015), and a curated collection of glycan determinants (Kawasaki, Nakao, and Tominaga 2008; Cummings 2009; Pérez et al. 2015; Mariethoz et al. 2016). Potential glycan structures highlighted in Glynsight can be forwarded to EpitopeExtractor to extract glycan determinants. The last tool of the pipeline is Glydin', a glycoepitope network viewer created for exploring and visualizing the relationship between glycan determinants based on shared monosaccharide composition or the glycosyltransferase needed to get one from the other. Glydin', like EpitopeExtractor, is based on GlyS3 and relies on the same

collection of curated glycoepitopes. EpitopeExtractor analysis results can be sent to Glydin' to be mapped on the network. Glynsight, EpitopeExtractor and Glydin' are web-based tools developed by the Proteome Informatics Group at SIB Swiss Institute of Bioinformatics. The first two have been developed using web components in synergy with D3.js where Glydin' is build with only HTML, CSS and D3.js. All three tools are part of the collection hosted by Glycomics@Expasy (www.expasy.org/glycomics) (manuscript in preparation).

We validated this approach while processing data found in supplementary material of Anugraham et al. (2014). In this study, the glycan profiles of healthy vs. ovarian cancer cells were generated, analyzed and compared. We show that most results presented in the article could be confirmed by viewing and inspecting the profiles displayed in Glynsight and suggest the use of this tool as a substantial time saver in extracting relevant information. We also illustrate how our pipeline leads to narrow down explanations on the role of differentially expressed glycans in the comparative mode. Finally and in the same pipeline, we match the results regarding glycan determinants and provide further contextual information for characterizing potential ligands.

Results

We used the published dataset (Anugraham et al. 2014) to illustrate and validate the use of the pipeline. This study on epithelial ovarian cancer (EOC) compared the membrane glycosylation of two ovarian noncancer epithelial with four cancerous cell lines. Data were extrapolated from a study published by the American Society for Biochemistry and Molecular Biology, which highlights the potential for specific structural and isomeric changes in *N*-glycans in detecting and potentially treating ovarian malignant tumors. We kept the headings of the paper to emphasize the similarity between the original results and those output by our pipeline.

Relative quantitation of *N*-glycans of membrane proteins from noncancer and cancer cell lines

The profiles created in Glynsight show the glycan compositions in nine groups according to the incrementally ordered content in monosaccharides (see [Material and methods](#)). In the individual mode, each experiment was displayed and every profile showed that glycans whose compositions contain 9–12 monosaccharides are more highly expressed than the others (Figure 1A). In the differential mode, we systematically compared the cancer and noncancer cell lines. The greatest quantitative difference in glycan compositions is found between HOSE 17.1 (noncancer) and SKOV3 (cancer) (Figure 1A). In qualitative terms the individual compositions of IGROV1 are the most distinct compared to the others. This is potentially reflecting its high carcinogenic properties previously found in nude mice (Bénard et al. 1985; Kurbacher et al. 2011; Buehler et al. 2013) (Figure 1B). Note that 33 out of 55 individual glycan compositions are present in all cell lines. Of the remaining 20, 17 were found exclusively in cancer cell lines (red—in IGROV1 are all expressed), and three are found only in both noncancer cell lines (H4N4F2, H5N4F2 and H5N4F3—blue).

In order to determine the possible quantitative differences between the glycans, in the membrane proteins of the cancer and noncancer cell lines, Anugraham and coauthors statistically analyzed the *N*-glycans common to all the cell lines and divided them into four classes: high mannose, hybrid, complex (neutral and sialylated) and core fucosylated (Anugraham et al. 2014). Such division is not necessary using Glynsight since the differential display immediately shows that the

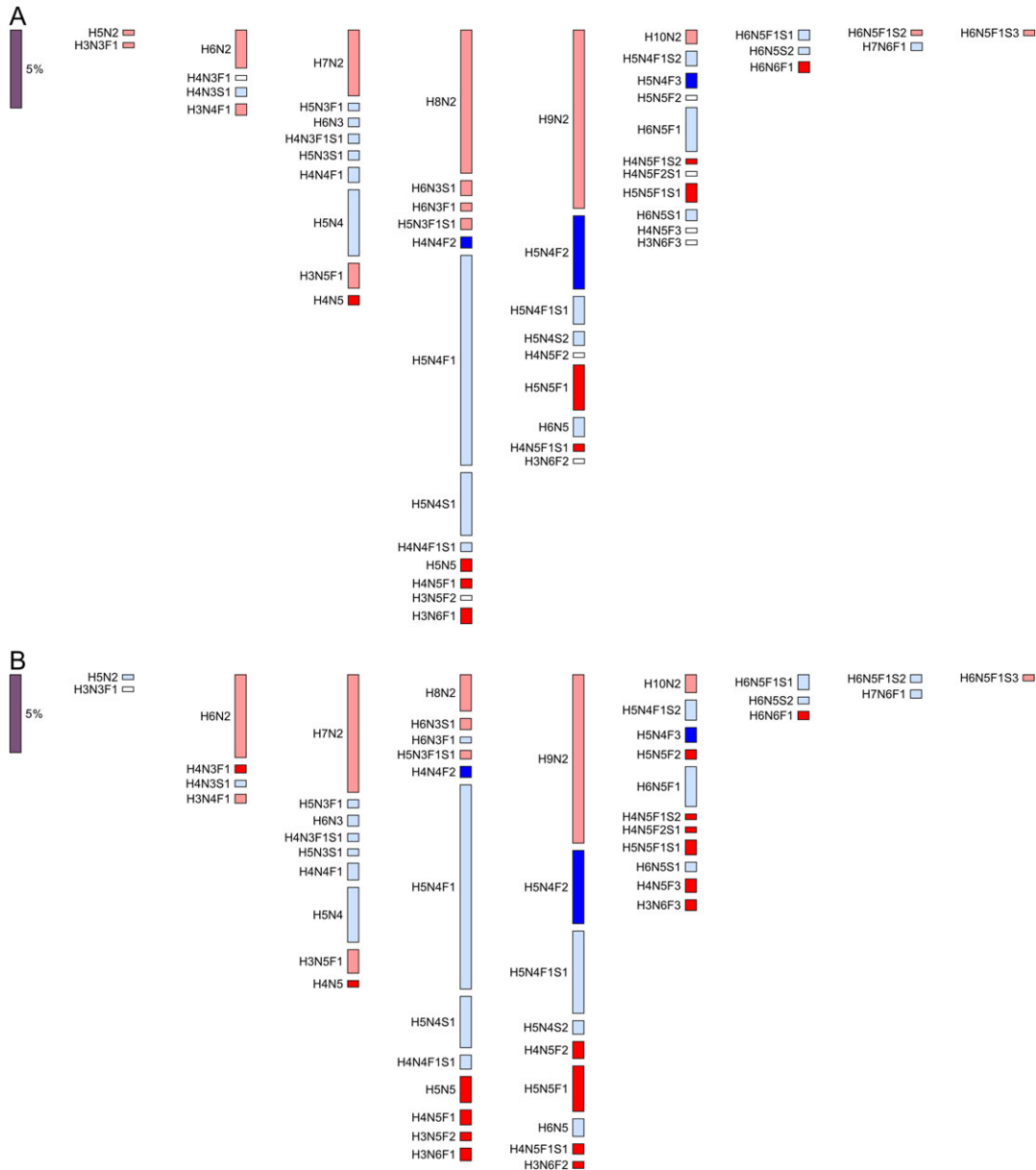


Fig. 1. Differential display of glycan compositions between HOSE 17.1 (noncancer cell line) and **(A)** SKOV3 (cancer cell line) / **(B)** IGROV1 (cancer cell line). Compositions are displayed in columns and in each column the overall number of monosaccharides is constant. The height of a bar represents the difference between the expression of a specific composition in HOSE 17.1 with that of the same composition in **(A)** SKOV3 and **(B)** IGROV1. Brighter blue (respectively, red) bars highlight differences arising from compositions unique to HOSE 17.1 (respectively, **(A)** SKOV3 and **(B)** IGROV1) while pale blue (respectively, red) bars show differences arising from compositions expressed in all experiments but higher (respectively, lower) in HOSE 17.1 than in **(A)** SKOV3 and **(B)** IGROV1. For example, H5N4F2 is over-expressed only in HOSE 17.1 (and 0 in **(A)** SKOV3 and **(B)** IGROV1) while H5N4F1 is over-expressed in HOSE 17.1 in contrast with its nonzero but lesser expression in **(A)** SKOV3 and **(B)** IGROV1.

compositions with a high mannose content (compositions with only Hex (H) and HexNAc (N), where $\text{Hex} > 4$ and $\text{HexNAc} = 2$) appear highly expressed in cancer cell lines compared to noncancer ones; the compositions with more than four HexNAc (*N*-acetylglucosamine and *N*-acetylgalactosamine) are less represented in the nontumor cell lines than the tumor.

The compositions involving fucosylation and sialylation were analyzed according to the method introduced by Hayes and coauthors for the conversion of raw data (Hayes et al. 2012). For each composition we therefore considered not only the type of monosaccharide but also the corresponding number of Fuc and *N*-acetylneuraminic acid (NeuAc)

monomers. To begin with, the relative intensities of fucosylated compositions, as reported in (Anugraham et al. 2014), were multiplied by their respective quantity of Fuc. At this point, in order to obtain the relative total intensity for each cell line, the intensities previously calculated on each composition were summed, and these sums were divided by the total amount of monosaccharides present in all fucosylated compositions. Six values were obtained, one for each cell line (Table SI). Then, the same procedure was applied to the sialylated compositions (Table SII). In the end, the expression of fucosylated compositions shows more variation in the cancer cell lines, mostly in IGROV1, but is higher in the two noncancer cell lines. In contrast, the expression of

sialylated compositions was not distinguishable but rather equally distributed in either cancer/noncancer cases.

Bisecting GlcNAc

Bisecting GlcNAc is added to a glycan structure through a reaction catalyzed by a specific enzyme encoded by the MGAT3 gene. As demonstrated by qRT-PCR analysis, the lower transcript abundance of MGAT3 observed for mouse and human normal ovary contrasts with the large increase in MGAT3 transcripts observed in all cases of epithelial endometrioid carcinoma analyzed in both mouse and human (Abbott et al. 2008). This means that this specific glycan synthesis reaction will be under-expressed in the noncancer cell lines compared to EOC cell lines. Overall, 25 compositions (HexNAc > 4), with a hypothetical bisecting GlcNAc were considered in our analysis (File S3A). Most of the glycan compositions with an odd number of HexNAc appeared, as expected, in the EOC cell lines. For example, H5N5F1 present only in the cancer lines, reaches a very high expression in OVCAR3 (5.93) and H5N5F2 is exclusively present in IGROV1. As a consequence, almost all the compositions containing five HexNAc are either more abundant or exclusively represented in cancer cells (Figure 2). In contrast, H6N5F1 showed a differential range value of 2.26 to -2.52 towards the nontumor HOSE 17.1 cell line when compared with the cancer ones (the range for HOSE 6.3 was still significant, greater than 1).

Such differential expression led us to consider the possible glycan structures known to match these specific compositions (H6N5F1). This information is recorded in the GlyConnect database (<https://glyconnect.expasy.org>) (manuscript in preparation). None of the reported 24 structures, searched in GlyConnect in “ALL” tissues, displayed a bisecting GlcNAc, which is rarely found in EOC cells. The same appears when “Urogenital System” is selected as specific tissue, narrowing down the structures to eleven (File S3B). Consequently,

H6N5F1 is more likely to stem from H6N5, identified as nonbisected and represented in pale blue, than from H5N5F1. As previously stated fucosylation, appears to be a prerogative of noncancer cells. In contrast, H5N5F1 is more likely to stem from the possible bisection of H5N4F1, than from the fucosylation of H5N5, through the addition of an *N*-acetylhexosamine (Figure 2), which is line with (Miwa et al. 2012), who observed that “the presence of a bisecting GlcNAc on a biantennary *N*-glycan terminating in GlcNAc prevents the subsequent action of the core fucosyltransferase FUT8”.

We proceeded with the analysis of other compositions. Overall, only nine were identified in the database as bisected. These compositions were only expressed in the cancer cell lines except for H3N5F1 that is weakly expressed in noncancer cell lines. The other 16 compositions match nonbisecting structures and are mostly expressed in cancer cell lines. Individual analysis of glycan compositions that are differentially expressed in cancer cell lines confirmed the association of nonbisecting GlcNAc with the disease. In particular, the most over-expressed compositions in cancer are exclusive to IGROV1 and SKOV3 cell lines.

LacdiNAc-type *N*-glycans (glycoepitopes)

To evaluate our full pipeline, we considered the profiles of HOSE 17.1 and SKOV3, since they show the most distinctive quantitative trends. Starting from Glynsight in differential mode (Figure 1A), the differentially expressed compositions in each profile were divided in two groups: cancer (bright and pale red bars) and noncancer (bright and pale blue bars). Using the connection with Glyconnect, we enriched the information from compositions to glycan structures, selecting all the *N*-glycans that are expressed in *Homo sapiens* and match a selected composition in one of the two groups. No restriction was imposed on tissue. In the end, the cancer-related group contains 163 structures whereas the noncancer group contains 245.

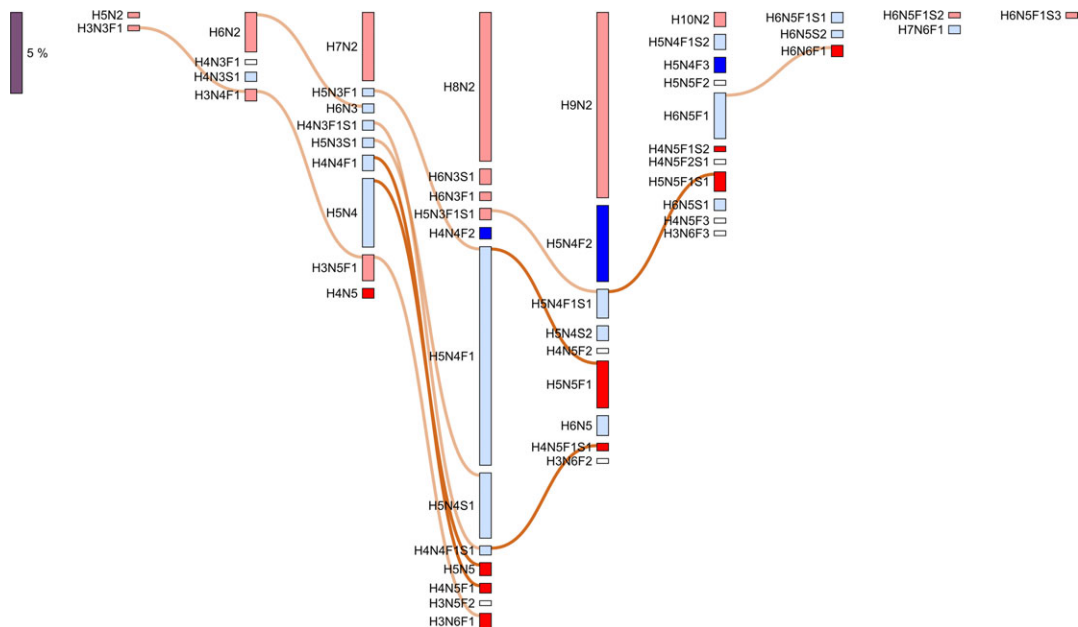


Fig. 2. Influence of HexNAc number variation shown in the differential display between HOSE 17.1 (noncancer cell line) and SKOV3 (cancer cell line). The differential display shown in Figure 1 is reproduced here and complemented with orange (chosen for the readability of the figure but modifiable as explained in [Material and methods](#)) connectors between different bars to single out the addition of HexNAc (N). User-selected brighter connections emphasize the addition of a HexNAc to compositions with four HexNAc (N4) and highlight the potential for a bisecting GlcNAc. A change in the bar color from pale blue (over-expressed in noncancer cell lines) to bright red (only expressed in cancer cell lines) can also be observed.

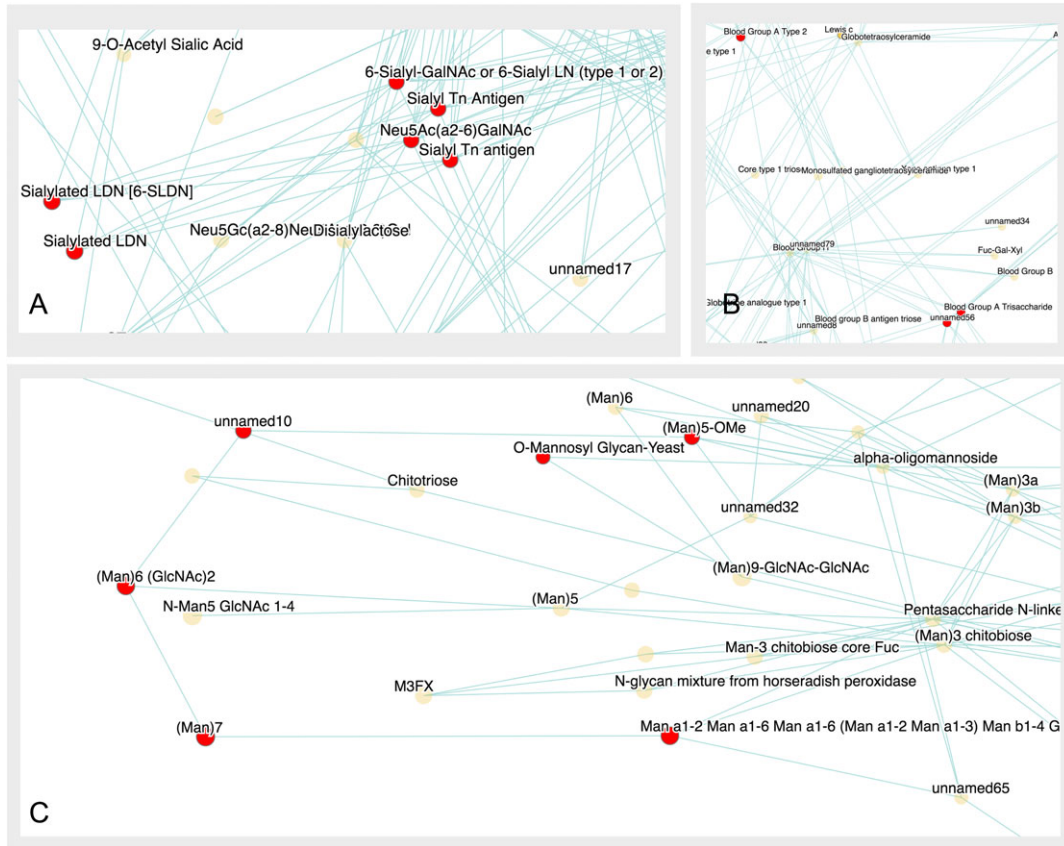


Fig. 3. Mapping of 19 cancer-related glycoepitopes on the Glydin' network. Glycoepitopes related to cancer composition (bright and pale red bars) from the differential display of SKOV3 and HOSE 17.1 (Figure 1A), have been extracted using EpitepeXtractor. The 19 glycoepitopes unique to cancer are mapped on the Glydin' network and appear as red nodes. They tend to cluster together. The three most populated clusters are shown: (A) this cluster contains Sialyl Tn Antigen and Sialyl LacdiNAc, (B) Blood group A like cluster and (C) High-mannose cluster.

Subsequently, the glycan determinants present in the two groups of glycans were submitted to EpitepeXtractor. The cancer structures matched 80 glycoepitopes and the noncancer, 82. These data can be directly sent to Glydin', the Glycoepitope network. Due to the high amount of glycoepitopes extracted, an accurate visual comparison is a challenge, both in EpitepeXtractor and in Glydin'. To ease this comparison, only the unique glycoepitopes distinctly identified in each group were considered, in an attempt to detect exclusive characteristics. This led to keep 19 and 20 glycan determinants, respectively, in the cancer and noncancer sets.

The 19 glycan determinants only found in the cancer set, are singled out as red nodes in the Glydin' network. Inspecting the subnetworks surrounding these nodes of interest, three main groups of glycan determinants emerged: (i) blood group A related determinants, (ii) Sialyl Tn antigen with Sialylated LacdiNAc and (iii) mannose related glycoepitopes (Figure 3). Similarly in the noncancer case, Lewis related glycoepitopes, in particular X, C and A-like, blood group B and H antigen appear as distinctive traits of noncancer glycan structures (Figure 4).

The above shows results obtained from compositions present in either of the two groups defined from Glynsight as cancer (red bars) and noncancer (blue bars). A more focused analysis of HOSE 17.1 and SKOV3 glyco-profiles was performed using only the compositions exclusively expressed either in cancer or noncancer cell lines, i.e. those displayed in bright red and bright blue bars in Glynsight. As shown in Figure 1A, this limits the compositions to {H4N5, H5N5, H4N5F1, H3N6F1, H5N5F1, H4N5F1S1, H4N5F1S2, H5N5F1S1, H6N6F1}

for the cancer group and to {H4N4F2, H5N4F2, H5N4F3} for the noncancer group. Following the same steps as described above led to single out the same glycoepitope subnetworks. However, comparing Figure 3A with Figure 6 shows that LacdiNAc (LDN in the figure) is only present in the second analysis (Figure 6) whereas both contain the sialylated LacdiNAc.

The presence of LacdiNAc in cancer-related glycans is brought out in these results and matches the observations made by Anugraham and coauthors from the structures identified by mass spectrometry. However, our pipeline also selects blood group antigens and Lewis related glycoepitopes as potential distinctive properties of respectively cancer and noncancer-related profiles. We noted that other studies on ovarian cancer (Christiansen et al. 2014) have reported the presence of Sialyl Tn antigen in association with cancer cells. This determinant is also distinguished as related to cancer in our analysis.

All the steps of the glycoepitope extraction analysis, including the IDs of GlyConnect structure and Glydin' glycoepitope, are available in the Supplementary data (File S3D). This file also includes the links for exploring the cancer and noncancer results in Glydin'.

Gene expression of specific glycosyltransferases in ovarian cancer cell lines

Anugraham and coauthors have also reported on differential enzymatic regulation of the glycosylation pathway in each cancer and noncancer cell lines, via transcriptomics experiments. As mentioned in

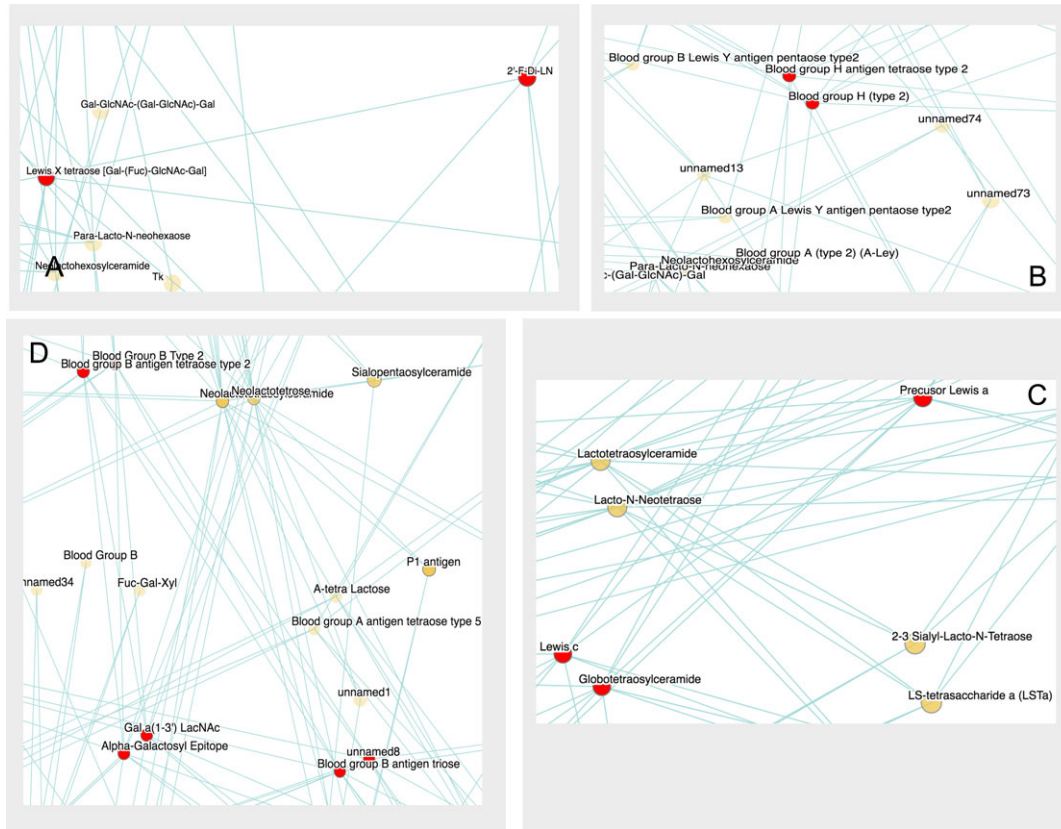


Fig. 4. Mapping of 20 noncancer-related glycoepitopes on the Glydin' network. Glycoepitopes were generated from noncancer-related glycan compositions shown as bright and pale blue bars in the Glynsight differential display of SKOV3 and HOSE 17.1 (Figure 1A) and have been extracted using EpitopeExtractor. The 20 glycoepitopes unique to noncancer are mapped on the Glydin' network and appear as red nodes. They tend to cluster together. The four most populated clusters are shown: (A) and (C) Lewis X/A clusters (B) Blood group H cluster and (D) Blood group B cluster.

Discussion, glycan compositions were searched in GlyConnect in order to match all human N-linked structures stored in the database, irrespective of the tissue information. In this section, we filtered the human N-linked structure set with the “Urogenital System” requirement. This smaller selection in Glynsight individual profiles of SKOV3 and OVCAR3 is exclusively composed of structures with bisecting GlcNAc.

This observation matches the statistically significant differential expression of the MGAT3 gene between the cancer and noncancer cell lines reported by Anugraham and coauthors. Furthermore, we have observed that all fucosylated glycan compositions are highly expressed in noncancer cell lines and that the core and non-core fucosylated glycans could be distinguished.

We examined the fucosylation linkage while dividing the compositions according to the fucose (Fuc) residue number (F1, F2 and F3). The mono-fucosylated compositions were searched in GlyConnect (N-linked, Homo sapiens and Urogenital System) and out of all the reported structures showing core fucosylation (140/151 structures—92,7%), a majority displayed a higher level of (α 1–6) fucosylation (File S3C). Six mono-fucosylated compositions (H6N5F1, H6N6F1, H5N4F1S1, H5N5F1S1, H5N4F1S2 and H6N5F1S3), matched structures with the non-core (α 1–2) and the (α 1–3) bond appeared to be on display (File S3C). Clicking on those structures redirects to entries in our database where the reference articles pointed at evidence of possible (α 1–2)-linked Fuc residue while not excluding core-fucosylated analogues (Stroop et al. 2000). The (α 1–3)-fucosylated N-glycans were

also found, representing sialyl Le(x) elements in addition to conventional N-glycans established for human serum transferrin (hST) (van Rooijen et al. 1998), and also in a tri-sialyl oligosaccharide isolated from human serum (Nakagawa et al. 1995).

The (α 1–6) fucosylation linkage is related to the FUT8 gene expression. The (α –1,6)-fucosylated structures are qualitatively associated with the highly tumorigenic cancer line IGROV1 and to cancer in general, while in noncancer essentially two compositions (H5N4F1 and H5N4F1S1) are the ones that dimmed and shifted our conclusions towards the noncancer cell lines because of their extremely high values.

Bi-fucosylated and tri-fucosylated compositions, i.e. H4N4F2, H5N4F2 and H5N4F3 (possibly derived from H4N4F1 and H5N4F1) were found only in noncancer cell lines while H4N5F2S1, H3N6F3, H5N5F2, H4N5F3, H3N6F2, H4N5F2 and H3N5F2 were shown uniquely in IGROV1. This supports either GlcNAc bisecting or FUT8 gene expression in cancer cells. Considering the possible non-core fucosylation linkages for F2 and F3, (α 1–2) and (α 1–3) were the other two options in structures matched in the GlyConnect database. It is not possible to determine which fucosyltransferase gene is related to these compositions, but looking at the quantitative RT-PCR results of the mRNA transcripts (Anugraham et al. 2014) we can suggest a correlation with the higher expression of the FUT2 enzyme and possibly link the high expression of FUT4 to (α –1,3)-fucosylation (Figure 5).

Anugraham and coauthors reported differences in the sialylation gene expression. Despite our ability to spot sialylated compositions,

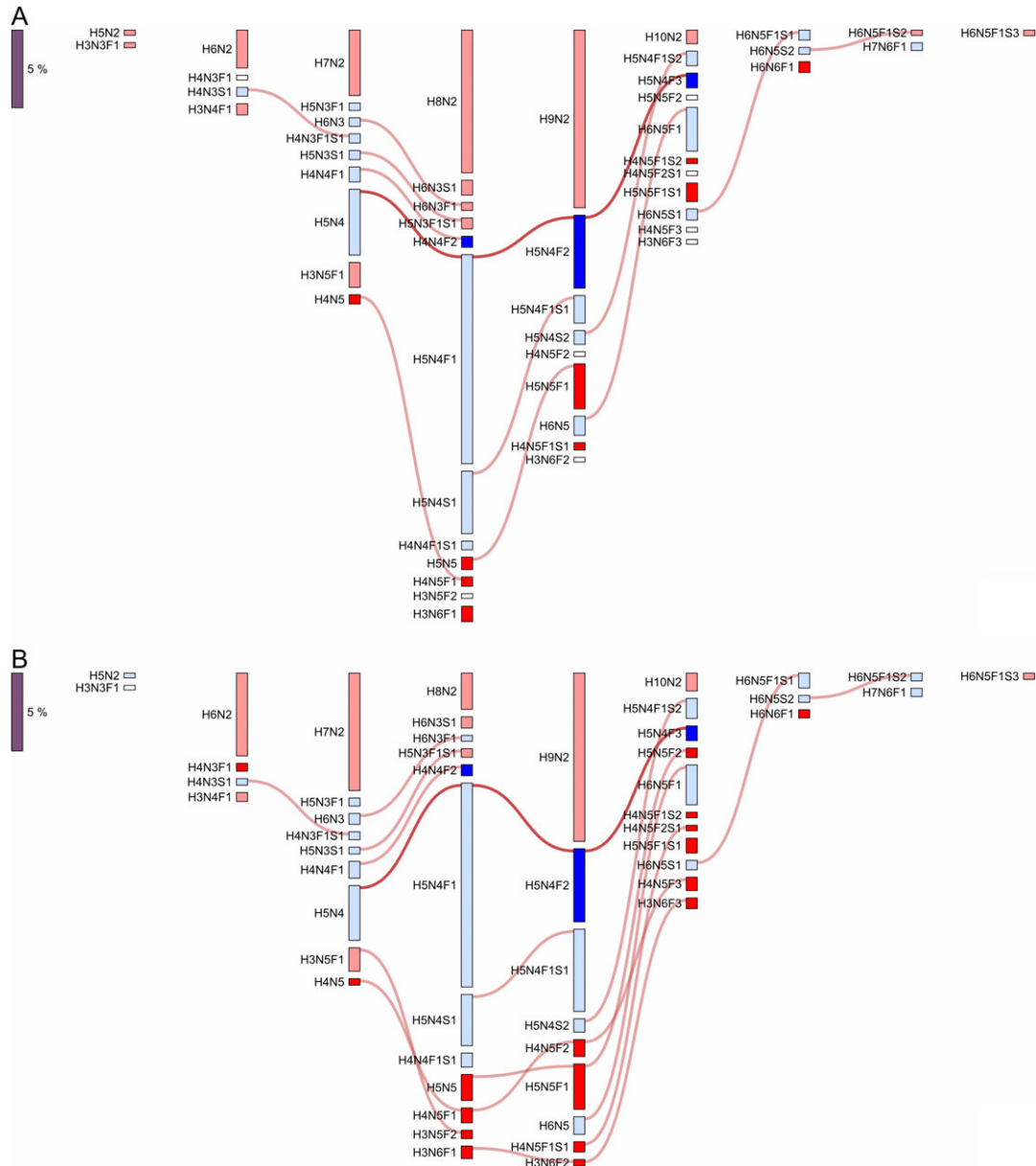


Fig. 5. Influence of fucosylation shown in the differential display between HOSE 17.1 (noncancer cell line) and **(A)** SKOV3 (cancer cell line) and **(B)** IGROV1 (cancer cell line). The differential display shown in Figure 1 is reproduced here and complemented with red connectors between bars to single out the addition of a fucose (fucosylation). The brightness of the connectors is user-defined. Connectors are faded by default and brightened upon a click. Fucosylated compositions are mostly over-expressed in noncancer cell lines. Connectors from H5N4 to H5N4F3 via H5N4F1 and H5N4F2 were brightened to emphasize compositions unique to noncancer HOSE 17.1 (bright blue) with respect to both **(A)** SKOV3 and **(B)** IGROV1. However, HOSE 17.1 vs. **(A)** SKOV3 and **(B)** IGROV1 shows a higher density of red connectors in **(B)**. The expression of more fucosylated glycans is possibly correlated with the higher expression of fucosyltransferases observed in IGROV1 cell line by (Anugraham et al., 2014).

exclusively in the cancer cell lines (H4N5F1S1, H4N5F2S1, H5N5F1S1, H4N5F1S2) we could not relate this information with the expression of ST3GAL or ST6GAL genes since both (α 1–3) and (α 2–6) sialylation cannot be distinguished in Glynsight.

Discussion

Validation of experimental results

In an effort to illustrate and validate the use of Glynsight, EpitopeXtractor and Glydin', we have analyzed a recently published

dataset on membrane glycosylation in EOC. Data were extracted from spreadsheets in Supplementary material and imported into Glynsight, the entry point of our pipeline.

Glynsight performed a fast and in-depth comparison of the profiles relative to six cell lines reported in the dataset. SKOV3 and IGROV1 appeared to be the most differentiated cell lines, IGROV1 being the only profile where all the glycan compositions were significantly expressed. This might lead to a subsequent and more careful analysis of this specific cancer cell line.

Three fucosylated compositions H4N4F2, H5N4F2 and H5N4F3, not previously discussed in the examined article (Anugraham et al.

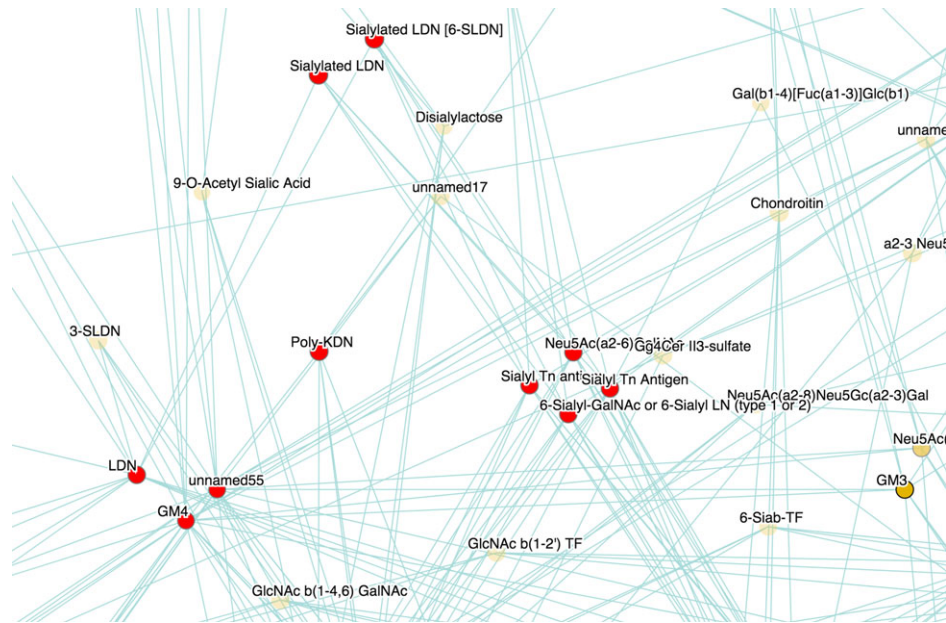


Fig. 6. LacdiNAc subnetwork in Glydin'. Glycan determinant cluster that emerges when mapping glycoepitopes from compositions exclusively expressed in cancer (bright red bars) related to the differential display of SKOV3 and HOSE 17.1 (Figure 1A). Glycoepitopes shown in red are the ones extracted by EpitopeXtractor. The presence of LacdiNAc (LDN) confirms the observation made in (Anugraham et al. 2014). Other red nodes can be seen as possible targets for further experimental studies.

2014), were identified as discriminant. These compositions can only be characteristically found in noncancer cells (Figure 5), in both HOSE cell lines. Furthermore, we confirmed a higher mannose content in cancer cell lines compared to noncancer as well as a GlcNAc bisection, contrary to fucosylated glycan compositions that were found to be expressed in noncancer cell lines. The GlcNAc bisection has been confirmed to be most likely related to the expression of MGAT3 gene, which encodes the enzyme adding the monosaccharide in a beta 1–4 linkage to the core of the N-glycans (Abbott et al. 2008; Anugraham et al. 2014), contrasting with fucosylated structures that were discovered to be qualitatively associated to cancer. Whereas increased MGAT3 transcripts were identified in the EOC through RT-PCR experiments (Anugraham et al. 2014), as supported by our Glynsight analysis, fucosylation and sialylation genes could not be described. As Glynsight links each composition to a collection of putative glycan structures via GlyConnect the structural variety of these structures makes it difficult to map unique enzyme information and correlate with gene expression data. We could however hypothesize on the FUT8 gene associated with the (α 1–6) fucosylation linkage. Structures containing Fuc(α 1–6) proved to be equally distributed in all highly expressed compositions.

Each structure is associated with tissue expression information that cannot yet be confidently leveraged due to the scarcity of structure-to-tissue correlations. The precision of the proposed approach will increase with regular updates of knowledge in GlyConnect via new annotated glycan structure import from public results. At the time of writing, GlyConnect contains 3404 glycan structures of which 1557 are expressed in human.

Novelty brought by the pipeline

Extracting and displaying glycan structures highlights trends in the possible monosaccharide arrangements of each composition. Such a collection of putative structures may not be comprehensive due to the

incompleteness of Glyconnect, but it enables functional interpretation through running EpitopeXtractor and executing in silico glycoepitope extraction. By simulating the identification of potential ligands, EpitopeXtractor combined with Glydin' helps estimating the binding characteristics of highly expressed compositions. This saves time and work required to map glycoepitopes. To our knowledge, this in silico glycoepitope extraction and mapping from differentially expressed glycan compositions is new in glycoinformatics. With the data tested, EpitopeXtractor marked the evidence of Lewis x/a as unique glycoepitopes, underlining a specific characteristic of noncancer glycan compositions. However, this observation does not reflect a reliable correlation between the presence of Lewis epitopes and noncancer cell lines. It partly contradicts a previous study (Escrevente et al. 2006), where m130 and GG ovarian carcinoma cell lines were shown to express high levels of Lewis x, sialyl Lewis a and Lewis b that have been associated with major cancer activities. Further experimental work would be needed to determine the specificity of Lewis x glycoepitopes. Nonetheless, EpitopeXtractor correctly reported the presence of LacdiNAc in cancer-related structures, which was proven by Anugraham and coauthors using mass spectrometry. In addition, Sialyl Tn Antigen emerged as a candidate to distinguish cancer from noncancer cell lines. These epitopes have been already described in relation to ovarian cancer in (Christiansen et al. 2014).

Potential for further applications

With this approach, glycoepitope analysis shows the potential for bringing out new hypotheses and helping scientists narrow down the range of glycan determinants of interest. Note that we currently mainly focus on mammalian N- or O-linked glycans.

Our pipeline, in particular Glynsight, can be tested as a discovery tool in the research field as well as quality control tool for pharmaceutical companies. Batch to batch glycosylation analysis has become more and more important to assess the quality of therapeutic monoclonal

antibodies (mAbs) (Planinc et al. 2017). Glynsight proposes an innovative way of displaying glycan profile data, which can be used as a signature for each batch. Moreover, the differential display can easily catch discrepancies among batch glycan profiles leading to accurate and precise quality assurance. For confidential data, a desktop version of Glynsight is available upon request.

Glycan profile interactive visualization can potentially be used in a wide range of studies, from glycoproteomics (Stadlmann et al. 2017) to therapeutic mAbs glycan analysis (Cymer et al. 2017), to compare as well as publish results avoiding the usage of endless spreadsheets and tables.

To conclude, Glynsight is tackling several issues, which arise in the analysis of glycan profiles. In addition, our strategy for performing fast glycan determinant extraction, using EpitepeXtractor and mapping with Glydin' suggests potential new targets for glycan biomarker discovery.

Future plans

In the future we plan to expand the GlyConnect knowledge base as well as to either enlarge or diversify pipelines adding new tools and new strategies. In the course of 2018, a centralized glycan epitope database will be connected with EpitepeXtractor and Glydin'. This will allow for users' contribution to knowledge of glycan determinants. We have also scheduled changes in the Glydin' display to accommodate more information. For example the edges of the network will be associated with enzyme knowledge while the nodes will be associated with expression data to highlight glycoepitopes in over or under-expressed glycans. Finally, we also plan to release a Glynsight version for the simultaneous comparison of larger collections of profiles. To feed our future plan further, we also collect comments and feedback from our collaborators and users since matching their needs is our primary goal. Suggestions are an endless source of ideas for developing better and smarter tools.

Material and methods

Data

We have selected a recently published dataset (Anugraham et al. 2014) from which we extracted the details in Supplementary material. We considered Table I in (Anugraham et al. 2014) and transformed all N-glycan structures into compositions (Table I). We left O-glycan data due to the small range that was included. We define glycan composition as the number of monosaccharides, which composes the glycan out of a list of 7 main building blocks: Hexose (Hex), N-acetylhexosamine (HexNAc), Fuc, NeuAc, N-Glycolylneuraminic acid (NeuGc) including Sulfate for O-linked. We use the simple notation shown in Table I. For instance the N-linked core is H3N2.

This data was then converted into the Glynsight format and loaded into our software. Potential isomers were not taken into consideration

Table I. Notation for monosaccharides

Monosaccharide	Encoding
Hex	H
HexNAc	N
Fuc	F
NeuGc	G
NeuAc	S
Sulfate	s

and consequently their quantity was added together to obtain a single value. The Glynsight session file is available in Supplementary data (File S1) and can be used to reproduce the results.

Tools

The pipeline which allows the user to explore data from compositions to glycoepitopes is split in three different web tools: Glynsight, EpitepeXtractor and Glydin. Each tool was designed to carry out a specific task and can be used either alone or in conjunction with the others.

The next sections give an in depth overview of each tool, showing the main features and how to use them.

Glynsight

Quantitative mass-profiling experiments allow the monosaccharide compositions of distinct sets of isobaric glycans (i.e. structures having the same mass) to be identified and the abundance of each isobaric set to be estimated, thereby producing a glycome in the form of a glycan composition-based profile. Usually, a profile consists in a list of independent glycan compositions associated with corresponding intensities. These profiles are the input of Glynsight.

The first step implemented in Glynsight is to bring out relationships between input compositions. To that end, compositions are incrementally ordered and structured in columns where all compositions with a constant overall number of monosaccharides are listed. For instance, if the profile is an N-glycome, then the first column of the profile is the N-core H3N2 (5 monosaccharides in total) and the next is H3N3 or H3N2F1, H4N2, H2N3F1 (6 monosaccharides in total), etc. In this way, the implicit inclusion of H3N2 in H3N3 or H3N2F1 is made obvious. More generally, this presentation of the profile highlights the monosaccharide shared between compositions. In this way, Glynsight shows the relationships between compositions in an interactive way. This allows the user to follow an increment in galactosylation or fucosylation, etc., as explained below.

The interface can be used to retrieve all the potential glycan structures, which match a particular composition. Potential structures are not generated using software but are retrieved from curated databases accessible via the GlyConnect platform (<https://glyconnect.expasy.org>). Glynsight is available online at <https://glycoproteome.expasy.org/glynsight/>.

Glycan profile upload

To begin with, the user needs to upload a set of glycan profiles. This operation is made simple through a multiple file upload in Glynsight under "Import Experiments" in "Manage Experiments" menu. Each glycan profile must be stored in a comma-separated value (CSV) type of file. All CSV files must contain two columns:

1. glycan composition and
2. glycan abundance as percentage or any other unit.

As shown in the example file (File S2), the glycan composition column must be named "composition" whereas glycan expression column must be named "quantification".

When the file is correctly imported, a grey tick will appear next to the filename and a new entry will be available in the experiment list situated on the left side of the Glynsight interface. By default the name of the file is taken as the name of the experiment, but clicking on "Rename Experiments" in "Manage experiments" allows the user to change any experiment name.

If a wrong experiment has been mistakenly uploaded, the button “Remove selected experiment” removes only the selected experiment (s) in the list. If a complete reset of Glynsight is needed, clicking on “Remove all experiments” will wipe out all the experiments uploaded.

Manage a session

A Glynsight session is a collection of uploaded glycan profiles. A processed collection can be downloaded using “Download session” button under the “Manage Session” menu. Glynsight packs all the information into a single JSON file ready to be shared and uploaded for future usage. Once the session file is uploaded, using “Upload session” button, Glynsight will automatically show the collection of glycan profiles under the “Experiment list”. In this case, reuploading all the glycan profiles is avoided, in an attempt to minimize mistakes.

Display mode

Glynsight has been designed for two main usages: visualize and compare glycan profiles. Each is characterized by a distinct mode and a characteristic display. A single glycan profile is visualized in the “individual display mode” while in the “differential display mode” the comparison of two profiles is shown.

The user can smoothly switch between the two modes by selecting the preferred one in the “Display mode” left menu. Switching display mode automatically triggers a change in the Experiment list enabling the selection of multiple experiments. The following sections explain how to use the individual and the differential displays.

Individual display

We assume here that “Individual Display” is selected in the left menu. Selecting a glycan profile in the experiment list triggers an update the viewer. Glynsight retrieves the data related to the profile from the local storage in the browser and updates the view to fit with the new data. As detailed above, the viewer shows compositions incrementally ordered and structured in columns. Within each column, they are ordered by increasing numbers of HexNAc from top to bottom.

A bar next to each composition represents the relative intensity in the glycan profile, whereas the bar color gives a visual indication of the expression level. Four different colors are used:

- Green corresponds to highly expressed glycan compositions
- Yellow corresponds to lowly expressed glycan compositions.
- White corresponds to unexpressed glycans.
- Grey corresponds to glycans present in other profiles but not in the one currently displayed.

Since the separation between highly or lowly expressed glycans cannot be fixed but it is relative to the experiment, the user can set the “Expression threshold” in the upper right part of the viewer under Scale Setting. This threshold defines the upper limit of the yellow bars and the lower bound of the green bars.

In addition to the “Expression threshold”, Scale Setting can also operate on the display scale represented by the leftmost purple bar and set by default to 5%. The value (number) and the unit (text) can be modified to reflect the quantification unit. In some cases, a larger value improves the interpretation of variations.

In the upper left part of the viewer, the Edge Selection setting is composed of six different switches with color pickers. Using the corresponding switch, the user can highlight a specific monosaccharide increment, which will be visible in the form of an edge (or connector) connecting two composition bars. Since each monosaccharide has an

associated color in a specific glycan encoding format, the color of the edge can be adapted using the color picker under the switch. By default, colors are following the Symbol Nomenclature for Glycans (SNFG) (e.g. red for Fuc, purple for sialic acid).

In the upper right corner of the viewer tab, a button with an arrow pointing down, allows the user to take a screenshot of the viewer for sharing or documenting purposes. The format of the image is set to scalable vector graphics. Custom settings of the viewer are kept when a new experiment is loaded. Consequently, selecting a new experiment will only modify the height of each bar reflecting the new data. It is important to note that all glycan compositions showed in the viewer are taken from the first imported profile. In other words, if the user uploads experiments where the second contains more compositions than the first, then the extra compositions will not show in the viewer. To avoid this problem, it is imperative to upload first, the profile with the largest set of compositions.

Differential display

Differential Display is an adapted version of the Individual Display for comparing two glycan profiles. In this mode, the list of experiments in the left side menu is duplicated to provide the user with the option of selecting an experiment in each of the two lists. Glynsight subtracts intensities of each pair of identical compositions. The viewer shows red and blue bars whose height for each composition represents the result of the subtraction. For any given composition, if its intensity in the first experiment is higher than in the second, then the bar is blue. In the opposite situation, the bar is red. We also introduced a distinction to spot compositions unique to an experiment. Brighter blue (resp. red) highlights differences arising from compositions unique to the first (resp. second) experiment while toned down blue (resp. red) shows differences arising from compositions expressed in both experiments but higher in the first (resp. second) than in the second. Note that white and grey bars are still present with the same meaning as previously defined.

Potential structures

To link compositions with potential glycan structures, Glynsight searches in GlyConnect, our dashboard for curated glycan data (manuscript in preparation), for all reported glycans that match a particular composition. This process is transparent to the user and is initiated by clicking on a composition of interest or its corresponding bar in the main viewer. Then, potential structures are visualized in the structure tab under the viewer. Results can be filtered by type of glycan attachment to protein, species and tissue. Each can be removed by clicking on the red cross in the corner of each image. Clicking on the image goes directly to the structure page of the database. Finally, the “Send to EpitopeXtractor” button sends the list of glycan structures to the next tool in the pipeline described in this paper. It performs the extraction of glycoepitopes/glycan determinants.

EpitopeXtractor

The part of a complex carbohydrate molecule that is recognized by a GBP is referred to as “glycoepitope” or “glycan determinant”. It is the glycan structure required for recognition by a GBP (Wang et al. 2014). Glycan determinants are uniquely recognized by antibodies or by lectins, the generic category of GBPs. Here, glycoepitope and glycan determinant are used as synonyms.

EpitopeXtractor is the second step of our pipeline as well as a standalone tool. Its purpose is to extract all the glycoepitopes

contained in a glycan structure or a set thereof. Several lists of glycan determinants have been described in the literature or listed in databases, often with overlaps, we have created a collection which summarizes the public knowledge about glycoepitopes. EpitopeXtractor is using this knowledge to extract glycoepitopes from glycan structures. A more detailed description of the sources selected to build our collection of glycoepitopes is given in the section related to the Glydin' tool.

EpitopeXtractor is powered by GlyS3 (Alocchi et al. 2015), another in house tool, designed to search glycan substructures. The branched structure of glycans complicates this task and we proposed a solution in GlyS3 taking into account possible structure fuzziness. In a nutshell, EpitopeXtractor performs substructure search, via GlyS3, for each glycoepitope in our curated list and outputs all matches.

EpitopeXtractor is available online at <http://glycoproteome.expasy.org/epextractor>. To our knowledge, it is the first online tool extracting glycoepitopes from one or more glycan structure(s) based on a multi-source and curated list of glycan determinants.

Glycan structure input

There are three ways of importing glycan structures to be processed for glycoepitope extraction:

- Draw a glycan using the graphical input using GlycanBuilder (Ceroni et al. 2007). Note that in the near future, SugarSketcher currently only prototyped (<https://glycoproteome.expasy.org/sugarsketcher/>) will supersede GlycanBuilder.
- Define a structure with the text input pasted or written in GlycoCT condensed (Herget et al. 2008). This format is widely used in glycoinformatics and many converters are available from other formats (Campbell et al. 2014).
- Insert a list of comma-separated glycan IDs from our in house database GlyConnect.

Extraction results

Running EpitopeXtractor can take several minutes due to the large amount of epitopes present in the standard collection (approximately 500 different glycoepitopes). When the extraction is completed, the results show all the SNFG coded images of the epitopes extracted from the input glycan(s). Clicking on each picture redirects the user to a tool called Glydin', a glycoepitope viewer presented in the next section.

When multiple glycans are submitted, the results are shown as a conceptual map highlighting the relation between selected glycoepitopes and input structures. This interactive map displays glycoepitopes in a central vertical list radiating to structures that contain the substructures. Hovering on a glycoepitope highlights all the structures including this particular substructure whereas moving the mouse on a structure will show all the glycoepitopes found in this selected glycan.

Visualization with Glydin'

By default EpitopeXtractor outputs a list of glycoepitopes associated with a set of glycans. However, the underlying relationships between glycoepitopes that partially share composition should also be highlighted and explored. This can be managed with Glydin', a tool we developed to map and display glycoepitope similarities. In the upper right corner of the Result window, the user can click on a button with a network icon to send all the extracted glycoepitopes to Glydin'. Then, a new browser page opens and shows the glycoepitope network of Glydin' where all the selected epitopes are singled out as red nodes.

Further detail on the purpose and usage of Glydin' is provided in the next section.

Glydin'

Glycoepitopes have been collected by different authors and recorded in different ways in a few unrelated databases or simply printed in tables of articles. In all cases, glycoepitopes are listed as independent entities despite their compositional similarity and the common set of enzymes required to synthesize them. We designed Glydin' (Glycan dynamics) to (i) merge distinct sources of information and standardize the description glycoepitopes provided in different formats and (ii) map the similarity between these entities in an interactive network. Glydin' is as a result, a tool to visualize and explore the relationships between glycoepitopes based on either their shared monosaccharide composition or the glycosyltransferase needed to get one from the other.

Glydin' is available online at <http://glycoproteome.expasy.org/epitopes/>.

Glycan determinant data sources

The extent of glycan determinants has not been yet defined. Several independent sources are available where information on glycoepitopes can be found. The following four resources cover the current knowledge on glycoepitopes:

- The article "The repertoire of glycan determinants in the human glycome" (Cummings 2009) provided a table of 122 glycoepitopes.
- The 2014 version of GlycoEpitope database (<http://www.glycoepitope.jp/>) (Kawasaki, Nakao, and Tominaga 2008) reviewed by Dr Catherine Hayes (formerly a member of the Dept of Medical Biochemistry and Cell Biology at the University of Gothenburg, Sweden) generated a list of 156 glycoepitopes.
- The SugarBind database (<https://sugarbind.expasy.org/>) (Mariethoz et al. 2016) contains 200 different glycoepitopes with described structures.
- The BiOligo database (<http://glyco3d.cermav.cnrs.fr/search.php?type=bioligo>), part of the Glyco3D portal, provides more than 250 annotated entries of glycan determinants (Pérez et al. 2015).

Preprocessing structural data

The four different sources of glycoepitopes cited above were integrated into one unique and nonredundant collection. However, the compilation of the sources was made difficult due to the variety of formats used to encode structures. Glycan determinant structures in the different sources are described with IUPAC formulae that are not standardized. Consequently, the same epitope referenced in two different sources may have two different encodings and will be considered as two different entities by an automatic text parser. Using several online converters, all glycoepitopes were first translated to GlycoCT, the most common structure encoding format. In many cases, the four sources provided names for the glycoepitopes and those with none were left as "unnamed". In the end, a list of 558 unique glycan determinants was compiled.

Data visualization

Glydin' displays the 558 glycoepitopes as nodes in a network. Two networks are mapped to emphasize two different types of shared properties:

1. In the “Substructure Based Map”, two nodes are connected if one glycoepitope is a substructure of the other.
2. In the “Enzyme Base Map”, two nodes are connected as the result of a glycosyltransferase adding a monosaccharide to the other.

Substructure network

Using GlyS3 (Alocci et al. 2015), the edge list representing the network was created. Each node was searched as a substructure in the remaining 557. Practically, when an epitope A is found to be a substructure of an epitope B, then node A and node B are connected, and an edge with the IDs of the epitopes A and B is added to the edge list. This resulted in a network where each node is connected to itself since it is a substructure of itself. Also, if a node A is connected to B and node B is connected to node C, then a redundant edge that links nodes A and C is created and overcrowds the network. Finally, to clean all the redundant edges, a transitive reduction of the graph was applied.

Enzymatic network

The enzymatic network can be represented as an evolution of the substructure network. Starting from the same entities, more constraints have been added to the substructure search. This network is rooted. Roots are single monosaccharides (Hex and HexNAc). An edge between two nodes is added only if one is a substructure of the other and if it covers a subregion that includes one of the roots. Including the root as part of the substructure is the key to shift from being a substructure to an extension. In the end, like the substructure network, a filtering step using the transitive reduction was applied to the graph.

Network properties

The “Substructure Based Map” shows a network as one large connected component whereas the “Enzyme Base Map” is a collection of connected components each stemming from one of the five main monosaccharides. In either map all nodes are labeled with their glycan determinant name. Hovering over a node highlights the glycoepitope and its neighboring nodes and shows their names. The glycoepitopes called “unnamed” were not given names in the original sources. Nodes have distinct shades of color. Glycoepitopes with the strongest shade are referenced in all four sources and appear in the foreground. Color fades as the number of source references decreases. In the background the lightest shade corresponds to nodes referenced in only one source. Upon clicking on a node, a small window that shows the corresponding glycoepitope details opens in the upper right corner and next to it, a target icon appears. Clicking on this target centers the network on the node of the glycoepitope detailed in the window. This window contains four sections. The first one shows the glycoepitope SNFG cartoon representation. The second one (“Referenced In”) specifies where it was referenced and links to the original sources. The third section (“Referenced As”) states the names that it was given in the sources, and the last one (“Related To”) contains the names of the connected glycoepitopes. Clicking on the link to a related glycoepitope will refresh the window with information of that new glycoepitope. A double click on a node highlights it and the nodes related to it while the others fade out to the background. A zoom option is also available. Zooming in when centered on a glycoepitope expands the graph around it. Above a certain zoom level, the glycoepitope names are automatically displayed for all the nodes.

Supplementary data

Supplementary data is available at *Glycobiology* online.

Funding

European Union FP7 Innovative Training Network [grant number 316,929] and by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation (SERI). ExPASy is maintained by the web team of the Swiss Institute of Bioinformatics and hosted at the Vital-IT Competency Center.

Acknowledgements

We thank Dr David Falk and Prof Manfred Wuhler for helpful discussions on the genesis of Glynsight and feedback on the early prototypes. We also acknowledge Prof Nicolle Packer and Dr Niclas Karlsson, our main collaborators in the Glycomics@ExPASy initiative.

Conflict of interest statement

None declared.

Abbreviations

CFG, consortium for functional glycomics; CSV, comma-separated value; EOC, epithelial ovarian cancer; Fuc, fucose; GBP, glycan binding protein; Hex, hexose; HexNAc, N-acetylhexosamine; mAbs, monoclonal antibodies; NeuAc, N-acetylneuraminic acid; NeuGc, N-glycolylneuraminic acid; SNFG, symbol nomenclature for glycans; SVG, scalable vector graphics.

References

- Abbott KL, Nairn AV, Hall EM, Horton MB, McDonald JF, Moremen KW, Dinulescu DM, Pierce M. 2008. Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. *Proteomics*. 8(16): 3210–3220, <https://doi.org/10.1002/pmic.200800157>.
- Adamczyk B, Jin C, Polom K, Muñoz P, Rojas-Macias MA, Zeeberg D, Borén M, Roviello F, Karlsson NG. 2018. Sample handling of gastric tissue and O-glycan alterations in paired gastric cancer and non-tumorigenic tissues. *Sci Rep*. 8(1):242, <https://doi.org/10.1038/s41598-017-18299-6>.
- Adamczyk B, Tharmalingam T, Rudd PM. 2012. Glycans as cancer biomarkers. *Biochim Biophys Acta*. 1820(9):1347–1353, <https://doi.org/10.1016/j.bbagen.2011.12.001>.
- Agravat SB, Saltz JH, Cummings RD, Smith DF. 2014. GlycoPattern: A web platform for glycan array mining. *Bioinformatics*. 30(23):3417–3418, <https://doi.org/10.1093/bioinformatics/btu559>.
- Alocci D, Mariethoz J, Horlacher O, Bolleman JT, Campbell MP, Lisacek F. 2015. Property graph vs RDF triple store: A comparison on glycan substructure search. *PLoS One*. 10(12):e0144578, <https://doi.org/10.1371/journal.pone.0144578>.
- Anugraham M, Jacob F, Nixdorf S, Everest-Dass AV, Heinzlmann-Schwarz V, Packer NH. 2014. Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: Glycan structures reflect gene expression and DNA methylation status. *Mol Cell Proteomics*. 13(9):2213–2232, <https://doi.org/10.1074/mcp.M113.037085>.
- Buehler M, Brian T, Leboucq A, Jacob F, Caduff R, Fink D, Goldstein DR, Heinzlmann-Schwarz V. 2013. Meta-analysis of microarray data identifies GAS6 expression as an independent predictor of poor survival in ovarian cancer. *BioMed Res Int*. 2013:238284, <https://doi.org/10.1155/2013/238284>.
- Bénard J, Da Silva J, De Blois MC, Boyer P, Duvillard P, Chiric E, Riou G. 1985. Characterization of a human ovarian adenocarcinoma line, IGROV1, in tissue culture and in nude mice. *Cancer Res*. 45(10): 4970–4979.

- Campbell MP, Ranzinger R, Lütke T, Mariethoz J, Hayes CA, Zhang J, Akune Y, Aoki-Kinoshita KF, Damerell D, Carta G et al. 2014. Toolboxes for a standardised and systematic study of glycans. *BMC Bioinf.* 15(Suppl 1):S9, <https://doi.org/10.1186/1471-2105-15-S1-S9>.
- Ceroni A, Dell A, Haslam SM. 2007. The GlycanBuilder: A fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code Biol Med.* 2(August):3, <https://doi.org/10.1186/1751-0473-2-3>.
- Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. 2008. GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res.* 7(4):1650–1659, <https://doi.org/10.1021/pr7008252>.
- Cholleti SR, Agravat S, Morris T, Saltz JH, Song X, Cummings RD, Smith DF. 2012. Automated motif discovery from glycan array data. *OMICS J Integr Biol.* 16(10):497–512, <https://doi.org/10.1089/omi.2012.0013>.
- Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. 2014. Cell surface protein glycosylation in cancer. *Proteomics.* 14(4–5): 525–546, <https://doi.org/10.1002/pmic.201300387>.
- Cummings RD. 2009. The repertoire of glycan determinants in the human glycome. *Mol BioSyst.* 5(10):1087–1104, <https://doi.org/10.1039/b907931a>.
- Cymer F, Beck H, Rohde A, Reusch D. 2017. Therapeutic monoclonal antibody N-glycosylation—Structure, function and therapeutic potential. *Biologicals.* November. <https://doi.org/10.1016/j.biologicals.2017.11.001>.
- Escrevente C, Machado E, Brito C, Reis CA, Stoeck A, Runz S, Marmé A, Altevogt P, Costa J. 2006. Different expression levels of alpha3/4 fucosyltransferases and lewis determinants in ovarian carcinoma tissues and cell lines. *Int J Oncol.* 29(3):557–566.
- Gudelj I, Keser T, Vučković F, Škaro V, Goreta SŠ, Pavić T, Dumić J, Primorac D, Lauc G, Gornik O. 2015. Estimation of human age using N-glycan profiles from bloodstains. *Int J Legal Med.* 129(5):955–961, <https://doi.org/10.1007/s00414-015-1162-x>.
- Hayes CA, Nemes S, Issa S, Jin C, Karlsson NG. 2012. Glycomic work-flow for analysis of mucin O-linked oligosaccharides. In: McGuckin M, Thornton D, editors. *Mucins, Methods in Molecular Biology.* Humana Press. p. 141–163, https://doi.org/10.1007/978-1-61779-513-8_8.
- Herget S, Ranzinger R, Maass K, Lieth C-WVD. 2008. GlycoCT—A unifying sequence format for carbohydrates. *Carbohydr Res.* 343(12):2162–2171, <https://doi.org/10.1016/j.carres.2008.03.011>.
- Ismail MN, Stone EL, Panico M, Lee SH, Luu Y, Ramirez K, Ho SB, Fukuda M, Marth JD, Haslam SM et al. 2011. High-sensitivity O-glycomic analysis of mice deficient in core 2 [beta]1,6-N-acetylglucosaminyltransferases. *Glycobiology.* 21(1):82–98, <https://doi.org/10.1093/glycob/cwq134>.
- Jansen BC, Falck D, de Haan N, Ederveen ALH, Razdrov G, Lauc G, Wührer M. 2016. LaCyTools: A targeted liquid chromatography-mass spectrometry data processing package for relative quantitation of glycopeptides. *J Proteome Res.* 15(7):2198–2210, <https://doi.org/10.1021/acs.jproteome.6b00171>.
- Jansen BC, Karli RR, Bondt A, Ederveen ALH, Palmblad M, Falck D, Wührer M. 2015. MassyTools: A high-throughput targeted data processing tool for relative quantitation and quality control developed for glycomic and glycoproteomic MALDI-MS. *J Proteome Res.* 14(12):5088–5098, <https://doi.org/10.1021/acs.jproteome.5b00658>.
- Jin C, Diarmuid TK, Skoog EC, Padra M, Adamczyk B, Vitizeva V, Thorell A, Venkatakrisnan V, Lindén SK, Karlsson NG. 2017. *Structural Diversity of Human Gastric Mucin Glycans.* Molecular & Cellular Proteomics: MCP, March. <https://doi.org/10.1074/mcp.M117.067983>.
- Joshi HJ, von der Lieth C-W, Packer NH, Wilkins MR. 2010. GlycoViewer: A tool for visual summary and comparative analysis of the glycome. *Nucleic Acids Research* 38 (Web Server issue). W667–W670, <https://doi.org/10.1093/nar/gkq446>.
- Kawasaki T, Nakao H, Tominaga T. 2008. GlycoEpitope: A database of carbohydrate epitopes and antibodies. In: *Experimental Glycoscience.* Tokyo: Springer. p. 429–431, https://doi.org/10.1007/978-4-431-77922-3_104.
- Kemna MJ, Plomp R, van Paassen P, Koeleman CAM, Jansen BC, Damoiseaux JGMC, Cohen Tervaert JW, Wührer M. 2017. Galactosylation and sialylation levels of IgG predict relapse in patients with PR3-ANCA associated vasculitis. *EBioMedicine.* 17(January):108–118, <https://doi.org/10.1016/j.ebiom.2017.01.033>.
- Keser T, Gornik O, Vučković F, Selak N, Pavić T, Lukić E, Gudelj I, Gašparović H, Biočina B, Tilin T et al. 2017. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia.* 60(12):2352–2360, <https://doi.org/10.1007/s00125-017-4426-9>.
- Knezevic A, Gornik O, Polasek O, Pucic M, Redzic I, Novokmet M, Rudd PM, Wright AF, Campbell H, Rudan I et al. 2010. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology.* 20(8):959–969, <https://doi.org/10.1093/glycob/cwq051>.
- Knezevic A, Polasek O, Gornik O, Rudan I, Campbell H, Hayward C, Wright A, Kolcic I, O'Donoghue N, Bones J et al. 2009. Variability, heritability and environmental determinants of human plasma N-glycome. *J Proteome Res.* 8(2):694–701, <https://doi.org/10.1021/pr800737u>.
- Konishi Y, Aoki-Kinoshita KF. 2012. The GlycomeAtlas tool for visualizing and querying glycome data. *Bioinformatics.* 28(21):2849–2850, <https://doi.org/10.1093/bioinformatics/bts516>.
- Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, Pučić-Baković M, Novokmet M, Mangino M, Thaqi K et al. 2014. Glycans are a novel biomarker of chronological and biological ages. *J Gerontol A Biol Sci Med Sci.* 69(7):779–789, <https://doi.org/10.1093/geronola/glt190>.
- Kurbacher CM, Korn C, Dixel S, Schween U, Kurbacher JA, Reichelt R, Arenz PN. 2011. Isolation and culture of ovarian cancer cells and cell lines. *Methods Mol Biol.* 731:161–180, https://doi.org/10.1007/978-1-61779-080-5_15.
- Mariethoz J, Khatib K, Alocci D, Campbell MP, Karlsson NG, Packer NH, Mullen EH, Lisacek F. 2016. SugarBindDB, a resource of glycan-mediated host-pathogen interactions. *Nucleic Acids Res.* 44(D1):D1243–D1250, <https://doi.org/10.1093/nar/gkv1247>.
- Miwa HE, Song Y, Alvarez R, Cummings RD, Stanley P. 2012. The bisecting GlcNAc in cell growth control and tumor progression. *Glycoconjugate J.* 29(8–9):609–618, <https://doi.org/10.1007/s10719-012-9373-6>.
- Nakagawa H, Kawamura Y, Kato K, Shimada I, Arata Y, Takahashi N. 1995. Identification of neutral and sialyl N-linked oligosaccharide structures from human serum glycoproteins using three kinds of high-performance liquid chromatography. *Anal Biochem.* 226(1):130–138, <https://doi.org/10.1006/abio.1995.1200>.
- North SJ, von Gunten S, Antonopoulos A, Trollope A, MacGlashan DW, Jang-Lee J, Dell A, Metcalfe DD, Kirshenbaum AS, Bochner BS et al. 2012. Glycomic analysis of human mast cells, eosinophils and basophils. *Glycobiology.* 22(1):12–22, <https://doi.org/10.1093/glycob/cwr089>.
- Planinc A, Dejaegher B, Heyden YV, Viaene J, Praet SV, Rappez F, Antwerpen PV, Delporte C. 2017. Batch-to-batch N-glycosylation study of infliximab, trastuzumab and bevacizumab, and stability study of bevacizumab. *Eur J Hosp Pharm.* 24(5):286–292, <https://doi.org/10.1136/ejpharm-2016-001022>.
- Pučić M, Knezevic A, Vidic J, Adamczyk B, Novokmet M, Polasek O, Gornik O, Supraha-Goreta S, Wormald MR, Redzic I et al. 2011. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics.* 10(10): M111.010090, <https://doi.org/10.1074/mcp.M111.010090>.
- Pérez S, Sarkar A, Rivet A, Breton C, Imbert A. 2015. Glyco3D: A portal for structural glycosciences. In: *Glycoinformatics, 241–58. Methods in Molecular Biology.* New York, NY: Humana Press, https://doi.org/10.1007/978-1-4939-2343-4_18.
- Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. 2005. Glycomics: An integrated systems approach to structure-function relationships of glycans. *Nat Methods.* 2(11):817–824, <https://doi.org/10.1038/nmeth807>.
- Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S, Sasisekharan R. 2006. Advancing glycomics: Implementation strategies at the consortium for functional glycomics. *Glycobiology.* 16(5):82R–90R, <https://doi.org/10.1093/glycob/cwj080>.
- Rooijen J. J. van, Jeschke U, Kamerling JP, Vliegthart JF. 1998. Expression of N-linked sialyl Le(x) determinants and O-Glycans in the carbohydrate moiety of human amniotic fluid transferrin during pregnancy. *Glycobiology.* 8(11): 1053–1064.

- Ruhaak LR, Deelder AM, Wuhrer M. 2009. Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry. *Anal Bioanal Chem.* 394(1):163–174, <https://doi.org/10.1007/s00216-009-2664-5>.
- Ruhaak LR, Miyamoto S, Lebrilla CB. 2013. Developments in the identification of glycan biomarkers for the detection of cancer. *Mol Cell Proteomics.* 12(4):846–855, <https://doi.org/10.1074/mcp.R112.026799>.
- Ruhaak LR, Zauner G, Huhn C, Bruggink C, Deelder AM, Wuhrer M. 2010. Glycan labeling strategies and their use in identification and quantification. *Anal Bioanal Chem.* 397(8):3457–3481, <https://doi.org/10.1007/s00216-010-3532-z>.
- Schnaar RL. 2016. Glycobiology simplified: Diverse roles of glycan recognition in inflammation. *J Leukocyte Biol.* 99(6):825–838, <https://doi.org/10.1189/jlb.3RI0116-021R>.
- Stadlmann J, Taubenschmid J, Wenzel D, Gattlinger A, Dürnberger G, Dusberger F, Elling U, Mach L, Mechtler K, Penninger JM. 2017. Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. *Nature.* 549(7673):538, <https://doi.org/10.1038/nature24015>.
- Stroop CJ, Weber W, Gerwig GJ, Nimtz M, Kamerling JP, Vliegenthart JF. 2000. Characterization of the carbohydrate chains of the secreted form of the human epidermal growth factor receptor. *Glycobiology.* 10(9):901–917.
- Taniguchi N, Kizuka Y. 2015. Glycans and cancer: Role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv Cancer Res.* 126:11–51, <https://doi.org/10.1016/bs.acr.2014.11.001>.
- Wang Y-C, Peterson SE, Loring JF. 2014. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* 24(2):143, <https://doi.org/10.1038/cr.2013.151>.