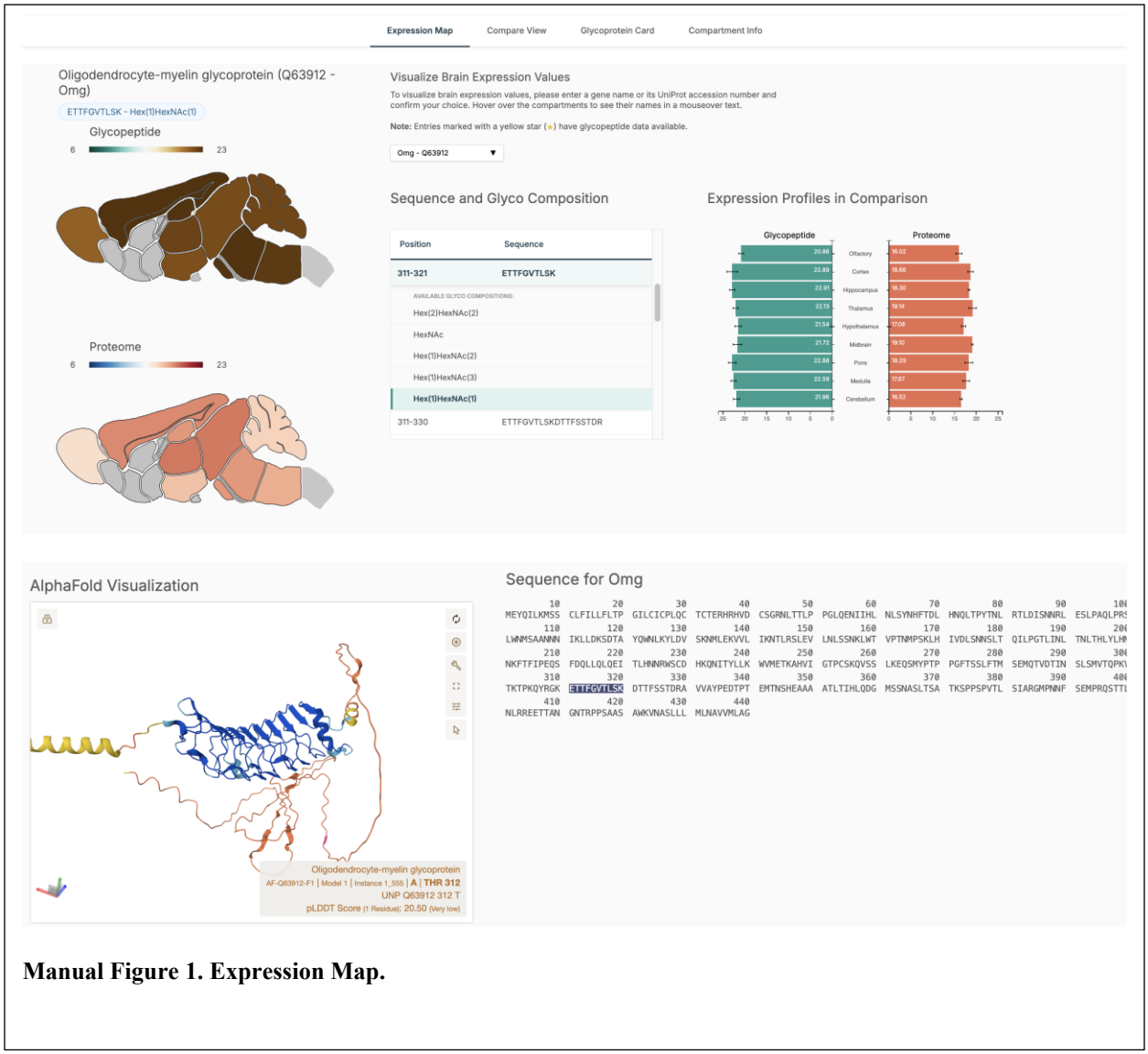
GlycoDB.org Manual

GlycoDB.org is organized into four analytical modules: *Expression Map*, *Compare View*, *Glycoprotein Card*, and *Compartment Info*. Together, these interfaces provide complementary perspectives on the dataset. Protein-centered views enable detailed exploration of individual proteins, glycopeptides, glycosylation sites, and glycan compositions, whereas region-centered views facilitate investigation of anatomical compartment-specific expression patterns and direct comparison between brain regions.

Expression Map

The *Expression Map* view provides a protein centered interface for exploration of the quantitative spatial O-glycoproteomics dataset. This module enables simultaneous visualization of glycopeptide- and protein-level abundance across nine anatomically defined mouse brain regions, allowing users to investigate how individual O-GalNAc glycosylation events are distributed throughout the brain. Users can search for their protein of interest (either by gene name or UniProt accession) to visualize glycopeptide and corresponding protein level abundance profiles. The interface displays the regional distribution of selected glycopeptides using both a color-coded schematic representation of the brain regions as well as quantitative bar charts, enabling direct comparison between glycopeptide abundance and overall protein expression.

To begin, enter a gene name or UniProt accession in the search field located at the top center of the page (for example, OMG) see Manual Figure 1 (Fig. M1). Once a protein is selected, all identified glycopeptides derived from that protein are listed in the "*Sequence and Glyco Composition*" table below the search field. Selecting a peptide sequence (for example, *ETTFGVTLSK*) displays all detected O-GalNAc glycan compositions associated with that peptide. Users can then choose a specific O-glycan composition, such as Hex(1)HexNAc(1) (1T), for further inspection.



Manual Figure 1. Expression Map.

Brain Expression Maps:

The upper-left panels show regional expression patterns across the investigated brain compartments: Upper brain schematic: abundance of the selected glycopeptide (designated – *Glycopeptide* (Fig. M1)); Lower brain schematic: corresponding protein abundance (designated – *Proteome* (Fig. M1))

Expression levels are represented using a color scale. Moving the cursor over a brain region displays a tooltip with the anatomical compartment name and the corresponding quantitative expression value.

Expression Profiles in Comparison:

The panel on the right displays comparative bar plots for all brain regions: Green bars: abundance of the selected glycopeptide of chosen glycoform; Orange bars: protein-level abundance. The range of the error bars is accessible via the tooltip that is shown while hovering over the respective colored bar. This side-by-side visualization allows rapid comparison of glycopeptide specific expression with total protein abundance across brain regions.

Sequence and Glyco Composition*:*

The *"Sequence and Glyco Composition"* table is the main interactive component of the Expression Map view. It lists all detected glycopeptides together with their sequence positions and associated glycan compositions. Selecting an entry updates all visualizations simultaneously, including: Glycopeptide brain expression maps; Quantitative glycopeptide abundance plots and Protein sequence highlighting.

Sequence for selected protein:

The lower-right panel displays the full amino acid sequence of the selected protein (e.g. "*Sequence of Omg*") using ProSeqViewer[1]. The currently selected glycopeptide is highlighted within the sequence, allowing users to identify its exact position within the protein.

The lower-left panel integrates the interactive three-dimensional structure of the protein directly from AlphaFold[2, 3]. The structure can be rotated using the mouse, and hovering over residues displays their amino acid positions. This provides structural context for the identified glycosylation sites.

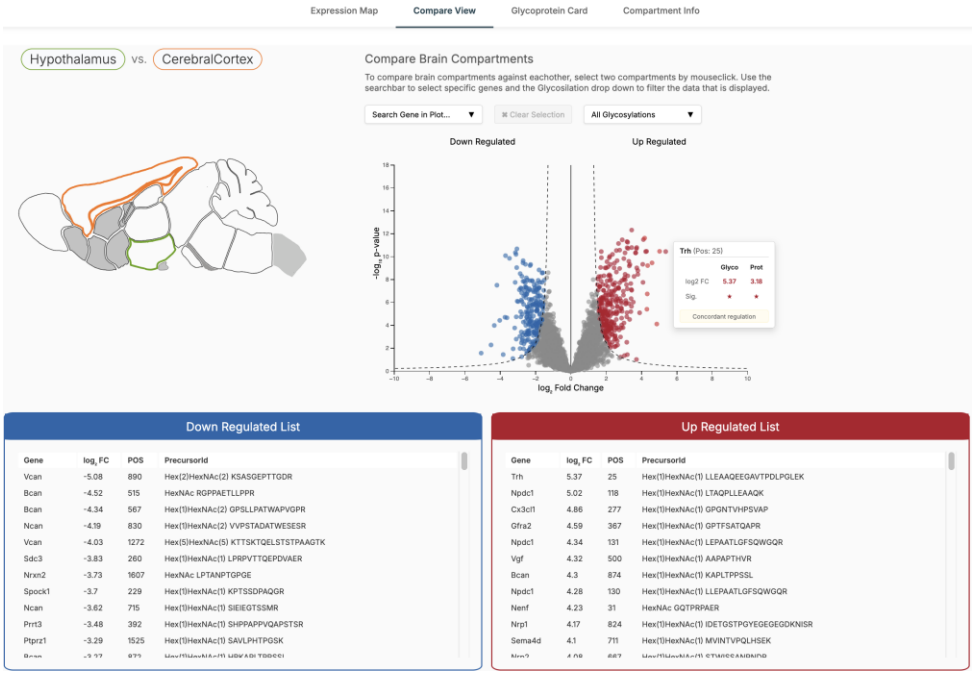
Compare View.

Compare view enables pairwise regional analysis of glycopeptide regulation between user-selected regions. In this view, two manually selected compartments on the brain schematic are highlighted and contrasted in an interactive volcano plot, allowing rapid identification of glycopeptides enriched in one region relative to the other. The comparison can be performed globally across all quantified glycopeptides or restricted to selected glycan classes. In addition, the view supports targeted interrogation of individual genes (proteins). When a gene of interest is selected, the corresponding glycopeptides are highlighted within the differential comparison, allowing users to inspect how multiple site- or glycoform-resolved features from the same protein behave across the

selected brain-region pair. Regulated features are additionally summarized in separate upregulated and downregulated tables below the plot, providing a structured entry point for downstream inspection.

***Part I: Global Compare view*:**

To begin, select *"Compare View"* from the upper navigation menu. By default, two random brain compartments are selected automatically. On the schematic mouse brain displayed on the left side of the page, users can select their brain compartment of choice by clicking on it. The selected regions will be highlighted on the brain schematic, and the comparison label (for example, *Hypothalamus vs Cerebral Cortex*) will appear above the diagram (Fig. M2A).



Manual Figure 2A. Global Compare View.

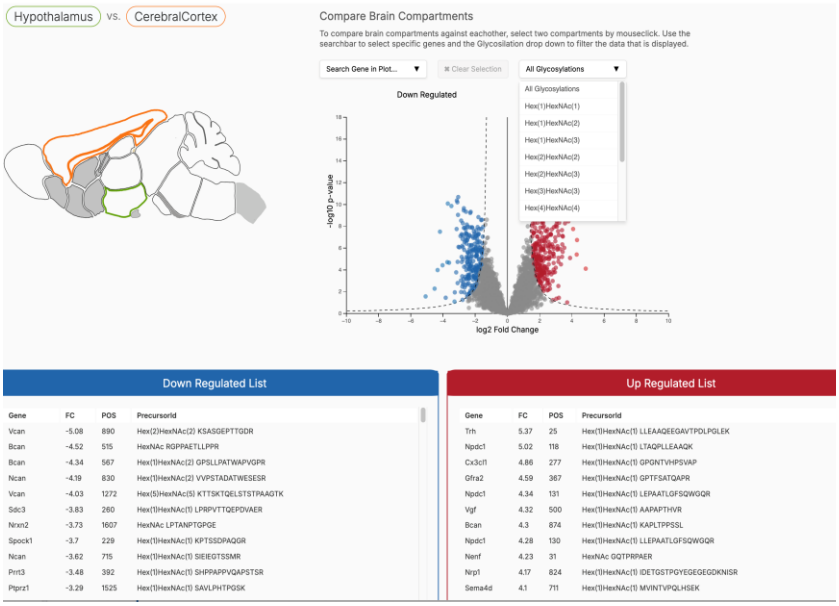
Volcano Plot Visualization.

The central panel displays O-glycopeptide abundance differences between the two selected brain regions as a volcano plot (Fig. M2A). Each dot represents a single O-glycopeptide. Red dots indicate glycopeptides significantly enriched in the first selected region (in this example, hypothalamus), blue dots indicate glycopeptides significantly enriched in the second selected region (cerebral cortex) and grey dots represent glycopeptides without statistically significant

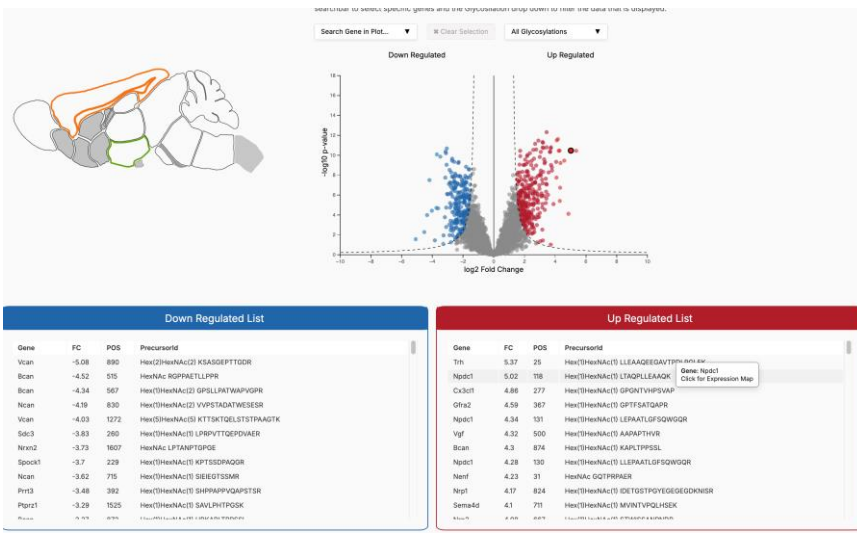
differences between the two regions. Moving the mouse cursor over an individual dot displays a window with detailed quantitative information, including (Fig. M2A): Protein or gene name (for example, TRH); N-terminal position of the glycopeptide within the protein sequence (Pos: 25); Glycopeptide fold change between regions (Log2FC – Glyco); Corresponding protein fold change (Log2FC – Prot.). Statistical significance (Sig.) indicators for both glycopeptide and protein measurements (Fig. M2A). Filled red and blue stars indicate statistically significant upregulation and downregulation, respectively, whereas unfilled stars indicate non-significant differences. In addition, the tooltip window displays a note indicating whether the observed regulation is concordant, discordant or glyco-specific.

Differentially Regulated Glycopeptide Tables

Below the volcano plot, two interactive tables summarize significantly regulated O-glycopeptides (Fig. M2A): Left table (blue): Down-Regulated List; Right table (red): Up-Regulated List. Each table contains the following columns: *Gene* - gene name (gene symbol); *FC* - fold-change values displayed on a Log2 scale; *POS* - amino acid position corresponding to the N-terminal residue of the glycopeptide within the protein sequence; *PrecursorID* - glycan composition together with the glycopeptide sequence. For example: *Hex(1)HexNAc(1) LLEAAGEEGAVTPDLPGLEK* (derived from TRH). By default, the tables display all significantly regulated O-GalNAc glycopeptides ranked from highest to lowest fold change.



Manual Figure 2B. Compare View – Selection of O-glycan.



Manual Figure 2C. Global Compare View – Selecting O-glycopeptide.

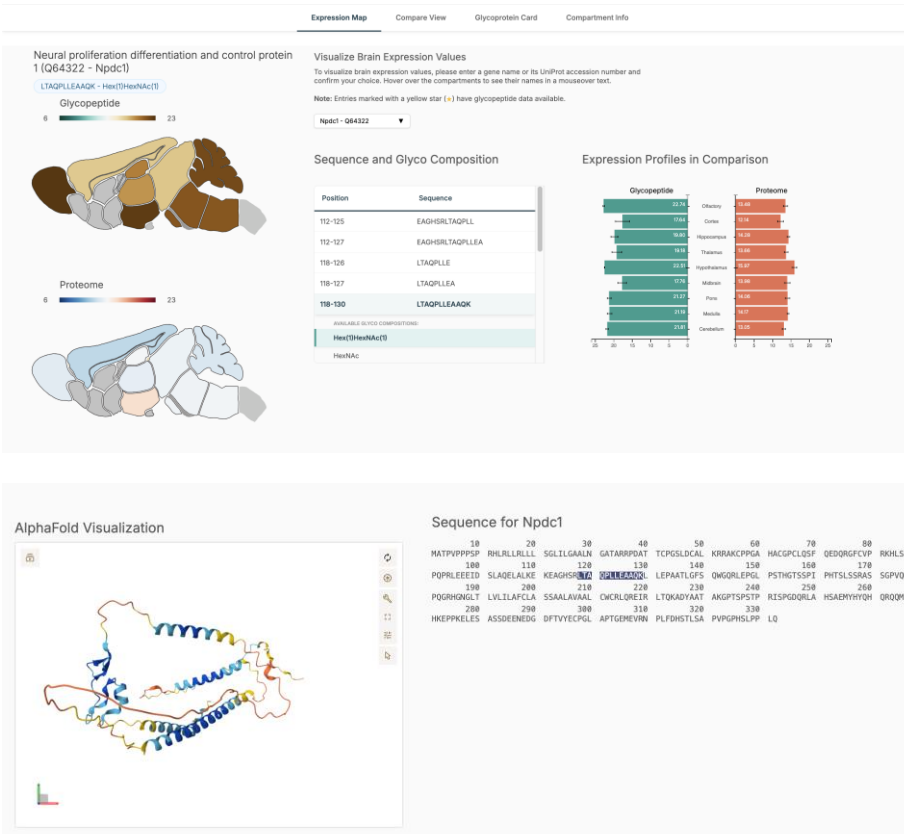
Filtering by Glycan Composition.

By default, the *Compare View* displays all detected O-glycan compositions and glycopeptide sequences. However, users can restrict the analysis to a specific glycan composition using the *"All Glycosylations"* selection panel located in the upper-right corner of the interface (Fig. M2B). Selecting a specific glycan composition filters the volcano plot and associated tables to display only glycopeptides carrying the selected glycan structure.

Interactive Navigation to Expression Map View.

The volcano plot and result tables are fully interactive. Hovering over a glycopeptide sequence within either table highlights and enlarges the corresponding dot in the volcano plot (for example NPDC1 on Fig. M2C). An additional navigation icon will then appear, allowing the user to open the selected glycopeptide in the *Expression Map View*.

Clicking the highlighted glycopeptide redirects the browser to the *Expression Map* interface for the corresponding protein and selected glycopeptide (Fig. M2D). As described in the *Expression Map* chapter, users can then explore the regional abundance distribution, glycan composition, sequence localization, and structural context of the selected glycopeptide. Users may also switch between other glycopeptides identified for the same protein directly within the *Expression Map* interface (see *Expression Map* chapter).



Manual Figure 2D. From Compare View to Expression Map.

Part II: Protein selective Compare view:

The Protein-Selective *Compare View* enables detailed investigation of changes in O-glycopeptide composition for a specific protein across two selected brain regions. In contrast to the global comparison view, which displays all detected glycopeptides, this mode focuses exclusively on glycosylation changes associated with a single protein of interest.

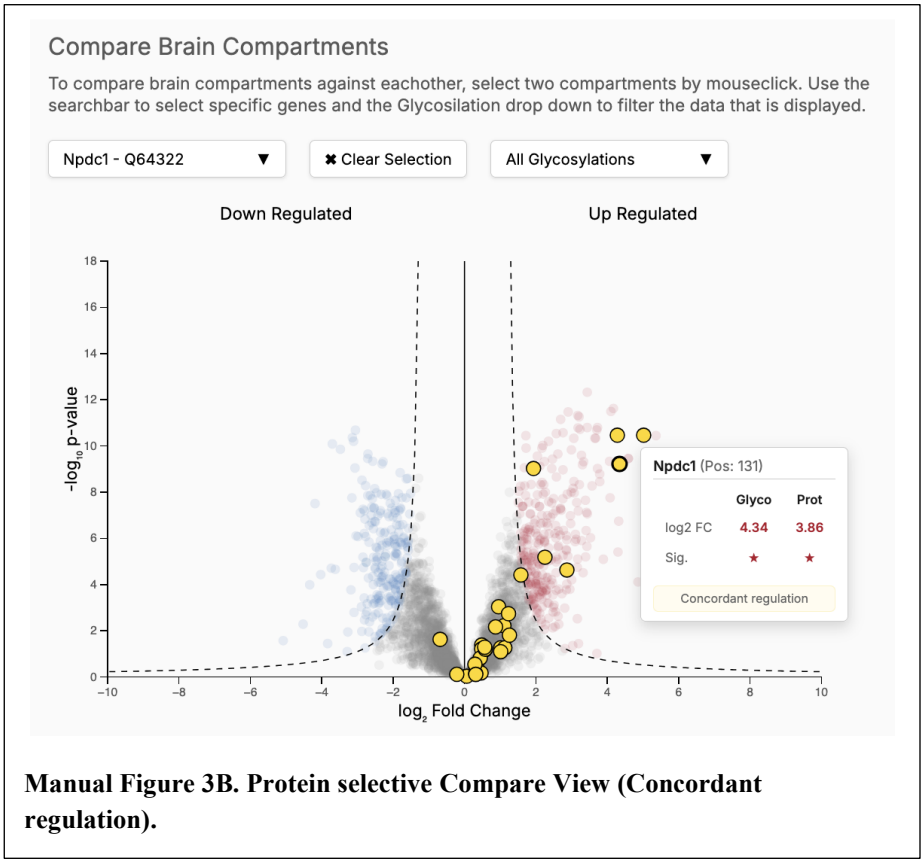
Manual Figure 3A. Protein selective Compare View.

To begin, enter a gene name or UniProt accession in the search field labeled *"Search Gene in Plot…"* located above the volcano plot. As an example, type Npdc1 (Fig. M3A). Next, select the Hypothalamus vs Cerebral Cortex comparison using the brain schematic, as described in Part I of the manual.

After filtering for NPDC1, only glycopeptides derived from this protein remain highlighted, while all unrelated glycopeptides are visually de-emphasized (Fig. M3A). In this example, six O-

glycopeptides exhibit statistically significant enrichment in the hypothalamus relative to the cerebral cortex, each with a Log2 fold change of approximately 2 or greater (Fig. M3A). No significantly downregulated NPDC1 glycopeptides are observed in this comparison.

As described previously, interactive summary tables below the volcano plot display all significantly regulated glycopeptides identified for the selected protein (Fig. M3A).



Manual Figure 3B. Protein selective Compare View (Concordant regulation).

A major feature of the *Compare View* (that was shown before but not emphasized) is the simultaneous visualization of glycopeptide and proteome level of regulation. Hovering over an individual data point in the volcano plot opens an annotation window displaying both glycopeptide and corresponding protein fold changes together with statistical significance information (Fig. M3B).

This enables users to distinguish between multiple biological regulation scenarios, including:

a) Concordant regulation – glycopeptide abundance changes that closely parallel overall protein abundance changes, indicating that O-glycosylation levels primarily reflect protein regulation;

b) Glyco-specific and protein-specific regulation – glycopeptide regulation that is substantially stronger or weaker than the corresponding protein-level change, suggesting selective regulation of O-GalNAc glycosylation in addition to, or independent from, changes in total protein abundance;

c) Discordant regulation – situations in which protein abundance changes in one direction while the corresponding O-glycopeptide changes in the opposite direction. For example, protein abundance may increase whereas the associated O-glycopeptide decreases, or vice versa.

In the NPDC1 example, concordant regulation of both protein abundance and O-glycopeptide expression is observed, with statistically significant upregulation in the hypothalamus relative to the cerebral cortex (Fig. M3B).

Additional Filtering and Navigation Options

In the *Compare View*, users can restrict the analysis to a specific glycan composition using the *"All Glycosylations"* selection panel located in the upper-right corner of the interface.

Users may also investigate additional proteins by entering a different gene name in the *"Search Gene in Plot…"* field. Clicking *"Clear Selection"* removes the protein specific filter and returns the interface to the Global Compare View.

Together, the *Compare View* provides a powerful framework for identifying region-specific O-GalNAc glycosylation patterns across the mouse brain. Different modules enable users to explore differential glycopeptide regulation at both the global dataset level and the level of individual proteins, while simultaneously integrating glycopeptide- and protein-level quantitative information. The interactive volcano plots, filtering options, and direct links to the *Expression Map* facilitate detailed investigation of region-enriched glycosylation events and their biological context.

However, because individual proteins can contain multiple glycopeptides, overlapping peptide sequences, and several glycan compositions, comparative visualizations may become increasingly complex for heavily glycosylated proteins. While the *Compare View* is highly effective for identifying statistically regulated features, it is less suited for obtaining a rapid and intuitive overview of the complete glycosylation architecture of a protein.

To address this challenge, GlycoDB.org includes the *Glycoprotein Card* module, which provides a compact, protein-centered summary of glycosylation features across brain regions. The Glycoprotein Card integrates quantitative glycopeptide information into an intuitive pseudoheatmap representation, allowing users to rapidly assess glycosylation occupancy, regional distribution, and glycoform diversity within a single unified view. This representation

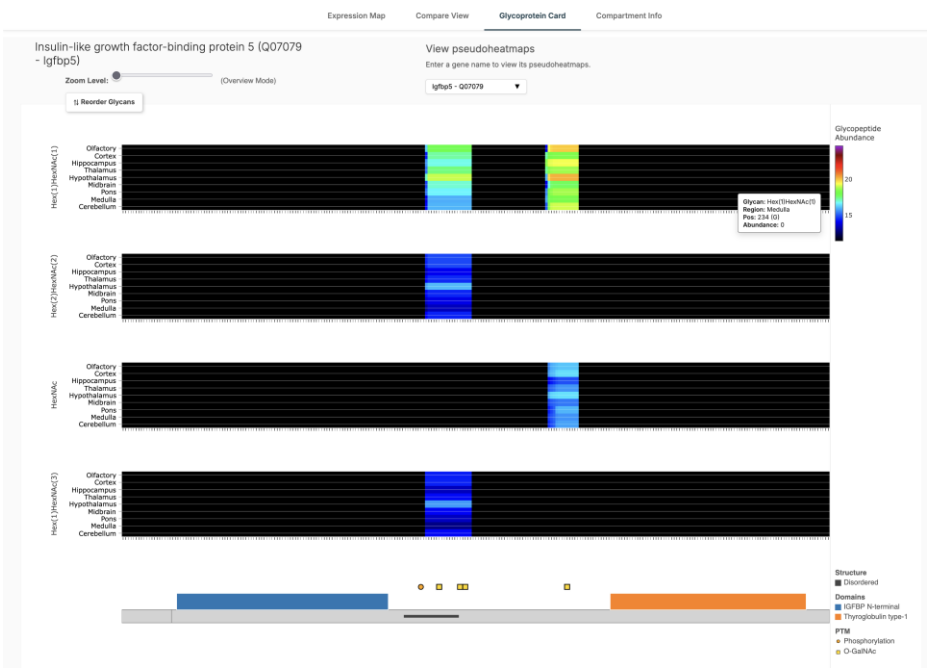
complements the detailed analytical capabilities of the Compare and Expression Map modules by offering a simplified overview of protein-specific glycosylation landscapes.

Glycoprotein Card

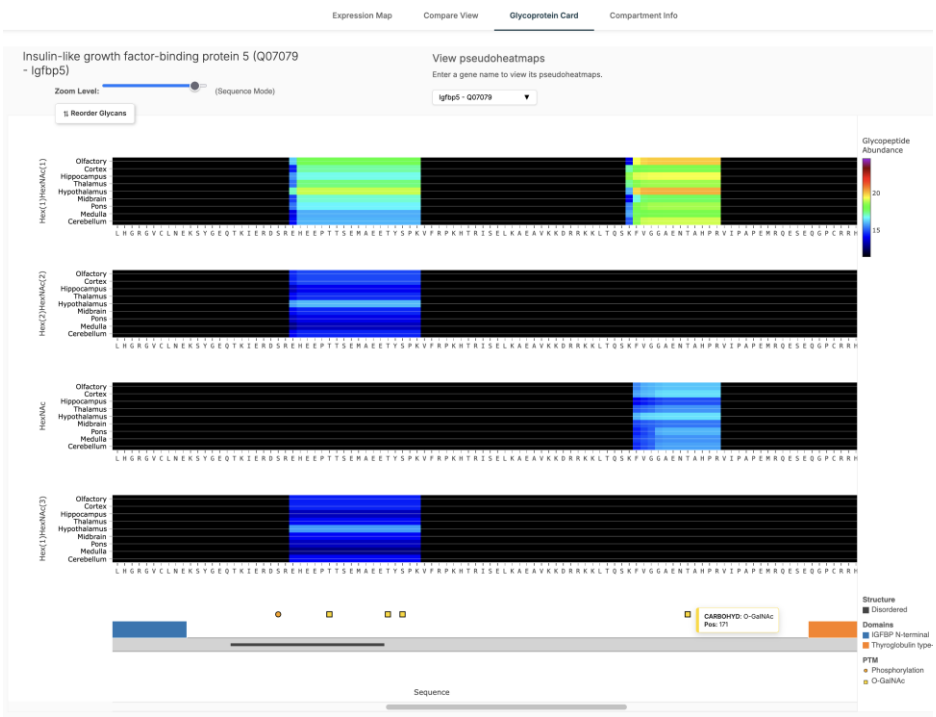
***Glycoprotein card condenses glycopeptide complexity into a pseudoheatmap view*.**

To provide a more compact and interpretable overview of protein-associated O-glycosylation, we implemented a *Glycoprotein Card* view that summarizes all detected glycopeptides for a selected protein in the form of pseudoheatmaps. This representation was designed to reduce the complexity arising from proteolytic digestion specificity, overlapping peptide backbones, and multiple glycan compositions, which can make data at glycopeptide-level difficult to interpret at a glance. In this view, the x-axis represents the linear protein sequence, whereas each row corresponds to one of the nine anatomically resolved brain compartments.

Similarly, to the Expression Map and Compare View, the user can select a protein of interest to display from the selection window at the top of the page by typing the gene name or selecting from the full list of available entities. Glycopeptide abundance is projected onto the protein sequence and visualized as a color-coded pseudoheatmap, with separate panels shown for individual glycan compositions (Fig. M4). Beneath the pseudoheatmaps, the protein sequence (if view allows) is displayed together with a schematic representation of protein domains, annotated phosphorylation



Manual Figure 4. Glycoprotein card.



Manual Figure 5. Glycoprotein card (zoomed in)

site (yellow circles), and mapped O-glycosylation sites (yellow rectangles) (Fig. M4). The interface further supports zooming and reordering of the pseudoheatmap panels by glycan (left upper corner), enabling users to inspect local regions of interest or reorganize the display according to glycoform patterns.

At higher zoom levels, individual amino acid residues and their corresponding modification sites can be examined in detail (Fig. M5). In addition, hovering the cursor over a modification in the schematic representation displays the exact position of the corresponding amino acid residue within the protein sequence (Fig. M5). This protein-centric representation

provides an alternative to both the raw *Expression Map* and repeated pairwise compartment comparisons. Rather than inspecting individual glycopeptides one by one or navigating a large number of regional volcano plots, users can directly assess whether glycosylation signals concentrate in particular sequence regions, whether similar patterns are observed across compartments, and whether distinct glycan classes localize to overlapping or separate parts of the same protein. In this way, the Glycoprotein Card facilitates rapid recognition of spatially patterned glycosylation features at the level of whole proteins (Fig. M4).

For example, the VGF (main manuscript Fig. 2) or IGFBP5 (Fig. M4, M5) Glycoprotein Cards reveal several discrete sequence regions with O-GalNAc signal, together with differences in their abundance across brain compartments and across glycan classes. Because all mapped glycopeptides are integrated into a common sequence-based layout, the view makes it easier to identify recurrent glycosylation hotspots and region-selective patterns that would be more difficult to extract from the raw peptide table alone. Thus, the pseudoheatmap view complements the detailed glycopeptide-centric and pairwise comparison views by providing a concise, protein-level summary of spatial O-GalNAc glycosylation.

Note: In the Glycoprotein Card pseudoheatmap window, the protein sequence is displayed by default from the N-terminus to the C-terminus. For large proteins, such as Neurexin 2 (NRXN2), where O-glycosylation sites are located toward the C-terminal region, the full protein sequence does not fit within the visible window. As a result, the displayed region appears empty of O-glycosylation modifications. To inspect O-glycosylation sites and regional differences located further toward the C-terminus, scroll down within the heatmap window to access the horizontal scrolling option. Use this control to move along the protein sequence toward the C-terminal region and visualize the corresponding O-glycosylation differences.

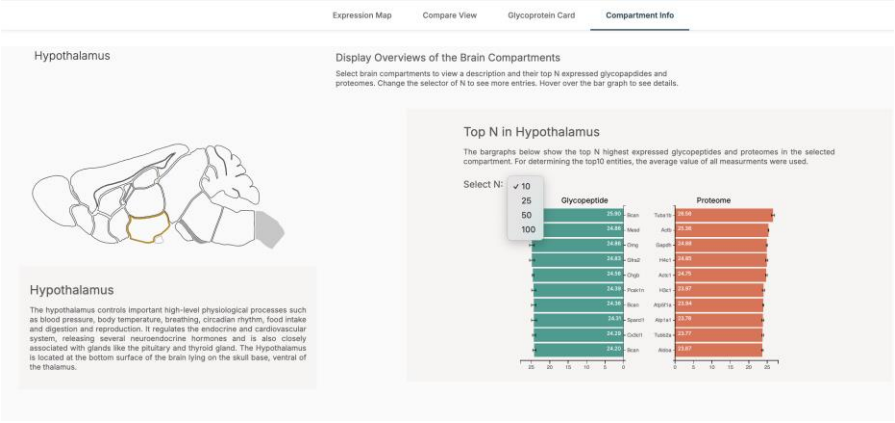
Compartment Info

The *Compartment Info* tab provides an overview of each anatomical brain region included in the spatial O-glycoproteomics atlas. It combines neuroanatomical context, biological background information, and quantitative summaries of the most abundant glycopeptides and proteins detected within a selected brain compartment.

The left side of the interface displays a schematic mouse brain map highlighting the anatomical region. The chosen compartment is outlined in mustard color (Fig. M6).

The right side of the interface presents the *Top-N in* Compartment overview, which displays the most abundant glycopeptides and proteins identified within the selected brain region. By default, the top 10 entries are shown, although users can expand the display to the top 25, 50, or 100 entries using the *Select N* dropdown menu (Fig. M6).

The *Glycopeptide* panel displays the highest-abundance O-glycopeptides detected in the selected compartment. Entries are ranked according to their average abundance across biological replicates, from the strongest to the lowest O-glycopeptide level. For each entry, the corresponding gene name and quantitative abundance value are shown, while error bars indicate the variability observed across biological replicates. Hovering over an individual bar reveals additional information, including the detected O-glycan composition, glycopeptide sequence, associated gene name, mean abundance value, and the observed abundance range within the selected compartment.



Manual Figure 6. Compartment Info.

The *Proteome* panel provides an analogous overview of the most abundant proteins quantified in the compartment. Proteins are ranked according to their average abundance across biological replicates, with error bars representing replicate variability. Hovering over a protein bar displays additional details, including the gene name, mean abundance value, and the observed abundance range.

Together, these visualizations provide a rapid overview of the molecular composition of each brain compartment, allowing users to compare highly abundant glycopeptides and proteins, identify

compartment-specific molecular signatures, and generate hypotheses regarding region-specific O-GalNAc glycosylation.

Conclusion.

GlycoDB.org was developed to make spatial O-glycoproteomics accessible, intuitive, and biologically interpretable. By integrating the *Compartment Info***,** *Expression View***,** *Compare View***,** and *Glycoprotein Card* interfaces, the platform enables exploration of O-GalNAc glycosylation of multiple complementary perspectives, from global neuroanatomical patterns to individual glycosites and glycoforms. The combination of quantitative visualization, differential analysis, structural mapping, and protein-centered summaries provides a comprehensive framework for investigating how glycosylation contributes to molecular specialization of the brain.

Beyond serving as a companion resource for the present study, GlycoDB.org establishes a foundation for future spatial glycoproteomics atlases and provides the community with a powerful platform for hypothesis generation, target prioritization, and mechanistic studies of glycosylation-dependent regulation in the nervous system.

References

1. Bevilacqua, M., Paladin, L., Tosatto, S. C. E. & Piovesan, D. ProSeqViewer: an interactive, responsive and efficient TypeScript library for visualization of sequences and alignments in web applications. *Bioinformatics* **38**, 1129-1130 https://doi.org/10.1093/bioinformatics/btab764 (2022).
2. Jumper, J.*, et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 https://doi.org/10.1038/s41586-021-03819-2 (2021).
3. Bertoni, D.*, et al.* AlphaFold Protein Structure Database 2025: a redesigned interface and updated structural coverage. *Nucleic Acids Res* **54**, D358-D362 https://doi.org/10.1093/nar/gkaf1226 (2026).