



ARTICLE

DOI: 10.1038/s41467-018-06950-3

OPEN

Systems glycomics of adult zebrafish identifies organ-specific sialylation and glycosylation patterns

Nao Yamakawa^{1,2}, Jorick Vanbeselaere¹, Lan-Yi Chang^{1,3}, Shin-Yi Yu¹, Lucie Ducrocq¹, Anne Harduin-Lepers ¹, Junichi Kurata⁴, Kiyoko F. Aoki-Kinoshita⁴, Chihiro Sato², Kay-Hooi Khoo³, Ken Kitajima² & Yann Guerardel ¹

The emergence of zebrafish *Danio rerio* as a versatile model organism provides the unique opportunity to monitor the functions of glycosylation throughout vertebrate embryogenesis, providing insights into human diseases caused by glycosylation defects. Using a combination of chemical modifications, enzymatic digestion and mass spectrometry analyses, we establish here the precise glycomic profiles of eight individual zebrafish organs and demonstrate that the protein glycosylation and glycosphingolipid expression patterns exhibits exquisite specificity. Concomitant expression screening of a wide array of enzymes involved in the synthesis and transfer of sialic acids shows that the presence of organ-specific sialylation motifs correlates with the localized activity of the corresponding glycan biosynthesis pathways. These findings provide a basis for the rational design of zebrafish lines expressing desired glycosylation profiles.

¹Université de Lille, CNRS, UMR 8576 - UGSF—Unité de Glycobiologie Structurale et Fonctionnelle, F- 59000 Lille, France. ²Bioscience and Biotechnology Center, Nagoya University, Nagoya 464-8601, Japan. ³Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan. ⁴Faculty of Science and Engineering, Soka University, Hachioji, Tokyo 192-8577, Japan. These authors contributed equally: Nao Yamakawa, Jorick Vanbeselaere. Correspondence and requests for materials should be addressed to Y.G. (email: yann.guerardel@univ-lille.fr)

First known as a home aquarium fish, zebrafish has been rapidly transformed by scientists into a versatile vertebrate model system to study a wide range of fundamental biological processes in vertebrates, including reproduction, development, evolutionary sciences, genetics, and neurobiology^{1–4}. Indeed, the availability of thousands of mutant variants allows *in vivo* dissection of developmental processes at single cell and molecular resolution. Furthermore, zebrafish exhibits numerous experimental advantages among which are short generation time, high fecundity rate, optic transparency, and fast generation time of functional organs. In addition to be an excellent platform for the study of fundamental biology, zebrafish emerged as a valuable system for modeling human disease⁵, including cardiovascular diseases^{6,7}, cancer onset and progression^{8–10}, muscular dystrophies¹¹, skeletal displasias¹², central nervous system disorders¹³, and inflammatory bowel disease¹⁴.

About 1–2% of the vertebrate's genome is believed to be involved in the synthesis and processing of glycans that play roles in most, if not all, biological processes. Of uttermost importance, the sialic acids pathway in higher vertebrates requires a large panel of enzymes with various subcellular localizations including the nuclear cytidine monophosphate (CMP)-5-*N*-acetylneuraminic acid (Neu5Ac) synthase (CMAS), the cytosolic UDP-*N*-acetylglucosamine (GlcNAc) 2-epimerase/*N*-acetylmannosamine kinase (GNE), the cytosolic cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH), the Golgi CMP-Neu5Ac transporter (SLC35A1), the Golgi sialyltransferases (ST), and sialidases (Neu)¹⁵. Thus, it comes as no surprise that any deregulation of glycan synthesis, trafficking, turn-over, or localization may induce variable clinical effects. In this context, zebrafish has recently been used not only to model but also to validate a wide variety of glycan-based diseases related to central nervous system development, cartilage formation, muscular dystrophies, skeletal disorders, neuromuscular transmission defects, neurocutaneous disorder, and inflammatory bowel disease among many others^{16–23}. In parallel, the development of innovative methods for imaging glycans *in vivo* during zebrafish organogenesis provided new insights into the biosynthesis and dynamics of glycans^{24–26}.

Despite the rush in using zebrafish to study glycosylation-driven Human pathologies, the current lack of a systems wide spatial glycomic view of zebrafish adult tissues and their determinants hinders a more fruitful functional dissection of the underlying molecular mechanisms. So far, only a restricted panel of glycans from the glycomes of zebrafish embryos and oocytes have ever been structurally defined^{27–31}. These earlier studies established that zebrafish has the potential of synthesizing not only human-like glyco-epitopes, such as Lewis-x, Gal β 1-4(Fuca1-3)GlcNAc, mono and di-sialylated T-antigens, \pm Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-6)GalNAc, but also species-specific glycan motifs, such as the Gal β 1-4(Neu5Ac/Glc α 2-3)Gal β 1-4(Fuca1-3)GlcNAc glyco-epitope and Fuca1-3GalNAc β 1-4 (Neu5Ac/Glc α 2-3)Gal β 1-3GalNAc O-glycans. However, no information on the glycosylation process in the adults was provided. Importantly, zebrafish shows a higher number of paralogous genes compared to Human, as exemplified by the ST (30 in zebrafish vs. 20 in Human)^{32,33}, due to either species-specific duplications of genes (e.g. ST3Gal1A, ST3Gal1B, ST3Gal1C, and ST3Gal1D)³⁴ or third round genome duplications that occurred in teleosts (e.g. ST6Gal2-r or ST3Gal3-r)^{15,34}, while some human paralogues are lost (e.g. ST3Gal6)³⁴.

To provide the requisite structural ground for better-guided correlation of paralogous gene activities and future genetic manipulation in functional analysis of zebrafish glycans, we have initiated a systematic organ-specific profiling of glycoconjugates and the expression patterns of enzymes involved in the terminal

glycosylation in adult zebrafish. Among key features, we identified a highest level of sialylation in intestine and ovary that is dominated by 2-keto-3-deoxynononic acid (Kdn) and 5-*N*-glycolyl neuraminic acid (Neu5Gc), respectively, along with a diverse range of zebrafish-specific terminal glycotopes, O-glycans, and glycosphingolipids (GSLs).

Results

Purification of glycans and GSLs. Eight organs, namely skin, liver, gill, brain, intestines, heart, testis, and ovaries were collected from three separate batches of 30 zebrafishes each and immediately processed according to standard procedure to isolate N-linked glycans (NGs), O-linked glycans (OGs), and GSLs^{28,29}. Released glycans were identified and semi-quantified by MS and MS/MS analyses. In parallel, we determined the sialic acid ratios within each of the three families of glycans from different organs, assessed the expression patterns of relevant glyco-genes, and established the tissue-specific expression of selected epitopes. The overall analytical workflow was summarized in Fig. 1.

Sialic acids composition of organs and glycans. Three different sialic acid species, Neu5Ac, Neu5Gc, and Kdn, were identified from zebrafish glycoconjugates. As shown in Fig. 2a, the quantity of total sialic acids observed in individual tissues was extremely variable, ranging from 0.6 to 7.8 nmol/mg of proteins for six organs out of eight and up to 22.1 and 41.1 nmol/mg for intestine and ovary. These values are consistent with those reported previously for fish gametes and testis (2–4 nmol/mg proteins), mice tissues (0.4–1.1 nmol/mg), and mouse melanoma cells (14 nmol/mg)^{35–37}. Not only the quantity but also the nature of sialic acids exhibited a wide organ-specific variability. The relative quantifications of sialic acids associated with glycans demonstrated similar overall patterns but differed from that observed in total tissues, as shown in Fig. 2b, c. Altogether, these observations confirm that sialic acid synthesis and transfer are differentially regulated in the different organs. The most striking feature of these organ-specific sialylation pattern is the overwhelming presence of Kdn in all glycoconjugates isolated from intestine (from 60% to 85%), with a more modest occurrence in liver, heart, gills, and skin. Another unusual feature that was further explored in the course of present study concerns the modest but relevant content of Neu5Gc in brain glycoconjugates (1–9%). Indeed, brain is considered to be largely devoid of Neu5Gc³⁸, although its presence has been unambiguously established in the brain glycoconjugates of horse³⁹.

Glycomic analysis and cognate sialylation enzymes. Altogether, we detected 249 NGs, 69 OGs, and 286 GSLs in the eight tissues, some of them being found in several organs while others were organ-specific. We confidently identified 100 different NGs including 95 complex and hybrid NGs (NG1–NG95, Supplementary Data 1), 33 OGs (OG1–OG33, Supplementary Data 2), and 172 GSLs (GL1–GL172, Supplementary Data 3). Out of the 172 GSLs identified, we delineated 52 glycan moieties (G1–G52, Supplementary Data 4) that could each be conjugated to several different ceramide anchors. The structural assignment of all glycans was mostly based on the specific CID MS/MS fragmentation patterns afforded by the $[M + Na]^+$ molecular ions of permethylated glycans in positive mode, as previously described and detailed in Supplementary Notes 1–4 section^{28,29,40}. All the identified NGs, OGs, and GSLs were included in Glycome Atlas that is freely available at <http://www.rings.t.soka.ac.jp/GlycomeAtlasV5/index.html>⁴¹. Each NG, OG, and GSL was converted to a LinearCode-like format and then incorporated into the Glycome Atlas database, which was augmented to include a

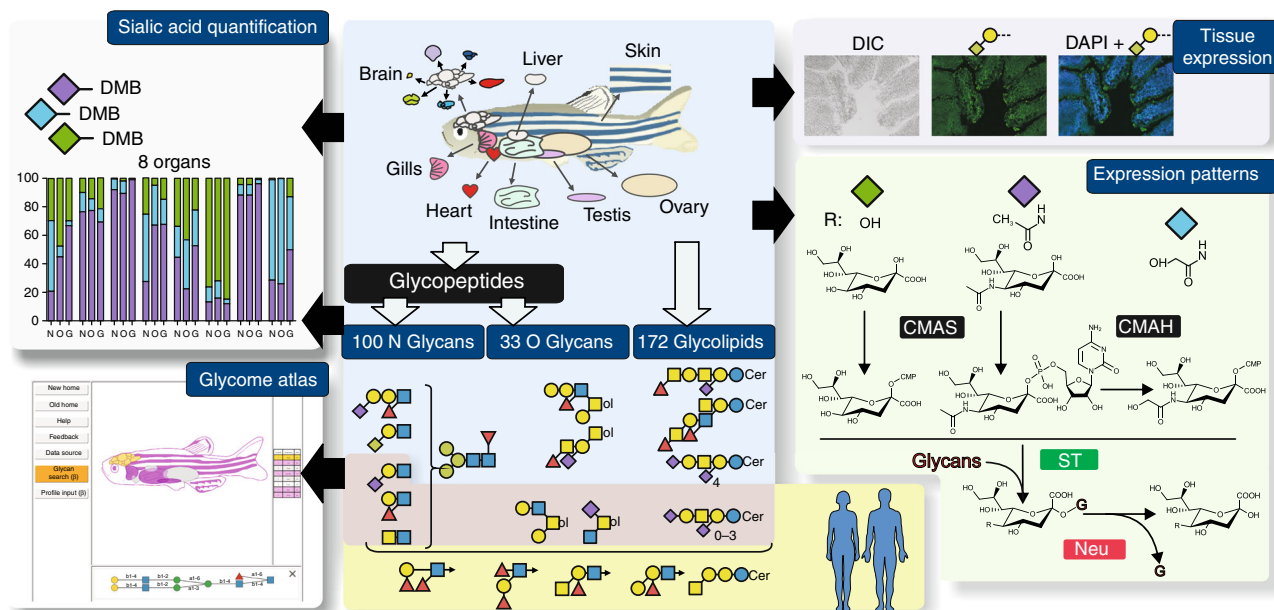


Fig. 1 Analytical workflow. NGs, OGs, and GSLs were purified from single batches of freshly dissected tissues before analysis of sialic acids, monosaccharides, and oligosaccharide sequences. Unique glycan structures identified and their spatial distribution were posted in an interactive database, Glycome Atlas (<http://rings.t.soka.ac.jp/GlycomeAtlasV5/index.html>), to provide easy access to the zebrafish glycome. In parallel, the expression pattern of enzymes involved in sialic acid metabolism was established, along with the tissue expression of specific sialylated epitopes. Zebrafish glycome was shown to express many human-type glyco-epitopes, such as type-2 Gal β 1-4GlcNAc β 1-based sialyl LacNAc and Le^x, simple sialylated O-glycans and major gangliosides, but differed by carrying additional glycosyl extensions and the absence of some blood group antigens (A, B), type-1 Gal β 1-3GlcNAc β 1-based epitopes, and globo-serie GSLs. CMAS CMP-Sia synthase, CMAH CMP-Sia hydroxylase, ST sialyltransferase, and Neu neuraminidase. Graphical representation is based on accepted conventions for glycans and monosaccharide nomenclature as follows: yellow circle, Gal; yellow square, GalNAc; blue circle, Glc; blue square, GlcNAc; green circle, Man; red triangle, Fuc; purple diamond, Neu5Ac; light blue diamond, Neu5Gc; green diamond, Kdn^{77,78}. The interglycosidic bonds between monosaccharides of antennae are represented using the conventional positions as in |, C2 position; /, C3 position; —, C4 position; \, C6 position

user interface for zebrafish. The original LinearCode format did not handle base monosaccharides, such as Hexose or HexNAc, so these were represented as Z and ZN, respectively. This zebrafish data including list of tissues was then added to the Glycome Atlas database. Then the zebrafish image was converted to SVG format so that specific tissues could be clicked to display the corresponding list of NGs, OGs, and GSLs. To use the updated GlycomeAtlas version 5, three buttons are available at the top for human, mouse, and now zebrafish. By clicking on zebrafish, an image of the zebrafish with organs outlined and a table of the eight organs for which glycans have been characterized will be displayed. By clicking on either the image or the table, the selected organ and row in the table will be highlighted in yellow, and the list of glycans profiled will be shown on the right. When a single glycan is clicked in this area, it will be highlighted and its detailed structure will be displayed below the image. Clicking again on another glycan will add the glycan to the selection list below the image. At the same time, the other organs in which the selected glycan(s) are found will also be highlighted in both the image and the table. This allows users to see the tissue specificity of the glycan(s) selected. Right-clicking on a selected glycan will display a pop-up menu for three options: (1) to copy the LinearCode[®] format for the glycan, which can be used for Glycan Search (available in the menu on the left), (2) open a window to the CFG if the glycan data is available there, or (3) open a window to GlyTouCan, if the glycan data is available⁴². The selected glycans can be deselected by clicking the X on the upper right.

Structural analysis of NGs by combined MS and MS/MS led to identification of 100 NGs out of which 9 were hybrid types and 86 complex types, as described in Supplementary Note 2. The

proposed structures and their occurrences were summarized in Supplementary Data 1 and their MS spectra collated in Supplementary Figure 1. The five oligomannosylated NGs assigned to ubiquitous signals at m/z 1580, 1784, 1988, 2192, and 2396 are the so-called Man-5–Man-9 high mannose (Man) type NGs. MS/MS analyses did not reveal any obvious deviation from the canonical structures and thus these five NGs will not be discussed further.

A comparison of the MALDI-MS spectra of NGs isolated from the eight organs established an unexpected level of tissue specificity from both qualitative and quantitative point of views (Supplementary Figure 1 and Data 1). It shows that the proportions of complex NGs versus oligomannosylated NGs are highly variable, ranging from 5% in ovary up to 66% in gills (Supplementary Figure 2). The very low proportion of complex NGs observed in ovary is highly reminiscent of what was observed in embryos²⁸. Individual complex and hybrid NGs were relatively quantified by integrating their corresponding MALDI-MS signals from three independent experiments and expressed as % total of complex and hybrid NGs within each sample. The relative abundance of 95 individual complex and hybrid type NGs is summarized as a heat map in Fig. 3 and presented with more details in Supplementary Figure 3. Within each organ, the detected structural heterogeneity varies from 9 distinct NGs in liver up to 46 in gills (Fig. 4a). Only three were identified in all the eight organs: NG89, NG90, and NG91 (Fig. 3, Supplementary Data 1). These are bi-antennary non-core-fucosylated structures carrying the previously defined unique zebrafish epitope Gal β 1-4 (Neu5Ac/Gc(α 2-3))[Gal β 1-4(Fuca1-3)GlcNAc²⁸ on both antennae, sialylated by Neu5Ac and Neu5Gc in 2:0 (NG89), 1:1

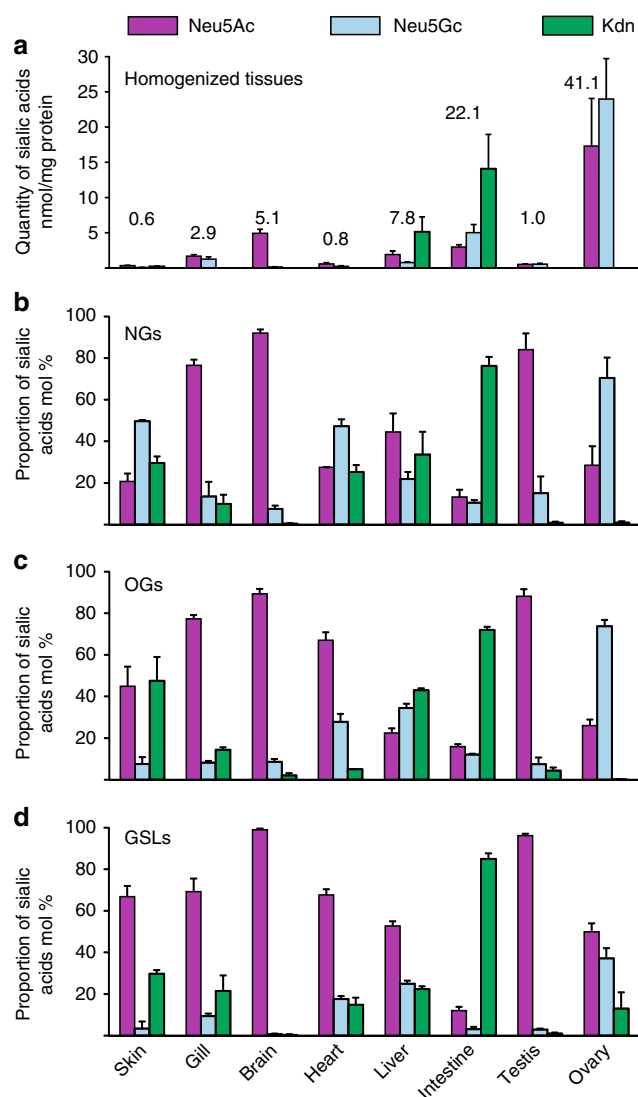


Fig. 2 Sialic acids composition. Individual sialic acids were identified and quantified by NP-HPLC-FL and LC-MS analysis of (a) total tissues, as well as purified (b) NGs, (c) OGs, and (d) GSLs. In total tissues (a), quantities were expressed as nmol of sialic acids per mg of proteins. In purified glycans (b–d), proportions of individual sialic acids were expressed in mol% because the yield of glycan purification from total tissues cannot be precisely established. Purple bar, Neu5Ac; blue bars, Neu5Gc; green bars, Kdn. Data are shown as mean \pm SE of at least three experiments

(NG90), and 0:2 (NG91) ratios, respectively. In contrast, the core fucosylated equivalent of NG89 (NG92) could only be identified in low amount in the skin.

Irrespective of the tissue origins, a vast majority of the identified NGs are fucosylated. However, the distribution of fucose between core fucose and Le^x-type is more variable. In most tissues, 40–60% of NGs are core-fucosylated and 50–70% carry a Le^x moiety but core fucosylation is notably overrepresented in brain (Fig. 4b). As suggested by the quantification of sialic acids in tissues and NGs (Fig. 2a, b), sialylation patterns of NGs also show important tissue variability both in terms of the degree and chemical nature of sialylation (Fig. 4c). It should be noted that the proportions of sialylated glycans as determined from MALDI-MS signal intensities do not always match the quantification data of sialic acids (Fig. 4c vs. Fig. 2a) because (1) the latter does not take into account individual intensities, (2) the nature of sialic acids affects the detected signal intensities of glycans in MALDI-MS,

and (3) MS analysis would bias against and thus underestimate the larger, potentially highly sialylated glycans with less favorable ionization, and detector response factor.

The vast majority of identified sialic acids, irrespective of their nature, was attached via α 2-3 linkage to internal Gal residue of the \pm Gal β 1-4(Neu5Ac/Glc α 2-3)Gal β 1-4(\pm Fuca1-3)GlcNAc motif. This unique sialylated motif was defined by MS/MS analysis and confirmed by GC/MS linkage analysis of total NG fraction that demonstrated the presence of C3-substituted, C4-substituted, and C3, C4-disubstituted Gal residues, whereas C6-substituted Gal residues could not be detected. However, we cannot preclude the possibility that minor Sia α 2-6Gal-containing N-Glycans may exist since our MS/MS data could not discriminate α 2-3- from α 2-6-sialylation on galactose (Gal). Hanzawa and collaborators have previously demonstrated that α 2-6-substituted N-glycans were the predominantly expressed sialylated N-glycans in devoloped embryos³¹. In accordance with their study, we have detected α 2-6 sialylation on the N-acetylgalactosamine (GalNAc) residue of LacdiNAc motifs in NG15, 24, 58, 59 that were exclusively observed in ovary in our study. However, they also showed that (1) adult-derived yolk does not contain α 2-6-linked sialic acid, and (2) α 2-6 sialylation decreases during embryonic development from 6 to 48 hpf, whereas α 2-3 sialylation increases in the later developmental stages. These observations are consistent with our current data and collectively suggest a developmental shift from embryonic α 2-6 to adulthood α 2-3-sialylation along with an overall significantly down-regulated sialylation to either absent or very minor level in all adult organs (NG20 and NG60). Contrary to Neu5Ac and Neu5Gc, Kdn was only found on the Gal residue of Gal(β 1,4)GlcNAc but never detected on the additionally galactosylated or fucosylated terminal disaccharide unit. At present, the sialyl linkage of Kdn to Gal could not be unambiguously established due to the low amount of Kdn-substituted glycans.

Altogether, our structural analyses established that the terminal glycosylation units decorating the zebrafish N-glycome are almost entirely based on the type 2 LacNAc (Gal β 1-4GlcNAc) antennae differentially substituted by Fuca1-3 and Sia α 2-3 on the GlcNAc and Gal residues, respectively, to yield the Le^x and SLe^x epitopes commonly expressed on mammalian glycoproteins. However, it can also be further substituted by a terminal Gal β 1-4 residue that has never been observed in Human and other mammals, to the best of our knowledge. In addition, a number of GalNAc β 1-3/4GlcNAc antennae that can be further sialylated on the terminal GalNAc residues were identified in several organs, with a clear prevalence in ovary, testis, and skin. Although the β 1-4 version of this unit, also called LacdiNAc, is abundant in invertebrates^{43,44}, it is less commonly found in mammalian species including Humans^{45,46}, where it is considered as a potential diagnosis marker for a number of cancers^{47,48}.

Thirty-three different mucin type OGs from eight organs could be identified by MS (Supplementary Figure 4) and MS/MS analyses of permethylated derivatives, as described in Supplementary Note 3. Their structures and relative quantifications are summarized in Fig. 5, Supplementary Figure 5, and Supplementary Data 2. These analyses did not lead to identification of other types of O-glycans, such as O-Man, O-Fuc, or O-Glc that may however be present at lower quantities. The major O-glycans are based on core-1 Gal(β 1-3)GalNAc-ol and core-2 Gal β 1-3(GlcNAc β 1-6)GalNAc-ol structures that are widely distributed among mammals including Humans. Few core-3 GlcNAc β 1-3GalNAc-ol and core-4 GlcNAc β 1-3(GlcNAc β 1-6)GalNAc-ol-based structures were additionally detected as very minor components. Interestingly, zebrafish exhibited some core extension and terminal glycosylation specificities never observed in Humans including those based on the GalNAc β 1-4Gal β 1-3GalNAc-ol sequence that

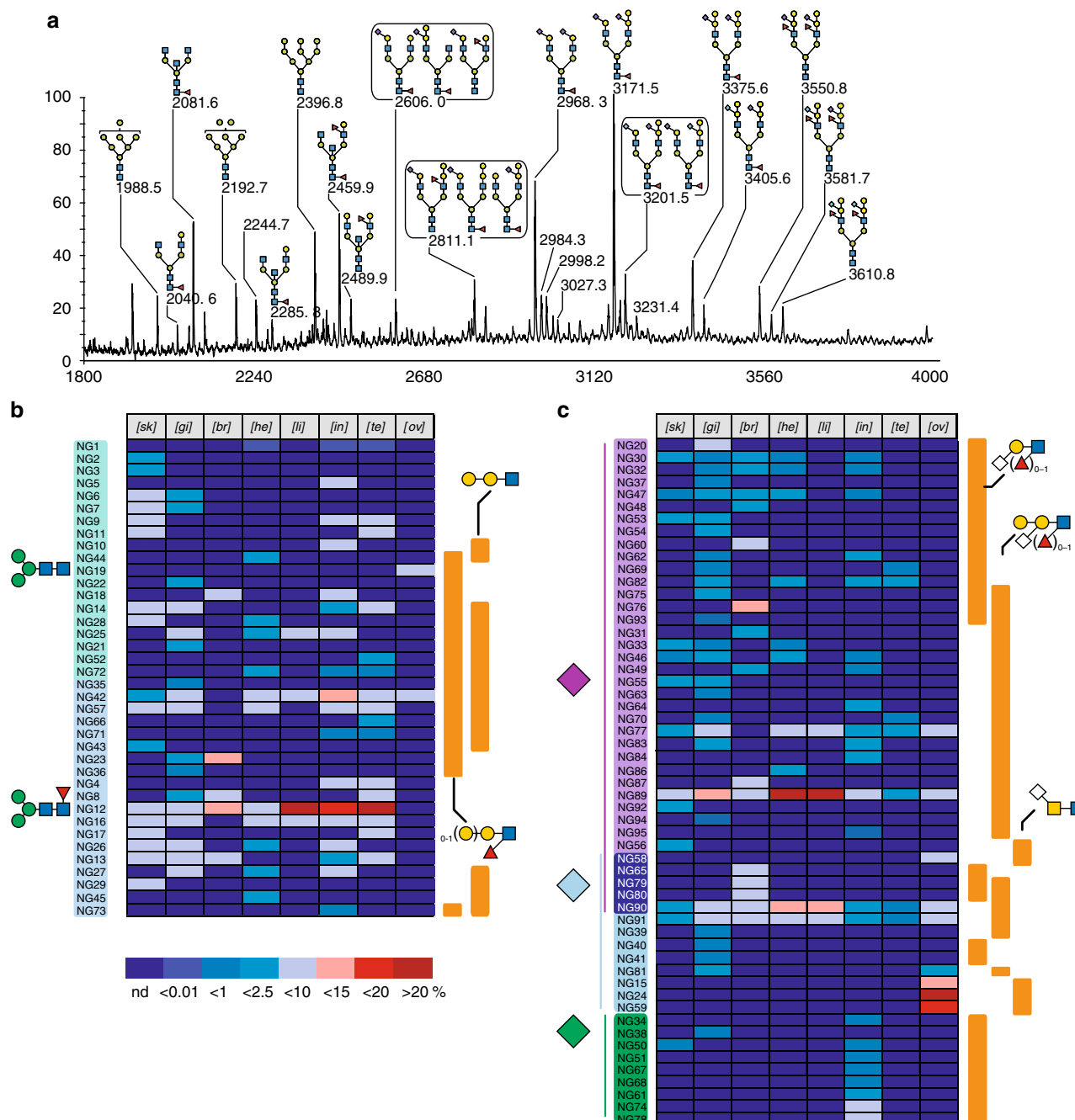


Fig. 3 N-glycome profiles. The chemical nature of individual glycans was established by a combination of MS/MS, GC/MS, and NMR, in comparison with previous work^{28,29} and reported in Supplementary Figure 1 and Supplementary Data 1. **a** Representative MALDI-TOF MS spectra of permethylated NGs isolated from brain. All MS spectra are provided in Supplementary Figure 1. **b** Tissues-specific distribution and relative quantification of neutral and **(c)** sialylated (right panel) NGs according to following glycan epitopes: core fucosylation, (Gal β 1-4)₀₋₁Gal β 1-4(Fuc α 1-3)GlcNAc β , Gal β 1-4Gal β 1-4GlcNAc β for neutral and nature of substituting sialic acids (Neu5Ac, Neu5Gc, Kdn), Sia α 2-3Gal β 1-4(Fuc α 1-3)₀₋₁GlcNAc β , Gal β 1-4(Sia α 2-3)Gal β 1,4(Fuc α 1-3)₀₋₁GlcNAc β , and Sia α 2-6GalNAc β 1,3/4GalNAc β . All Sia residues (Neu5Ac, Neu5Gc, Kdn) linked to Gal were considered as α 2,3 linked although minor α 2,6-Gal may co-exist as described in the text. Relative quantifications of individual compounds were extracted from the dataset presented in Supplementary Figure 3

was very abundant in testis and ovary, and the Gal β 1-4Gal β 1-4 (Fuc α 1-3)GlcNAc sequence that is reminiscent of the one observed in N-glycans.

As for NGs, OGs exhibited a high degree of tissue specificity (Fig. 5, Supplementary Figure 5, Supplementary Data 2). The level of fucosylation varied from 20% in brain to 80% in ovary, which suggests a tight organ-specific regulation of the fucosylation process. In accordance with the total sialic acid quantification,

Neu5Ac appears to be the most widely distributed sialic acid among OGs in all tissues except ovary in which Neu5Gc predominates (Figs. 2c, 5b, c). Kdn-sialylated glycans could only be firmly identified by MS and MS/MS analyses in intestine as five core-1 OGs (OG1, OG10, OG11, OG12, and OG13, Fig. 5b). Despite the relatively lower heterogeneity of O-glycosylation compared to N-glycosylation, we could not identify any glycan that is common to all organs. The most widely distributed OGs

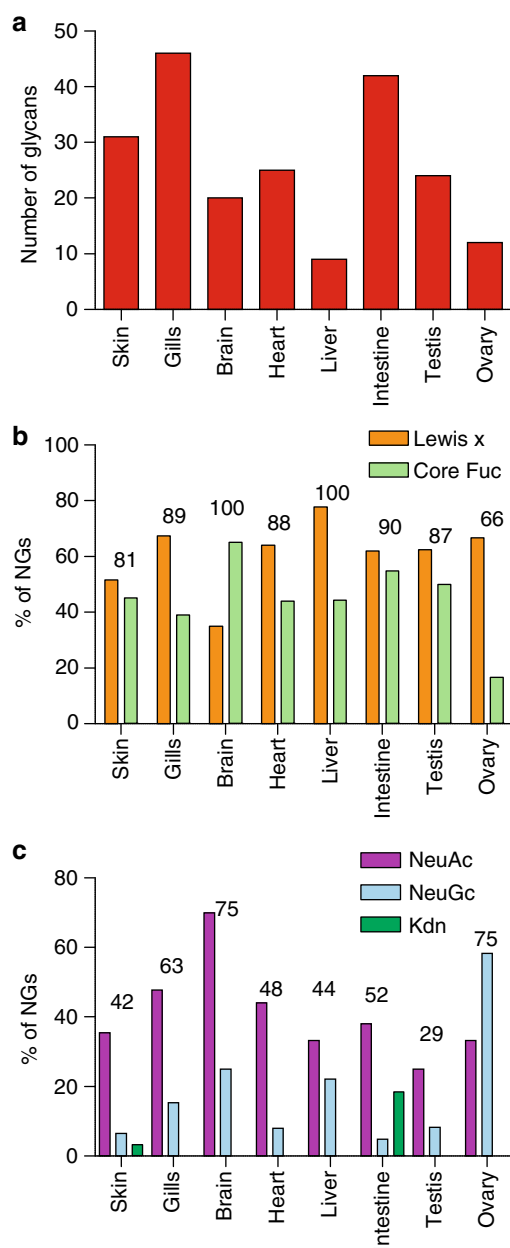


Fig. 4 Comparison of NGs features between organs. **a** Number of NGs confidently identified in each organ. **b** Fucosylation pattern expressed as the proportion of NGs substituted by one or more fucoses on the chitobiose disaccharides (green bars, core Fuc), and/or in Lewis-x epitope (orange bars); numbers indicate the overall percentage of fucosylated NGs. **c** Sialylation pattern expressed as the percentage of NGs substituted by any of the three sialic acids (purple bars, Neu5Ac; light blue bars, Neu5Gc; green bars, Kdn); numbers indicate the overall percentage of sialylated NGs relative to total NGs in the corresponding organs. It should be noted that a single NG may bear both Lewis-x and core Fuc, as well as different types of sialic acids

are Gal β 1-3(Neu5Ac α 2-6)GalNAc-ol (OG2), and Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-6)GalNAc-ol (OG15) in seven organs out of eight. Adult organs exhibited an overall much higher structural diversity than early embryos that are dominated by core 1 O-glycans OG18 and 20. It is noteworthy that the O-glycome profile of ovary is very similar to that of 1 h post-fertilized embryos. Other than these major glycans, two very minor Le^x-extended core 2 O-glycans and the oligosialylated versions of

NeuGc-substituted OG20 were identified in embryos but not observed in adult tissues²⁸. These minor O-glycans are believed to be mostly associated with soluble glycoconjugates in the perivitelline space, but not embryonic tissue²⁷. Their disappearance in adult organs is in agreement with decreasing oligosialylation of glycoproteins along zebrafish embryonic development²⁷.

As for N-glycans and O-glycans, glycolipids were extracted, purified, and permethylated prior to MS analysis (Fig. 6, Supplementary Figure 6, Supplementary Note 4). In addition, monosaccharide and lipid composition were confirmed by GC/MS. The structural analysis data are summarized in Supplementary Data 3 and 4. In all samples, we observed a high proportion of Hex-Ceramide (Hex-Cer) and lacto-Ceramide (Lac-Cer) that represent about 80% of total GSLs. Alongside Hex-Cer and Lac-Cer, we identified a complex pattern of GSLs dominated by gangliosides (Gg) and lacto-ganglioside (LcGg) series. Based on known biosynthetic pathways of the glycan moieties of GSLs in vertebrates and MS/MS analysis, we assigned 46 GSLs to Gala series (Gal β 1-1'Cer), 20 to Hemato series (Gal β 1-4Glc β 1-1'Cer), 1 to Lacto/Neolacto series (Gal β 1-3/4GlcNAc β 1-4Gal β 1-4Glc β 1-1'Cer), 49 to Lactoganglio series (GlcNAc β 1-3(GalNAc β 1-4)Gal β 1-4Glc β 1-1'Cer), and 56 to Ganglio series (GalNAc β 1-4Gal β 1-4Glc β 1-1'Cer) (Supplementary Data 3). The 172 GSLs (GL1-172) were made of 52 different carbohydrate moieties (G1-52) substituted by multiple ceramides, as summarized in Supplementary Data 4.

As revealed by MS mapping, GSLs exhibited a very high degree of organ-specificity (Fig. 6b, Supplementary Data 3 and 4) contributed by both glycan and ceramide moieties. The relative quantification of sphingoid bases revealed that d18:1 sphingosine was the major component for all organs (Fig. 6c). Additionally, significant proportions of t18:0 phytosphinganine were observed in skin, liver, intestine, testis, and ovary, whereas d18:0 sphinganine was identified in skin and testis. MS/MS analyses showed that the d18:1/C16:0 and d18:1/C24:0 combinations were the prevalent forms of ceramides, except for brain in which GSLs were almost exclusively substituted by the unusual d18:1/C18:0. Glycans moieties also exhibited a high level of organ-specificity with respect to the proportions of different GSLs families, level of fucosylation, and sialic acids diversity. As observed for NGs and OGs, the number of GSLs varies from one organ to the other (from 27 GSLs in liver to 43 in ovary) and their distribution in structurally related families is totally different (Fig. 6c). In particular, ovary stands out by the prevalence of the LactoGanglio series (>60%), whereas this family of GSLs is absent in brain and heart and below 20% in remaining organs. It is noteworthy that the fucosylation level of GSLs in different organs was more variable than in OGs and NGs (Fig. 6c). Indeed, the proportion of fucosylated GSLs varied from 0% in brain to about 40% in ovaries. Finally, the distribution of sialic acids in each organ was comparable to OGs and NGs, showing a prevalence of Neu5Ac-sialylated GSLs in all organs (Fig. 6c). The only discrepancy is the ovary's GSLs that were predominantly substituted by Neu5Gc whereas OGs and NGs were predominantly substituted by Neu5Ac. Finally, as observed for NGs and OGs, intestine was the only organ in which Kdn sialylated a significant proportion of GSLs.

Considering the prevalence of sialic acids in zebrafish glycome, we paid a particular attention to the sialylation machinery. As a first step, we screened databases for the *in silico* identification of *Danio rerio* ST genes, sialidases (*neu*), CMP-NeuAc synthase (*cmas*), and the CMP-NeuAc hydroxylase (*cmah*)^{33,49}. We identified 11 α 2,6-ST (*st6gal* and *st6galnac*), 10 α 2,3-ST (*st3gal*), 7 α 2,8-ST (*st8sia*), 2 CMAS (*cmas1* and *cmas2*), 7 sialidases (1 *neu1*, 5 *neu3* and 1 *neu4*), and 1 *cmah*, and oligonucleotide primer pairs were synthesized as described previously for each of

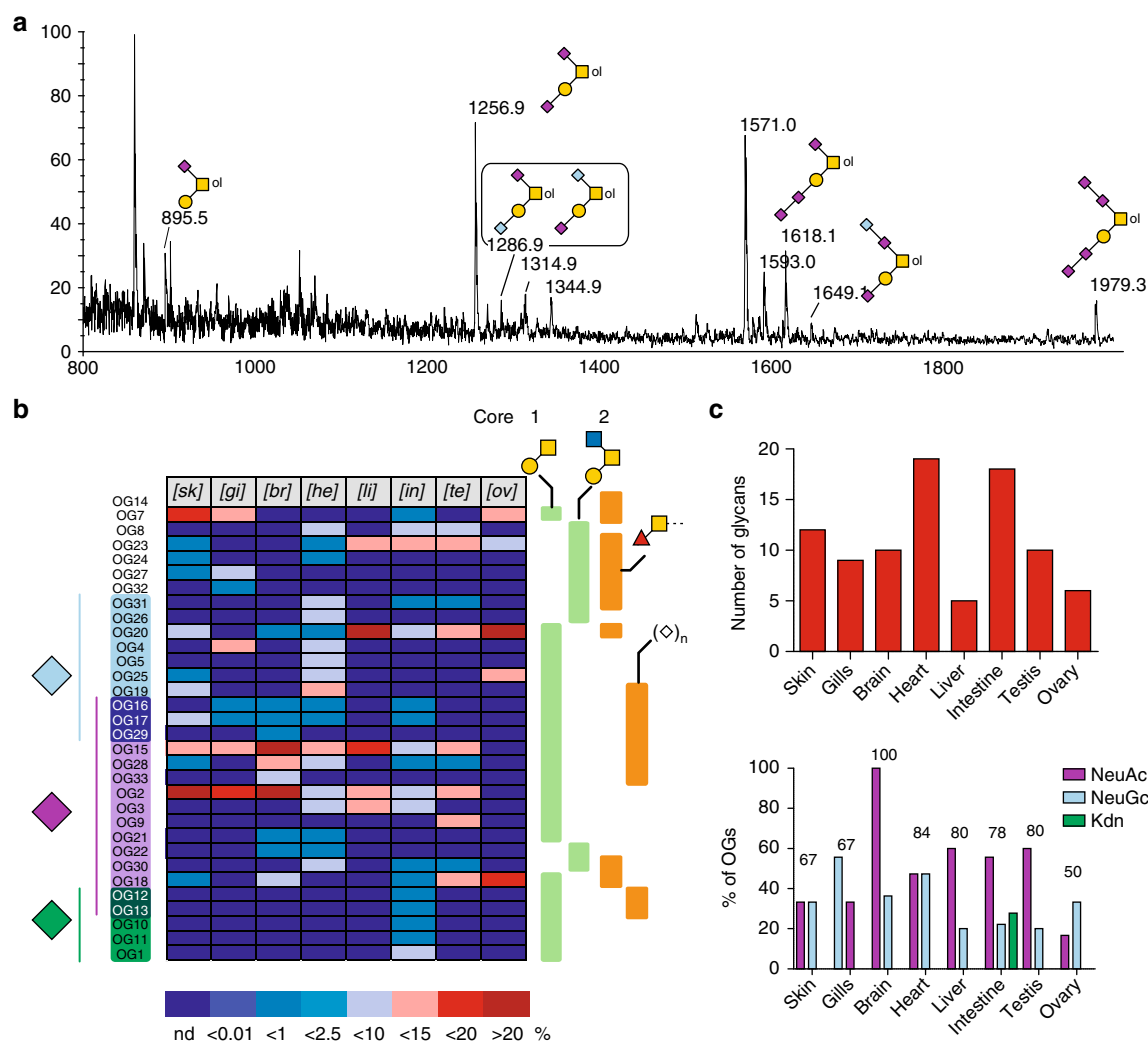


Fig. 5 O-glycome profiles. **a** Representative MALDI-TOF MS spectra of permethylated OGs isolated from brain. All MS spectra are provided in Supplementary Figure 4. **b** Tissues-specific distribution and relative quantification of OGs according to following glycan epitopes: nature of substituting sialic acids (Neu5Ac, Neu5Gc, Kdn), core-1 Gal β 1-3GalNAc, Core-2 Gal β 1-3(GlcNAc β 1-6)GalNAc, Fuc α 1-3GalNAc β , and oligosialylation. The nature of individual glycans was established by a combination of MS/MS, GC/MS, and NMR, in comparison with previous work^{28,29}. Tissue-specific distribution of glycans is presented in Supplementary Data 2. Relative quantifications of individual compounds were extracted from dataset presented in Supplementary Figure 5. **c** Number of OGs identified in each organ (upper panel) and proportion of sialylated OGs in each organ expressed as the percentage of OGs substituted by any of the three sialic acids (purple bars, Neu5Ac; light blue bars, Neu5Gc; green bars, Kdn); numbers indicate the overall percentage of sialylated OGs relative to total OGs in the corresponding organs. It should be noted that a single OG may bear different types of sialic acids

the ST²⁹, sialidases⁵⁰, CMAS, and CMAH⁵¹ (Supplementary Data 5). RT-PCR was used to establish the profile of expression of these genes in nine adult zebrafish tissues, i.e. ovary, testis, higher intestine, lower intestine, brain, heart, gill, skin, and liver (Fig. 7 and Supplementary Figures 7, 8). The recently described α 2,3-ST *st3gal1*-related genes (*st3gal1C* and *st3gal1D*), *st3gal3*³⁴ and the membrane-associated sialidase *neu4*⁵⁰ were not expressed at detectable level, whereas the sialidase genes *neu3.5* and *neu1*⁵⁰, the teleost-specific α 2,3-sialyltransferase gene *st3gal3-r*³⁴, and the 2 synthases *cmah* and *cmah* were expressed in all examined tissues. The *st3gal5* gene also known as the GM3 synthase is expressed in all tissues examined, which correlates well with the wide expression of GM3 and GM2 Gg. In addition, the *st8sia1* and *st8sia5* genes known as GD3 and GT3 synthases (SAT-II and SAT-V, respectively) shows a profile of expression restricted to brain, in agreement with diSia/oligoSia-containing GSLs observed in brain (Fig. 6b). Mono- α 2,3-sialylated core 1 structures (OGs) found in heart likely arise as a consequence of ST3Gal II activity since none of the *st3gal1*-related genes were expressed in heart. In

intestine, the monosialylated and disialylated OGs structures observed likely result from the enzymatic activity of ST6GalNAc IA, ST6GalNAc IB, ST3Gal IA since high expression of the corresponding genes was observed in this tissue. In addition, diSia epitopes on OGs were also identified in intestine, brain, and heart that probably resulted from ST8Sia VI enzymatic activity previously described for the mouse and human enzymes^{52,53}.

Intestine sialome. Among the organs that were screened, some exhibited remarkable or unexpected features. One of the most surprising features was the marked prevalence of Kdn for intestine (Fig. 2). Accordingly, monoclonal antibody raised against Kdn(α 2,3)Gal epitope (KDN3G) showed a strong response for two high molecular weight broad bands around 125 kDa and above 250 kDa in intestine, as well as in liver and heart although of lower intensities (Fig. 8a). Intestine also contained mucin-like components too large to reach separating gels (Supplementary Figure 9). Immunochemical detection of

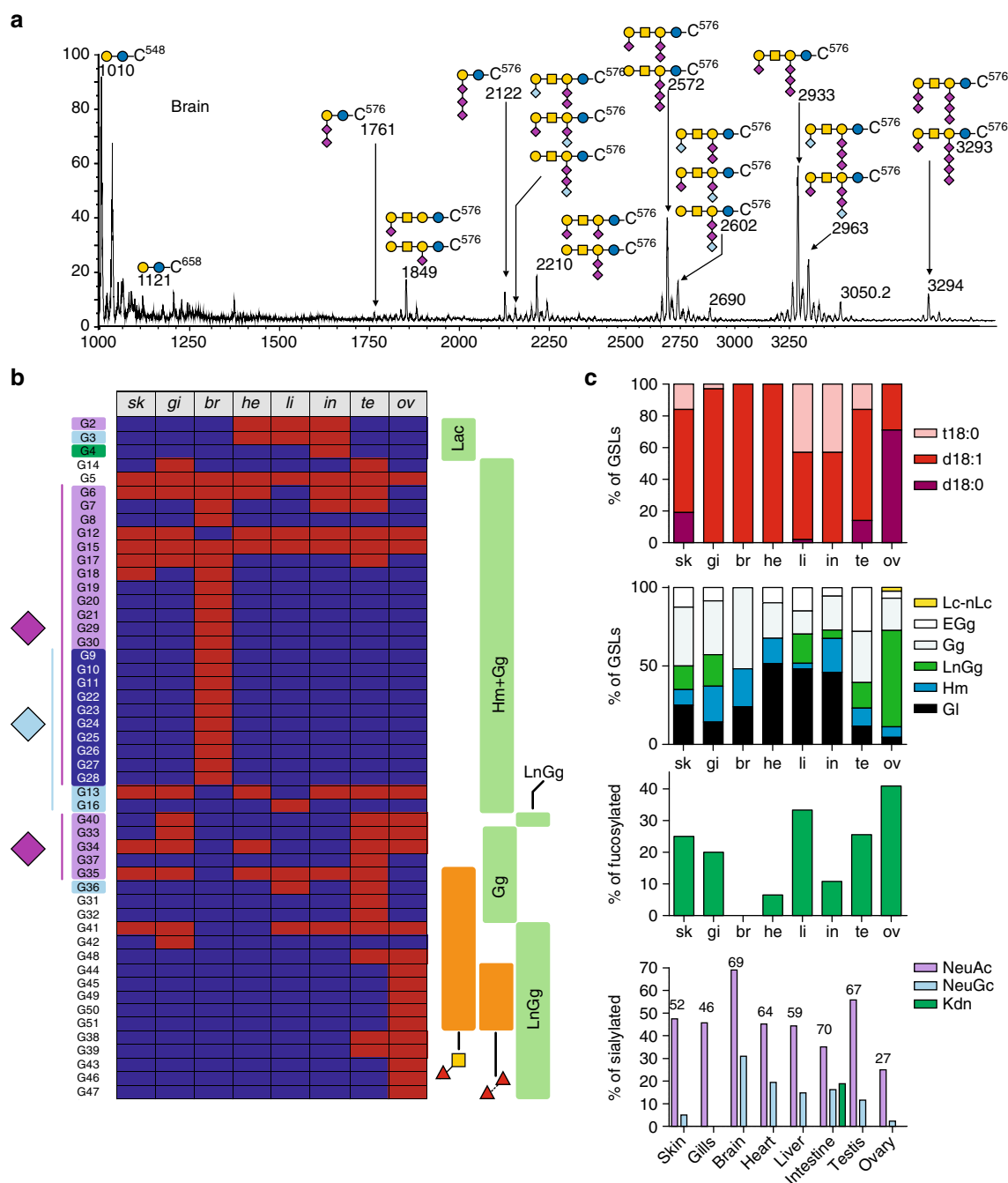


Fig. 6 Glyco-lipidome profiles. **a** Representative MALDI-TOF MS spectra of permethylated GSLs isolated from brain. All MS spectra are provided in Supplementary Figure 6. **b** Tissue-specific distribution of the glycan moieties of GSLs according to following glycan epitopes: nature of substituting sialic acids (Neu5Ac, Neu5Gc, Kdn), the GSL families (Lacto; hemato and ganglio Hm + Gg; extended galglio ExGg; lacto-ganglio LnGg), Fuc α 1,3GalNAc β , and oligo-fucosylation. The presence of multiple ceramides prevents reliable quantification. The identified species are depicted in red and the non-detected ones in dark-blue. **c** From top to bottom panels: distribution of ceramides expressed as the percentage of GSLs bearing one or the other sphingoid base (dark red bars, d18:0; medium red bars, d18:1; light red bars, t18:0); distribution of families expressed as the percentage of GSLs from one or the other family; occurrence of fucosylation expressed as the proportion of GSLs substituted by at least one fucose residue; sialylation pattern expressed as the percentage of GSLs substituted by any of the three sialic acids (purple bar, Neu5Ac; light blue bars, Neu5Gc; green bars, Kdn); numbers indicate the overall percentage of sialylated GSLs relative to total GSLs in the corresponding organs. It should be noted that a single GSL may harbor different types of sialic acids

Kdn using KDN3G antibody showed a specific signal at the surface of epithelial cells lining the lumen of zebrafish intestine (Fig. 8b). They appear as intense dot-like particles, in agreement with their putative attribution to aggregates of secreted mucins. However, it should be noted that the presence of Kdn is not restricted to OGs, so KDN3G may also target NG-

substituted glycoproteins and membrane glycolipids. Compared to intestinal epithelial cells, hepatocytes cells surface exhibited evenly distributed KDN3G-reactive epitopes, suggesting that Kdn was not carried on mucin-like secreted glycoproteins, in accordance with WB analysis (Supplementary Figure 9b).

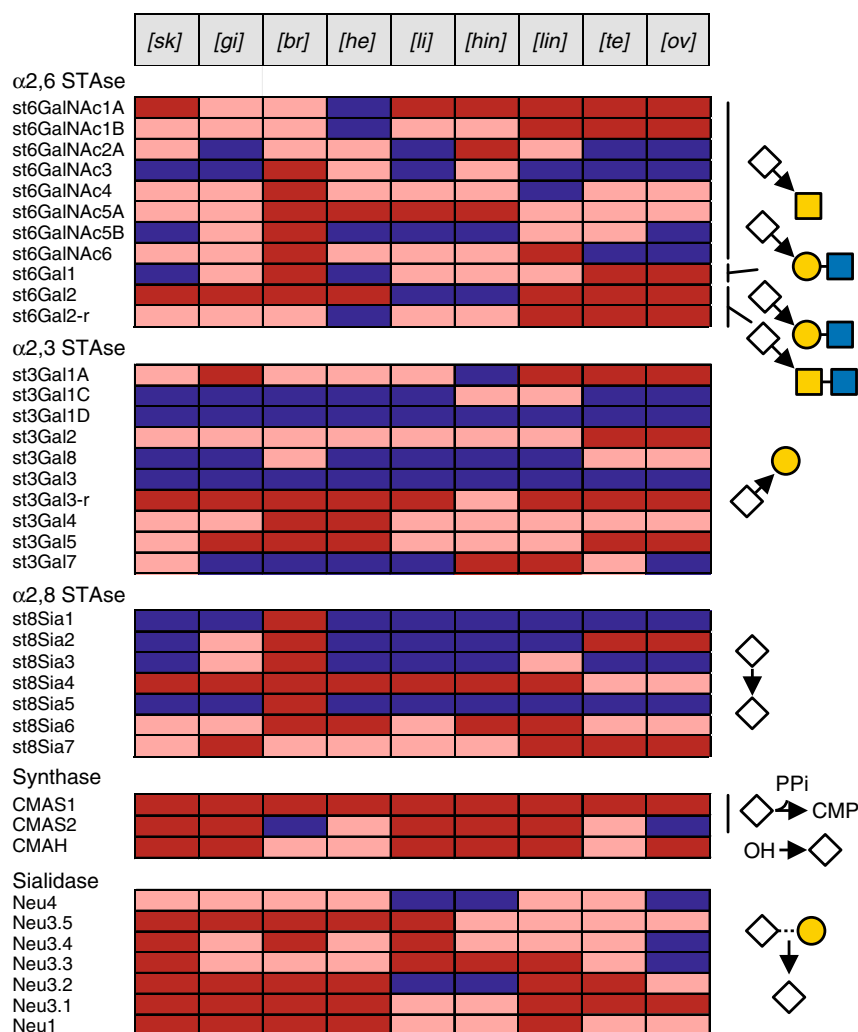


Fig. 7 Sialo-genes expression. Expression profiles of sialyltransferases, CMP-sialic acid synthases (CMAS), sialidases (Neu), CMP-sialic acid hydroxylases (CMAH), and sialidases genes in various zebrafish tissues based on RT-PCR data (Supplementary Figure 7). Expression levels of single mRNAs were depicted according to their relative intensities in each organ as non-detected (blue), intermediate intensity (pink), and strong intensity (red). Putative activities based on phylogenetic analyses are depicted at the right side of the map

Altogether, mass spectrometry, sialic acids quantification, Western-blot, and immunochemistry analyses established that intestine is the only organ that contains significant amount of Kdn-substituted glycans. In order to provide clues about the molecular bases that drive such exquisite organ specificity we tried to identify Kdn in the input and output of zebrafish. Whereas sialic acid analysis did not detect appreciable amount of Kdn in the fish food, it demonstrated that the intestine lumen contained significant amount of sialic acids (Neu5Ac 3 pmol/mg, Neu5Gc 5 pmol/mg, and Kdn 37 pmol/mg of dry material). Although variable from fish to fish, similar proportions of sialic acid species were recovered in the excreted feces. Feces components were separated by Nycodenz® density gradient using a recently developed procedure⁵⁴. Composition analysis showed that Kdn was highly enriched in purified microbiota, but not in the insoluble food debris (Fig. 8c). Furthermore, a vast majority of Kdn appeared to be associated with glycoconjugates, presumably bacterial polysaccharides, rather than present as free form. This strongly suggests that the microbiota harbored in the fish intestine independently synthesizes large amounts of Kdn. This view is reinforced by the discovery that the intestinal symbiont *Bacteroides thetaiotaomicron* can synthesize Kdn and transfer it to its capsular polysaccharides⁵⁵. However, considering the structure of Kdn-substituted glycans, including

those established for OGs, NGs, and GSLs, the identified glycans undoubtedly originate from zebrafish and not microbiota. Thus, it is conceivable that the high content of Kdn in intestine results from the direct scavenging of Kdn present in the lumen by the epithelial cells to build up zebrafish glycans in an energy saving manner (Fig. 8d). The scavenging of sialic acids through the intestinal barrier has been amply documented and could at least partially explain the clear tropism of Kdn for the intestinal tissue⁵⁶. However, it is hypothesized here that it may directly influence the intestinal glycosylation pattern through its recycling to epithelial mucin-type glycoconjugates, among other components. Alternatively, Kdn might also be directly transferred to extracellular glycoconjugates by exogenous bacterial sialyltransferase activities, but the structural similarity of Neu5Ac and Kdn-substituted glycans is clearly in favor of an endogenous sialyltransferase activity. Scavenged Kdn may also be secondarily transported to other organs through blood circulation particularly to liver, to be used in glycoconjugates metabolism. A complementary line of explanation may arise from the expression of two CMP-sialic acid synthases (CMAS) in zebrafish. Indeed, in contrast to other vertebrates, the genome of teleost fishes from the Otocephala clade contains two homologs of *Cmas* genes, *dreCmas1* and *dreCmas2* in zebrafish⁵¹. In zebrafish, both genes seem to be expressed as functional enzymes

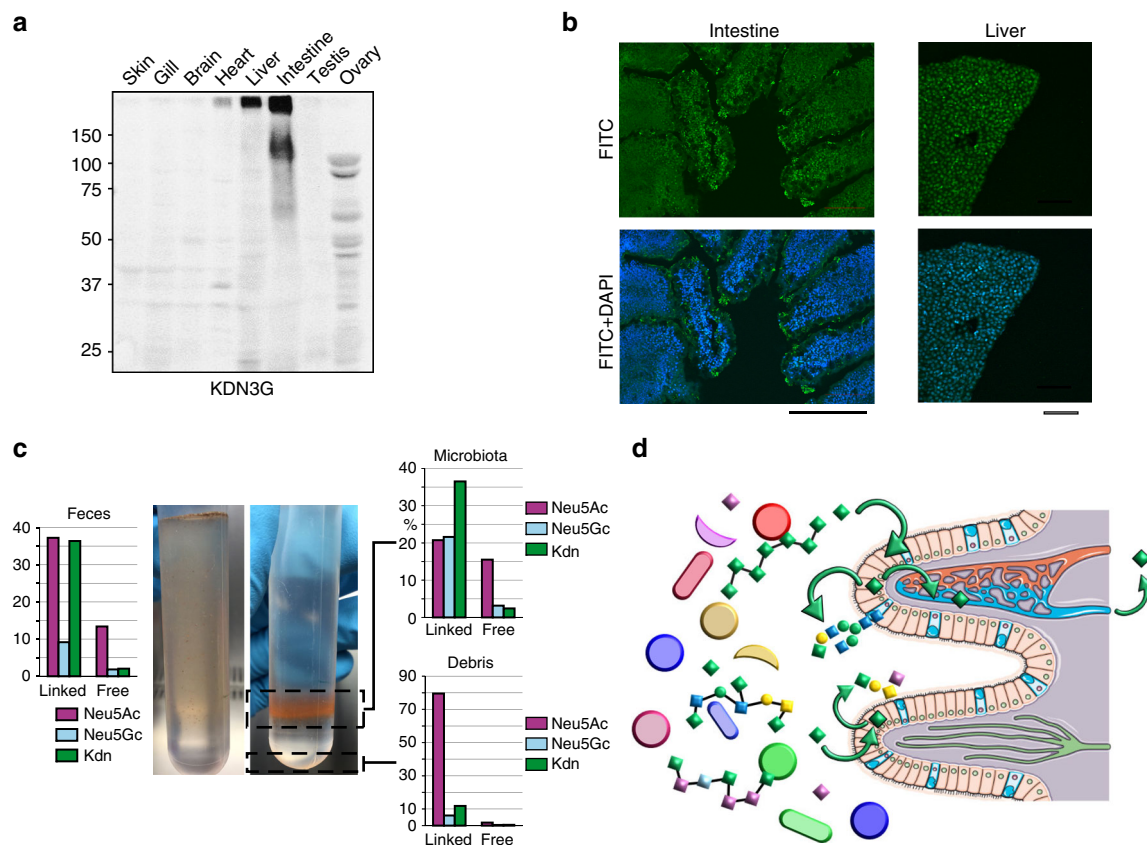


Fig. 8 Distribution of Kdn in organs. **a** WB screening of organs using Kdn(α 2,3)Gal specific-epitope KDN3G antibody raised against GM3-Kdn and characterized to recognize a purified KDN-containing glycoprotein (Gp-Kdn) derived from rainbow trout ovarian fluid^{11,76}. Control experiments using IgM/G are shown in Supplementary Figure 9b. **b** Histochemistry staining using DAPI and KDN3G showed the presence of Kdn(α 2,3)Gal at the surface of intestine epithelium and liver cells. Scale bar corresponds to 100 μ m for intestine and 50 μ m for liver. Control experiments using IgG are shown in Supplementary Figure 10. **c** Sialic acid composition analysis of feces before (left panel) and after (right panel) purification of by Nycodenz[®] gradient centrifugation shows a preferential association of Kdn with glycoconjugates of microbiota. **d** Hypothesized recycling of bacteria associated Kdn (green diamonds) by epithelial cells into endogenous glycoconjugates and subsequent transfer into blood circulation

that differ not only by their subcellular localization but also by their substrate specificity. Whereas dreCmas1 exhibits highest activity toward Neu5Ac, dreCmas2 preferentially activates Kdn⁵¹. Based on this finding, we have assessed the presence of the two paralogous genes in the different tissues by RT-PCR. As shown in Fig. 7 and Supplementary Figure 7, whereas dreCmas1 was ubiquitously expressed at similar levels in all organs, dreCmas2 exhibited a more restricted expression pattern. In particular, it was barely observed in ovary, brain, and heart but showed a robust expression in intestine, liver, and skin, as well as an intermediate expression in gills and testis. Interestingly, organs that showed the highest expression levels are precisely the ones that seem to contain the highest levels of Kdn associated with their glycoconjugates. This observation does not provide a single explanation for the tropism of Kdn toward a restricted set of organs, but one may speculate that the availability of Kdn substrate have driven the expression of a Kdn-specific *Cmas* and that the two observations are connected to the differential synthesis of Kdn sialoconjugates. Interestingly, the recently described *st3gal5*-related gene renamed *st3gal7*³⁴, which corresponds to the GM4 synthase⁵⁷, shows a profile of expression rather similar to dreCmas2 suggesting that *st3gal7* may also exhibit a certain specificity toward the transfer of Kdn on Gal-Cer moiety.

Brain sialome. The glycome of brain also stands out owing to its unique sialylation pattern. First, brain is the only organ in which glycans harbor substantial amount of oligosialylated motifs. In

particular, about 90% of the glycan moieties of GSLs from brain are sialylated and 75% are substituted by Sia₂ to Sia₄ epitopes. The presence of abundant oligosialylation in brain GSLs was confirmed by HPLC analysis of DMB-derivatized oligosialyl motifs, as shown in Supplementary Figure 11. Oligosialylation extends either from the Lac-Cer core of the hemato series (GL60-66; G7-11) or from the internal Gal residue of ganglio series trisaccharide core (GL95-106; G19-30) up to the newly described GP1d (GL106; G30). Gg are well known to represent the major type of GSLs in the brain of all inspected vertebrates so far. However, whereas the GSLs of mammalian brains, including humans, are dominated by mono-sialylated, di-sialylated, and tri-sialylated GSLs GM1, GD1a, GD1b, and GT1b, zebrafish brain is dominated by the tri-sialylated, tetra-sialylated and penta-sialylated GSLs GT1b, GT1c, GQ1c, and GP1d. The prevalence of higher form of sialylation in fish GSLs has already been documented in the brain of several fish species including codfish, skate, and dogfish^{58–60}. Similar to GSLs, all the OGs isolated from brains are sialylated, harboring up to four sialic acids. Brain synthesizes a very limited panel of OGs made of differentially sialylated Gal(β 1-3)GalNAc-ol core-1 molecules sialylated on the C6 position of GalNAc-ol by either Neu5Ac or Neu5Gc and potentially substituted on the C3 position of Gal residue by one or two sialic acids or on the C8 position of Sia(α 2-6)GalNAc-ol. Accordingly, brain expresses the most diverse panel of ST (Fig. 7 and Supplementary Figures 7, 8) being the only organ that expresses all eight members of *st6galnac* genes and three members of the *st6gal*

genes, which would explain the prevalence of sialylated OGs in brain. Similarly, the restricted expression of ST8Sia1 and ST8Sia3, which are believed to be involved in the elongation of sialylated chain of GD3 and GT3, is in agreement with the specific synthesis of b-series, c-series, and d-series of GSLs in brain. Finally, although the analytical techniques did not allow direct observation of polysialylated epitopes, both *st8sia2* and *st8sia4* genes responsible for the polysialylation of neural cell adhesion molecules (N-CAM), were expressed in the zebrafish brain as described previously in the rainbow trout brain⁶¹ and in accordance with previous immunochemistry study⁶².

As mentioned before, Neu5Gc is known to be expressed in minute amounts in the brain of vertebrates and its presence in neural tissues is still open to debates³⁸. Here, as many before us, we established that brain was the organ that contains the lowest amount of Neu5Gc, ranging from 1% of total sialic acids in GSLs to 8% in OGs⁶³. The low level of Neu5Gc synthesis in brain was substantiated by the relative low expression of CMP-Neu5Ac hydroxylase (CMAH) compared to other tissues (Fig. 7 and Supplementary Figure 7) as previously observed⁶⁴. Although low, the amount of Neu5Gc is not negligible and its presence in individual glycans was consistently proven by MS/MS analysis. Of particular relevance, the presence of three major NGs bearing the so-called zebrafish epitopes Gal β 1-4(NeuAc/Glc α 2-3)Gal β 1-4 (Fuc α 1-3)GlcNAc (NG89-91) and several Neu5Gc-substituted Gg (GL98-99, GL101, GL103-104) was firmly established. As shown in Supplementary Figure 12, the comparison of the MS/MS fragmentation patterns of GT1 molecules (GL96 and GL97 versus GL98, GL99, GL100, and GL101) and GP1 molecules (GL102 versus GL103 and 104) unambiguously established that Neu5Gc can replace Neu5Ac at multiple positions. The presence of Neu5Gc-substituted Gg in brain has already been firmly established in different species, which leaves no doubt concerning the capacity of brain tissue to process Neu5Gc^{39,65,66}. Although structural analysis of total brain cannot unambiguously answer to the nagging question about the localization of Neu5Gc-substituted glycans, the data provided here supports the localization of Neu5Gc in neural tissues rather than in endothelial tissues or blood cells in two ways. First, the glycan moieties of the Neu5Gc-substituted glycans identified in zebrafish brain are identical to those of the Neu5Ac-substituted glycans. This is of particular relevance for GSLs, because oligosialylated Gg are known to be very specific of neural tissues. As shown in Fig. 6a and Supplementary Data 3, the pattern of Neu5Gc-GSLs is identical to Neu5Ac-GSLs. Second, not only the glycan moieties of Neu5Ac and Neu5Gc containing GSLs are identical, but also their lipid moieties. Indeed, Neu5Gc containing GSLs are made of the specific d18:1/C18:0 ceramide (*m/z* 576) that was exclusively observed in brain GSLs. Thus, similarities between Neu5Gc and Neu5Ac containing glycans strongly suggest that neural tissue synthesizes small but significant amounts of Neu5Gc containing glycoconjugates.

Finally, considering the specificities of brain glycomes, we investigated if some of these features were shared among the different regions of brain. Toward this aim, we have separated the zebrafish brain into six defined regions as previously described⁶⁷, including olfactory bulb, telencephalon, optic tectum, cerebellum, medulla, and hypothalamus. Considering the minute amount of tissues at disposition, we focused on the NGs. In agreement with previous observations, the Neu5Gc proportion in individual regions was lower than 1%. Analyses of Neu5Gc content using nine replicates did not show significant differences between regions. Similarly, MS analysis of N-glycomes did not reveal major differences between regions (Supplementary Figure 13). Indeed, most of the major NGs identified in whole brain could be observed in individual regions albeit in different proportions. As

for whole brain, most regions showed a high proportion of complex NGs except cerebellum, which showed a higher proportion of oligomannosylated NGs than complex NGs. MS/MS analysis of major complex NGs confirmed the prevalence of core fucosylated NGs, as was observed in whole brain. In particular, NG76 that appears as the major brain-specific NG was observed in all regions, albeit also in different proportions. The presence of Neu5Gc-substituted NGs could be established by MS and MS/MS in some region, in particular in olfactory bulb (Supplementary Figure 13). Among those, the ubiquitous Neu5Gc containing NG90 and NG91 could be clearly identified exclusively in the olfactory bulb from three different independent preparations.

Discussion

In conclusion, the present study established that zebrafish synthesizes a rich and diverse glycome that share numerous features with higher vertebrates including the expression of fucosylated and sialylated terminal LacNAc and LacdiNAc epitopes, simple sialylated O-glycans and Gg expressions. Zebrafish glycans are decorated with species-specific and organ-specific glycans epitopes that arise from unique extensions of mammalian structural motifs, notably the Gal(β 1,4)-extended version of LacNAc unit and GalNAc(β 1,4)-extended core-1 OGs. On the other hand, zebrafish lacks common mammalian epitopes, such as those based on type-1 chain (Gal β 1-3GlcNAc, Le^a, Le^b, etc.) and major GSLs families like globosides. More importantly, the present report demonstrated that systems glycomics could uncover so far untapped physiological processes, such as scavenging and metabolic incorporation of bacterial monosaccharides in intestine. Altogether, we believe that the glycosylation map of zebrafish, in conjunction with the expression pattern of specific glyco-enzymes, provides a unique opportunity for the zebrafish biology and glycobiology communities to address the glycans functions on a solid structural background.

Methods

Zebrafish handling. Wild-type *Danio rerio* AB stock was purchased from the Zebrafish International Resource Center (Eugene, OR, <https://zebrafish.org/home/guide.php>). Zebrafish were raised in a zebTEC (Techniplast) Stand Alone Zebrafish Housing system at 28 °C⁶⁸. The zebrafish handling was conducted according to the French and European Union guidelines for the handling of laboratory animals and approved by the local Ethics Committee (APAFiS approval no. 2018011722529804). Young adult zebrafish 6–9 months were sampled and euthanized by immersion in a solution of tricaine methane sulfonate (MS-222, 300 mg/L) for at least 10 min. Zebrafish were dissected and organs from 20 adults (mixed male and females, except for testis and ovaries) were pooled before extraction.

Purification of glycans. Freshly dissected tissues were suspended in water, homogenized with a mechanical disperser and freeze dried. Dry material was sequentially extracted three times by chloroform/methanol (2:1, v/v) and chloroform/methanol (1:2, v/v). Pellets contained glycoproteins whereas supernatant contained glycolipids. Glycoprotein fraction was suspended in Tris/HCl and centrifuged at low speed. Supernatant was precipitated with 70% cold ethanol overnight and centrifuged at 1200×g. Total glycoprotein fraction was re-suspended in a solution of 6 M guanidinium chloride and 5 mM EDTA in 0.1 M Tris/HCl, pH 8, and agitated for 4 h at 4 °C. Dithiothreitol was then added to a final concentration of 20 mM and incubated for 5 h at 37 °C, followed by addition of iodoacetamide to a final concentration of 50 mM and further incubated overnight in the dark at room temperature. Reduced/alkylated sample was dialyzed against water at 4 °C for 3 days and lyophilized. The recovered protein samples were then sequentially digested by TPCK-treated trypsin for 5 h and chymotrypsin overnight at 37 °C, in 50 mM ammonium bicarbonate buffer, pH 8.4. Crude peptide fraction was separated from hydrophilic components on a C18 Sep-Pak cartridge (Waters) equilibrated in 5% acetic acid by extensive washing in the same solvent and eluted with a step gradient of 20%, 40%, and 60% propan-1-ol in 5% acetic acid. Pooled propan-1-ol fraction was dried and subjected to N-glycosidase F (Roche) digestion in 50 mM ammonium bicarbonate buffer pH 8.4, overnight at 37 °C. The released N-glycans were separated from peptides using the same C18 Sep-Pak procedure as described above. To liberate O-glycans, retained peptide fraction from C18 Sep-Pak was submitted to alkaline reductive elimination in 100 mM NaOH containing 1.0 M sodium borohydride at 37 °C for 72 h. The reaction was stopped by addition

of Dowex 50 × 8 cation-exchange resin (25–50 mesh, H⁺ form) at 4 °C until pH 6.5 and, after evaporation to dryness, boric acid was distilled as methyl ester in the presence of methanol. Total material was then submitted to cation-exchange chromatography on a Dowex 50 × 2 column (200–400 mesh, H⁺ form) to remove residual peptides. Glycolipid fraction was de-O-acylated with 0.1 M NaOH in chloroform/methanol (1:1, v/v) at 37 °C, 2 h and purified on a C18 Sep-Pak cartridge (Waters). Cartridge was washed with methanol/water (1:1, v/v) and eluted successively by methanol, chloroform/methanol (1:1, v/v) and (2:1, v/v). For the purification of N-glycans from brain regions, an alternative methodology based on the extraction of glycoproteins by Triton X-100 has been used⁶⁹. In brief, tissues were suspended in 200 µL of 1% triton in PBS and sonicated three times for 10 s. Cell debris were removed by centrifugation at 10,000×g, 4 °C for 10 min and a final concentration of 10 mM dithiothreitol and 50 mM iodoacetamide in 100 µL PBS were added sequentially to the supernatant, each followed by incubation at 37 °C for 1 h. Then 44 µL of 100% trichloroacetic acid was added and left at –20 °C for 30 min. The pellet collected by centrifugation at 10,000×g for 30 min was washed with 100% acetone three times and dried under a stream of nitrogen. Material was then subjected to trypsin and PNGase treatment and the N-glycans purified using the conditions aforementioned.

Structural analysis of glycans. Following their purification, NGs, OGs, and GSLs were permethylated using the NaOH/dimethyl sulfoxide reagent⁷⁰. The permethylated derivatives were then extracted in chloroform and repeatedly washed with water. MALDI-TOF and MALDI-TOF/TOF spectra were acquired on 4800 Proteomics Analyzer mass spectrometer (Applied Biosystems, Framingham, MA, USA) in reflectron positive or negative mode by delayed extraction using an acceleration mode of 20 kV, a pulse delay of 200 ns and grid voltage of 66%. Samples were prepared by mixing directly on the target 1 µL of oligosaccharide solution (1–5 pmol) with 1 µL of 2,5 dihydroxybenzoic acid matrix solution (10 mg/mL in CH₃OH/H₂O, 50/50, vol/vol). Between 50 and 100 scans were averaged for every spectrum.

The nature of each signal was established by calculating their individual monosaccharide, and optionally ceramide, contents. All identified carbohydrates-containing signals were fragmented by CID on a MALDI-TOF/TOF mass spectrometer to establish the sequences of corresponding oligosaccharides. The fragmentation patterns were all manually sorted out and eventually confirmed by automated identification tools (glycoworkbench). Only compounds presenting consistent MS/MS spectra were included into the final glycan list. In many instances, the structures of individual compounds were deduced from MS/MS spectra containing several isomers. In these cases, we confirmed the presence of a compound exclusively when a set of specific fragments could be unambiguously attributed. When needed, the structural features inferred from mass spectrometry approach were confirmed by further complementary techniques including enzymatic degradation, HPLC separation of oligosialic acids, linkage analysis by GC/MS and recognition by specific antibodies. The final structural attribution took into account known biosynthetic pathways of vertebrate glycans and previous in depth studies of zebrafish glycans using MS/MS, NMR, GC/MS, and specific glycosidases^{27–29}. For the sake of comparison, all glycans were relatively quantified as percentages of all identified glycan within each sample by integrating corresponding MALDI-MS signals from three independent experiments⁷¹. Whenever an individual signal contained several isomeric compounds, the relative proportion of each isomer was inferred from the MS/MS signal intensities of structurally related motifs.

Analysis of sialic acids. Intact sialic acids were released by mild hydrolysis in 0.1 N TFA at 80 °C for 1 h and coupled to DMB²⁹. For DMB derivatization, sialic acids reacted with a volume of DMB reagent at 50 °C for 2 h 30 min, then separated isocratically on a C18 reverse phase HPLC column (250 × 4.6 mm, 5 µm, Vydac) by a solvent mixture of acetonitrile/methanol/water (7:9:84) and identified by referring to the elution positions of standard Neu5Ac, Neu5Gc, and Kdn derivatives⁷². Individual sialic acid derivatives were quantified by integration of fluorescence signals after HPLC separation, plotted against standard curves of corresponding authentic standards. In parallel, the nature of sialic acids was confirmed by analyzing MS and MS/MS fragmentation patterns by LC/MS based on known patterns^{73,74}.

Analysis of monosaccharides. Lipid moiety of glycolipids was analyzed by GC/MS as trimethylsilylated (TMS) derivative. Glycolipids were methanolysed (1 mL of MeOH/HCl 0.5 M) at 80 °C during 16 h. The fatty acid methyl esters (FAME) were extracted by washing the methanol fraction by heptane three times. The methanol fraction was N-re-acetylated by 20 µL of acetic anhydride then dried under nitrogen and derivatized by 30 µL BisSilylTriFluoro-Acetamide (BSTFA, Sigma-Aldrich) in 30 µL of pyridine for 2 h at room temperature before analysis by GC-MS.

Derivatisation of lipids. Lipid moiety of glycolipids was analyzed by GC/MS as TMS derivatives. Glycolipids were methanolysed in 500 µL of 0.5 M of 82% aq. MeOH/HCl at 80 °C for 16 h⁷⁵. The FAME were extracted three times from the methanol fraction by 250 µL of heptane; which were pooled and concentrated to 100 µL of heptane prior to its GC/MS injections. Then, the methanol fraction was

neutralized to pH 6.5 using Ag₂CO₃ prior to N-re-acetylation by the addition of 20 µL of acetic anhydride at RT overnight. The methanol fraction was filtered and dried under nitrogen in order to derivatize the long chain base (LCB) by 20 µL of bis(1-trifluoroacetamide) (BSTFA) in 20 µL of pyridine at RT during 2 h; prior to its GC/MS injection.

Analysis of lipid derivatives by GC-MS. The FAME and LCB derivatives were analyzed using a Thermo Trace Ultra gas chromatograph interfaced with a Finnigan Automass II mass spectrometer. The lipid derivatives were injected (inlet temperature 280 °C) to a non-polar Alltech[®] EconoCap[™] Capillary-Column-1 (0.25 µm film, 0.25 mm 25 m). The mass spectrometer was used in the electron ionization mode (70 eV, source temperature 150 °C) mode with automatic ionization to produce a total ion chromatogram (TIC). The FAME temperature program was 120–280 °C at 5 °C/min and 10 min left at 280 °C. The LCB-TMS temperature program was 120–240 °C at 2 °C/min, 10 min left at 240, 240–280 °C at 4 °C/min and 10 min left at 280 °C.

Histochemistry. Tissues were excised from the zebrafish after fasting, washed with PBS, and fixed with 4% PFA for overnight. After washing with PBS, the tissues were serially soaked into 5%, 10%, and 15% sucrose in PBS for 1 h each and finally into 20% sucrose in PBS for overnight. They were then incubated with the OCT solution (OCT compound: 20% sucrose in PBS = 1:1) for 1 h. All procedures were performed at 4 °C. The tissues were soaked into the OCT compound and frozen at –80 °C. The frozen tissues were sliced into sections (14 µm thick), put onto the slide glass, and dried at 37 °C. The dried sections were washed five times with PBS and blocked with 5% goat serum in PBS for 1 h. Then the sections were incubated with primary antibody KDN3G⁷⁶ (1:100 dilution, 10 µg/mL) at 4 °C for overnight. After washing five times with PBS for 15 min each, they were incubated with secondary antibody (Alexa-488-conjugated anti-mouse IgG, 1:400 dilution; Thermo Fischer Scientific Ref A11001) at 4 °C for 1 h, followed by washing six times with PBS for 15 min each. Nuclear staining was performed by incubating with DAPI solution (1 µg/mL, 37 °C for 15 min). The sections were finally washed with PBS, and covered by cover glass with a drop of mounting fluid. Observation was performed under a fluoromicroscope (BX51, Olympus) or confocal scanning fluorescent microscope (LSAM 710; Karl Zeiss).

Western blots. Tissues were homogenized with PBS containing 1% Triton X-100, 1 µg/mL each of aprotinin, leupeptin, and pepstatin, 2 µg/mL of antipain, and 5 mM EDTA and incubated on ice for 1 h. After centrifugation of the homogenates, the protein concentration of the supernatants was measured by the bicinchoninic acid (BCA) assay. The supernatants were dissolved in Laemmli buffer containing 5% mercaptoethanol and incubated at 60 °C for 20 min. The samples were subjected to SDS-PAGE (10% or 7.5% polyacrylamide gels), followed by CBB staining and Western blotting using PVDF membranes. For immunostaining, PVDF membranes were blocked with PBS containing 0.05% Tween 20 (PBST) and 1% BSA at 25 °C for 1 h, and incubated overnight with or without the primary antibody, KDN3G (1:1000 dilution, 10 µg/mL) at 4 °C overnight, followed by incubation with the secondary antibody, peroxidase-conjugated anti-mouse IgG + IgM (1:5000 dilution, 0.4 µg/mL; American Qualex, Ref A108PS) at 37 °C for 45 min. Visualization was performed with chemiluminescent reagents (GE Healthcare).

RNA extraction, cDNA synthesis, and RT-PCR analysis. Total RNA was extracted from zebrafish tissues using the Nucleospin[®] RNA II kit (Macherey-Nagel, Hoerd, France). A proteinase K digestion step (55 °C, 10 min) and phenol/chloroform extraction were inserted in the protocol after dounce homogenization of the tissues and before column purification of total RNA. RNA integrity was further assessed using the RNA 6000 Nano LabChip Kit on an Agilent Bioanalyzer (Agilent Technologies, Stratagene, La Jolla, CA). For subsequent PCR amplifications, first strand cDNA was synthesized from total RNA using an oligo(dT) primer and the AffinityScript Q-PCR cDNA synthesis kit (Agilent technologies) according to the manufacturer protocol. Oligonucleotide primers were designed (Eurogentec, Herstal, Belgium) in the open-reading frame of the zebrafish sequences (Supplementary Data 5). PCR amplifications were carried out for 35 cycles with the Taq Core kit DNA polymerase (Qiagen, Courtaboeuf, France) with annealing temperatures ranging from 48 to 68 °C. Amplified fragments were subjected to 2% agarose gel electrophoresis, visualized by ethidium bromide, gel-extracted and subcloned in the pCR[®]2.1-TOPO vector and nucleotide sequences were confirmed by sequencing (Genoscreen, Lille, France). The uncropped agarose gel electrophoresis are shown in Supplementary Figure 8.

Data availability

All the identified NGs, OGs and GSLs were included in Glycome Atlas that is freely available at <http://www.rings.t.soka.ac.jp/GlycomeAtlasV5/index.html>. The data that support the findings of this study are available from the corresponding author upon request.

Received: 5 October 2017 Accepted: 26 September 2018

Published online: 07 November 2018

References

- Golling, G. et al. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat. Genet.* **31**, 135–140 (2002).
- Broughton, R. E., Milam, J. E. & Roe, B. A. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res.* **11**, 1958–1967 (2001).
- Bassi, I. et al. The zebrafish: an emerging animal model for investigating the hypothalamic regulation of reproduction. *Minerva Endocrinol.* **41**, 250–265 (2016).
- Hortopan, G. A., Dinday, M. T. & Baraban, S. C. Spontaneous seizures and altered gene expression in GABA signaling pathways in a mind bomb mutant Zebrafish. *J. Neurosci.* **30**, 13718–13728 (2010).
- Santoriello, C. & Zon, L. I. Hooked! Modeling human disease in zebrafish. *J. Clin. Invest.* **122**, 2337–2343 (2012).
- Bournele, D. & Beis, D. Zebrafish models of cardiovascular disease. *Heart Fail. Rev.* **21**, 803–813 (2016).
- Matrone, G., Tucker, C. S. & Denvir, M. A. Cardiomyocyte proliferation in zebrafish and mammals: lessons for human disease. *Cell. Mol. Life Sci.* <https://doi.org/10.1007/s00018-016-2404-x> (2016).
- Ceol, C. J. et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513–517 (2011).
- Patton, E. E. et al. BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol.* **15**, 249–254 (2005).
- Rudner, L. A. et al. Shared acquired genomic changes in zebrafish and human T-ALL. *Oncogene* **30**, 4289–4296 (2011).
- Kawahara, G., Guyon, J. R., Nakamura, Y. & Kunkel, L. M. Zebrafish models for human FKRP muscular dystrophies. *Hum. Mol. Genet.* **19**, 623–633 (2010).
- Shaheen, R. et al. A founder CEP120 mutation in Jeune asphyxiating thoracic dystrophy expands the role of centriolar proteins in skeletal ciliopathies. *Hum. Mol. Genet.* **24**, 1410–1419 (2015).
- Xi, Y., Noble, S. & Ekker, M. Modeling neurodegeneration in zebrafish. *Curr. Neurol. Neurosci. Rep.* **11**, 274–282 (2011).
- Pack, M. I. B. D. Fishing for missing heritability in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 318–320 (2015).
- Teppa, R. E., Petit, D., Plechakova, O., Cogez, V. & Harduin-Lepers, A. Phylogenetic-derived insights into the evolution of sialylation in Eukaryotes: comprehensive analysis of vertebrate β -galactoside α 2,3/6-sialyltransferases (ST3Gal and ST6Gal). *Int. J. Mol. Sci.* **17** (2016).
- Song, Y. et al. Neural and synaptic defects in slytherin, a zebrafish model for human congenital disorders of glycosylation. *PLoS ONE* **5**, e13743 (2010).
- Bentrop, J., Marx, M., Schattschneider, S., Rivera-Milla, E. & Bastmeyer, M. Molecular evolution and expression of zebrafish St8SialIII, an α -2,8-sialyltransferase involved in myotome development. *Dev. Dyn.* **237**, 808–818 (2008).
- Thornhill, P., Bassett, D., Lochmüller, H., Bushby, K. & Straub, V. Developmental defects in a zebrafish model for muscular dystrophies associated with the loss of fukutin-related protein (FKRP). *Brain J. Neurol.* **131**, 1551–1561 (2008).
- Manzini, M. C. et al. Exome sequencing and functional validation in Zebrafish identify GTDC2 mutations as a cause of Walker–Warburg syndrome. *Am. J. Hum. Genet.* **91**, 541–547 (2012).
- Carss, K. J. et al. Mutations in GDP-mannose pyrophosphorylase B cause congenital and limb-girdle muscular dystrophies associated with hypoglycosylation of α -dystroglycan. *Am. J. Hum. Genet.* **93**, 29–41 (2013).
- Wiweger, M. I. et al. Cartilage ultrastructure in proteoglycan-deficient zebrafish mutants brings to light new candidate genes for human skeletal disorders. *J. Pathol.* **223**, 531–542 (2011).
- Thakur, P. C., Davison, J. M., Stuckenholtz, C., Lu, L. & Bahary, N. Dysregulated phosphatidylinositol signaling promotes endoplasmic reticulum-stress-mediated intestinal mucosal injury and inflammation in zebrafish. *Dis. Model. Mech.* **7**, 93–106 (2013).
- Boccutto, L. et al. A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. *Hum. Mol. Genet.* **23**, 418–433 (2014).
- Dehnert, K. W. et al. Imaging the sialome during zebrafish development with copper-free click chemistry. *ChemBiochem Eur. J. Chem. Biol.* **13**, 353–357 (2012).
- Laughlin, S. T., Baskin, J. M., Amacher, S. L. & Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **320**, 664–667 (2008).
- Dehnert, K. W. et al. Metabolic labeling of fucosylated glycans in developing zebrafish. *ACS Chem. Biol.* **6**, 547–552 (2011).
- Chang, L.-Y. et al. Developmental regulation of oligosialylation in zebrafish. *Glycoconj. J.* **26**, 247–261 (2009).
- Guérardel, Y., Chang, L.-Y., Maes, E., Huang, C.-J. & Khoo, K.-H. Glycomic survey mapping of zebrafish identifies unique sialylation pattern. *Glycobiology* **16**, 244–257 (2006).
- Vanbeselaere, J. et al. Mapping the expressed glycome and glycosyltransferases of zebrafish liver cells as a relevant model system for glycosylation studies. *J. Proteome Res.* **11**, 2164–2177 (2012).
- Takemoto, T., Natsuka, S., Nakakita, S.-I. & Hase, S. Expression of complex-type N-glycans in developmental periods of zebrafish embryo. *Glycoconj. J.* **22**, 21–26 (2005).
- Hanzawa, K., Suzuki, N. & Natsuka, S. Structures and developmental alterations of N-glycans of zebrafish embryos. *Glycobiology* **27**, 228–245 (2017).
- Harduin-Lepers, A. et al. The human sialyltransferase family. *Biochimie* **83**, 727–737 (2001).
- Harduin-Lepers, A. Comprehensive analysis of sialyltransferases in vertebrate genomes. *Glycobiol. Insights* **2**, 29–61 (2010).
- Petit, D. et al. Integrative view of α 2,3-sialyltransferases (ST3Gal) molecular and functional evolution in deuterostomes: significance of lineage-specific losses. *Mol. Biol. Evol.* **32**, 906–927 (2015).
- Angata, T., Nakata, D., Matsuda, T. & Kitajima, K. Elevated expression of free deaminoneuraminic acid in mammalian cells cultured in mannose-rich media. *Biochem. Biophys. Res. Commun.* **261**, 326–331 (1999).
- Go, S., Sato, C., Furuhashi, K. & Kitajima, K. Oral ingestion of mannose alters the expression level of deaminoneuraminic acid (KDN) in mouse organs. *Glycoconj. J.* **23**, 411–421 (2006).
- Angata, T., Nakata, D., Matsuda, T., Kitajima, K. & Troy, F. A. Biosynthesis of KDN (2-Keto-3-deoxy-d- glycerol -D- galactol -nononic acid): identification and characterization of A kdn-9-phosphate synthetase activity FROM trout testis. *J. Biol. Chem.* **274**, 22949–22956 (1999).
- Davies, L. R. L. & Varki, A. Why is N-glycolylneuraminic acid rare in the vertebrate brain? *Top. Curr. Chem.* **366**, 31–54 (2015).
- Mikami, T. et al. Further characterization of equine brain gangliosides: the presence of GM3 having N-glycolyl neuraminic acid in the central nervous system. *J. Biochem.* **123**, 487–491 (1998).
- Yu, S.-Y., Wu, S.-W. & Khoo, K.-H. Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type N-glycans. *Glycoconj. J.* **23**, 355–369 (2006).
- Konishi, Y. & Aoki-Kinoshita, K. F. The GlycomeAtlas tool for visualizing and querying glycome data. *Bioinformatics* **28**, 2849–2850 (2012).
- Tiemeyer, M. et al. GlyTouCan: an accessible glycan structure repository. *Glycobiology* **27**, 915–919 (2017).
- Sasaki, N. et al. Drosophila beta 1,4-N-acetylgalactosaminyltransferase-A synthesizes the LacdiNAc structures on several glycoproteins and glycosphingolipids. *Biochem. Biophys. Res. Commun.* **354**, 522–527 (2007).
- Khoo, K. H., Chatterjee, D., Caulfield, J. P., Morris, H. R. & Dell, A. Structural mapping of the glycans from the egg glycoproteins of *Schistosoma mansoni* and *Schistosoma japonicum*: identification of novel core structures and terminal sequences. *Glycobiology* **7**, 663–677 (1997).
- Sato, T. et al. Molecular cloning and characterization of a novel human beta 1,4-N-acetylgalactosaminyltransferase, beta 4GalNAc-T3, responsible for the synthesis of N,N'-diacetyllactosylamine, galNAc beta 1-4GlcNAc. *J. Biol. Chem.* **278**, 47534–47544 (2003).
- Do, K. Y., Do, S. I. & Cummings, R. D. Differential expression of LacdiNAc sequences (GalNAc beta 1-4GlcNAc-R) in glycoproteins synthesized by Chinese hamster ovary and human 293 cells. *Glycobiology* **7**, 183–194 (1997).
- Kitamura, N. et al. Prognostic significance of reduced expression of beta-N-acetylgalactosaminylated N-linked oligosaccharides in human breast cancer. *Int. J. Cancer* **105**, 533–541 (2003).
- Hirano, K., Matsuda, A., Shirai, T. & Furukawa, K. Expression of LacdiNAc groups on N-glycans among human tumors is complex. *Biomed. Res. Int.* **2014**, 981627 (2014).
- Petit, D., Teppa, E., Cenci, U., Ball, S. & Harduin-Lepers, A. Reconstruction of the sialylation pathway in the ancestor of eukaryotes. *Sci. Rep.* **8**, 2946 (2018).
- Manzoni, M. et al. Molecular cloning and biochemical characterization of sialidases from zebrafish (*Danio rerio*). *Biochem. J.* **408**, 395–406 (2007).
- Schaper, W. et al. Identification and biochemical characterization of two functional CMP-sialic acid synthetases in *Danio rerio*. *J. Biol. Chem.* **287**, 13239–13248 (2012).
- Takashima, S. et al. Molecular cloning and expression of a sixth type of alpha 2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans. *J. Biol. Chem.* **277**, 24030–24038 (2002).

53. Teinturier-Lelièvre, M. et al. Molecular cloning and expression of a human hST8Sia VI (alpha2,8-sialyltransferase) responsible for the synthesis of the diSia motif on O-glycosylproteins. *Biochem. J.* **392**, 665–674 (2005).
54. Hevia, A., Delgado, S., Margolles, A. & Sánchez, B. Application of density gradient for the isolation of the fecal microbial stool component and the potential use thereof. *Sci. Rep.* **5**, 16807 (2015).
55. Wang, L., Lu, Z., Allen, K. N., Mariano, P. S. & Dunaway-Mariano, D. Human symbiont *Bacteroides thetaiotaomicron* synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). *Chem. Biol.* **15**, 893–897 (2008).
56. Bardor, M., Nguyen, D. H., Diaz, S. & Varki, A. Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. *J. Biol. Chem.* **280**, 4228–4237 (2005).
57. Chisada, S. et al. Zebrafish and mouse alpha2,3-sialyltransferases responsible for synthesizing GM4 ganglioside. *J. Biol. Chem.* **284**, 30534–30546 (2009).
58. Nakamura, K. Gangliosides of dogfish (*Squalus acanthias*) brain. *Neurochem. Int.* **30**, 593–604 (1997).
59. Nakamura, K., Kojima, H., Suzuki, M., Suzuki, A. & Tamai, Y. Novel polysialogangliosides of skate brain structural determination of tetra, penta and hexasialogangliosides with a NeuAc-GalNAc linkage. *Eur. J. Biochem.* **267**, 5198–5208 (2000).
60. Ando, S. & Yu, R. K. Isolation and characterization of two isomers of brain tetrasialogangliosides. *J. Biol. Chem.* **254**, 12224–12229 (1979).
61. Asahina, S. et al. Involvement of the alpha2,8-polysialyltransferases II/STX and IV/PST in the biosynthesis of polysialic acid chains on the O-linked glycoproteins in rainbow trout ovary. *J. Biochem.* **140**, 687–701 (2006).
62. Adolf, B. et al. Conserved and acquired features of adult neurogenesis in the zebrafish telencephalon. *Dev. Biol.* **295**, 278–293 (2006).
63. Davies, L. R. L. et al. Metabolism of vertebrate amino sugars with N-glycolyl groups: resistance of alpha2,8-linked N-glycolylneuraminic acid to enzymatic cleavage. *J. Biol. Chem.* **287**, 28917–28931 (2012).
64. Malych, Y. N., Shaw, L. & Schauer, R. The role of CMP-N-acetylneuraminic acid hydroxylase in determining the level of N-glycolylneuraminic acid in porcine tissues. *Glycoconj. J.* **15**, 885–893 (1998).
65. Nakao, T., Kon, K., Ando, S. & Hirabayashi, Y. A NeuGc-containing trisialoganglioside of bovine brain. *Biochim. Biophys. Acta-Lipids Lipid Metab.* **1086**, 305–309 (1991).
66. Casellato, R., Brocca, P., Li, S. C., Li, Y. T. & Sonnino, S. Isolation and structural characterization of N-acetyl- and N-glycolylneuraminic-acid-containing GalNAc-GD1a isomers, IV4GalNAcIV3Neu5AcII3Neu5GcGgOse4Cer and IV4GalNAcIV3Neu5GcII3Neu5AcGgOse4Cer, from bovine brain. *Eur. J. Biochem.* **234**, 786–793 (1995).
67. Gupta, T. & Mullins, M. C. Dissection of organs from the adult zebrafish. *J. Vis. Exp.* <https://doi.org/10.3791/1717> (2010).
68. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio)* (University of Oregon Press, Institute of Neuroscience, University of Oregon (Eugene, OR) 2007).
69. Chen, J.-Y. et al. Concerted mass spectrometry-based glycomic approach for precision mapping of sulfo sialylated N-glycans on human peripheral blood mononuclear cells and lymphocytes. *Glycobiology* **28**, 9–20 (2018).
70. Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–217 (1984).
71. Aoki, K. et al. Dynamic developmental elaboration of N-linked glycan complexity in the *Drosophila melanogaster* embryo. *J. Biol. Chem.* **282**, 9127–9142 (2007).
72. Hara, S., Takemori, Y., Yamaguchi, M., Nakamura, M. & Ohkura, Y. Fluorometric high-performance liquid chromatography of N-acetyl- and N-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids. *Anal. Biochem.* **164**, 138–145 (1987).
73. Klein, A. et al. New sialic acids from biological sources identified by a comprehensive and sensitive approach: liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology* **7**, 421–432 (1997).
74. Noel, M. et al. Probing the CMP-sialic acid donor specificity of two human beta-d-galactoside sialyltransferases (ST3Gal I and ST6Gal I) selectively acting on O- and N-glycosylproteins. *Chembiochem Eur. J. Chem. Biol.* **18**, 1251–1259 (2017).
75. Gaver, R. C. & Sweeley, C. C. Methods for methanolysis of sphingolipids and direct determination of long-chain bases by gas chromatography. *J. Am. Oil Chem. Soc.* **42**, 294–298 (1965).
76. Song, Y., Kitajima, K. & Inoue, Y. Monoclonal antibody specific to alpha-2->3-linked deaminated neuraminyl beta-galactosyl sequence. *Glycobiology* **3**, 31–36 (1993).
77. Varki, A. & Kornfeld, S. Historical Background and Overview. in *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, New York, 2015).
78. Varki, A. et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324 (2015).

Acknowledgements

This work was supported by the Centre National de La Recherche Scientifique (CNRS), the ANR-2010-BLAN-120401 grant (project Galfish) from the Agence Nationale de la Recherche, SPS Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation G2205 (to K.K.) and Mizutani Foundation for Glycoscience (to Y.G.). We are indebted to the Plateforme d'Analyse des Glycoconjugués (PAGés, <http://plateforme-pages.univ-lille1.fr/>) and to the Research Federation FRABio (Univ.Lille, CNRS, FR 3688, FRABio, Biochimie Structurale et Fonctionnelle des assemblages Biomoléculaires) for providing the scientific and technical environment conducive to achieving this work.

Author contributions

N.Y., J.V., L.Y.C., L.C., S.Y.Y., J.K., and Y.G. performed the experiments; A.H.L., K.A.K., C.S., K.H.K., K.K., and Y.G. designed the study and wrote the article, all authors critically reviewed the article and approved the final manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-06950-3>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Supplementary Informations

Systems glycomics of adult zebrafish identifies organ specific sialylation and glycosylation patterns

Nao Yamakawa^{1,2,†}, Jorick Vanbeselaere^{1,†}, Lan-Yi Chang^{1,3}, Shin-Yi Yu¹, Lucie Ducrocq¹, Anne Harduin-Lepers¹, Junichi Kurata⁴, Kiyoko F. Aoki-Kinoshita⁴, Chihiro Sato², Kay-Hooi Khoo³, Ken Kitajima², Yann Guerardel^{1,*}

¹ Univ. Lille, CNRS, UMR 8576 – UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France

² Bioscience and Biotechnology Center, Nagoya University, Nagoya 464-8601, Japan

³ Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

⁴ Faculty of Science and Engineering, Soka University, Hachioji, Tokyo 192-8577, Japan

Supplementary note 1

Analytical approach - Several technological approaches including LC-MS and MALDI-MS of native, fluorescence-labelled and permethylated glycans provide a reliable access to glycome. Here, straight analysis of permethylated glycans by MALDI-MS/MS was favored because it provides (1) overall sensitivity increase, (2) better discrimination of isomeric structures owing to specific cleavage¹, (3) reliable results for quantification over a wide range of glycosylated molecules²⁻⁴ and (4) a single analytical workflow for the analysis of both glycoproteins and glycolipids derived glycans^{5,6}. All MS data were acquired using standard procedures (DHB 10 mg/mL in MeOH/H₂O for permethylated glycans) that naturally result in the generation of predominant production of sodiated adducts in positive mode for both MS and MS/MS experiment. MALDI-MS spectra for permethylated glycans always produced sodiated species and sodiated parent only produced sodiated fragment ions and vice versa and their patterns are quite distinctive. The only exception to that rule was the B ions resulting from single sialic acid cleavage that generated [M+H]⁺ adducts.

Supplementary note 2

Structural analysis of NGs - Out of the 95 complex and hybrid NGs, 52 were substituted by at least one sialic acids: 41 by Neu5Ac, 12 by Neu5Gc and 5 by Kdn. Five of them were substituted by both Neu5Ac and Neu5Gc, but Kdn-substituted NGs were never substituted by another type of sialic acids (Fig. 3C). Seventy four NGs were substituted by at least one fucose residue, out of which 50 present at least one fucose on a LacNAc motif, and 45 on the chitobiose core. The most prominent fragments associated to CID MS/MS of permethylated glycans are always derived from the B/Y cleavage ions of HexNAc residues. This cleavage yield informative fragments on the presence and location of LacNAc motifs (B ion at m/z 486) that support the common Lewis type antigens (B ion at m/z 660 and 834). The multiple or recurrent cleavages of non-reducing terminal LacNAc epitopes shed light on the branching pattern owing to the presence or not of a remaining O-methyl group. The most obvious use of this is the possibility to distinguish the number of antennae versus the presence of poly-LacNAc motifs and was used here to identify bi-, tri- and tetra-antennary NGs. Then, oxonium-type elimination of the C3-substituent from the GlcNAc residue allows a discrimination of the type-1 (Gal(β 1,3)GlcNAc) versus type-2 (Gal(β 1,4)GlcNAc) motif, which in turn permit to identify Lewis x (Lex, Gal(β 1,4)[Fuc(α 1,3)]GlcNAc), Lewis a (Lea, Gal(β 1,3)[Fuc(α 1,4)]GlcNAc), Lewis y (Ley, Fuc(α 1,2)Gal(β 1,4)[Fuc(α 1,3)]GlcNAc) and Lewis b (Leb, Fuc(α 1,2)Gal(β 1,3)[Fuc(α 1,4)]GlcNAc) epitopes, even from a complex mixture of NGs. Based on these, we exclusively observed type-2 LacNAc motifs and corresponding Lex epitope. No indication of type-1, Lea or Leb epitopes could be gathered so far. Positioning of Fuc residues on either LacNAc branches or chitobiose core were easily distinguished owing to the presence of fucosylated LacNAc based-B ions (m/z at 660, 864, 1225, 1255) on one hand or the presence of fucosylated Man₃GlcNAc₂ Z-ion (m/z at 1303 and 1317) as well as the Y₁ ion at m/z 464 resulting from the cleavage of the fucosylated chitobiose core. High proportion the complex NGs (55 out of 86) showed Z-type fragments at m/z 690, 864, 1225 or 1255 that do not fit with usual LacNAc derived epitopes. Using a combination of MS/MS fragmentation, enzymatic digestion, chemical degradation and GC/MS linkage analysis we have previously demonstrated that these fragments were all derived from the Gal(β 1,4)Gal(β 1,4)GlcNAc motif that could be further fucosylated and sialylated to generate the unusual and so far zebrafish-specific Gal(β 1,4)(NeuAc/Gc(α 2,3))Gal(β 1,4)[Fuc(α 1,3)]GlcNAc epitope^{5,6}. As shown in Supplementary Fig. 14, single NGs can be either substituted exclusively by Neu5Ac, Neu5Gc or a by combination of Neu5Ac and Neu5Gc. In addition, five NGs substituted by one or two Kdn residues were identified at m/z 2809 (NG50 and NG51), 2983 (NG61), 3129 (NG74) and 3187 (NG78). Kdn residue appear to be exclusively localized on the terminal Gal residue of the branch LacNAc, as shown by the observation

of B ion at m/z 806 and C ion at m/z 579 in all corresponding fragmentation patterns (Supplementary Fig. 15).

Finally, substitution by HexNAc₂ motifs, tentatively assigned to LacDiNAc (GalNAc(β1,3/4)GlcNAc) motif, could be observed in 10 NGs (NG11, NG14-15, NG17, NG24, NG28, NG43, NG58-59, NG63) according to successive Z-type loss of external (M-259) and internal HexNAc (M-245) residues. It is noteworthy that we could observe the corresponding release of B-type fragment ion resulting from cleavage of internal HexNAc residue only when the HexNAc₂ motif was sialylated either by Neu5Ac (m/z 888) or Neu5Gc (m/z 918) but not when unsialylated. When sialylated, the sialylation exclusively occurred on the terminal non-reducing HexNAc residues in α2-6 position, as shown by the B-ions at m/z 643 and 673, respectively and in accordance with previous data from Hanzawa and collaborators (Supplementary Fig. 16)⁷.

Supplementary note 3

Structural analysis of OGs - Mass spectrometry sequencing of OGs released by reductive β -elimination established the presence of mucin type OGs, but no O-mannosylated or O-fucosylated glycans that are known to be minor compounds in vertebrates. The nature of the OG core was inferred for each compound from the combination of GalNAc-ol containing Y- and Z-type fragments, in particular the cleavage of Gal(β 1,3)GalNAc-ol and Sia(α 2,6)GalNAc-ol linkages for core-1 (Gal(β 1,3)GalNAc-ol) as well as Gal(β 1,3)GalNAc-ol and GlcNAc(α 1,6)GalNAc-ol linkages for core-2 (Gal(β 1,3)[GlcNAc(β 1,6)]GalNAc-ol). Altogether, about 80% of identified OGs exhibited a core-1 and 15% a core-2. A single low intensity core-3 (GlcNAc(β 1,3)GalNAc-ol) OG OG6 at m/z 936 and core-4 (GlcNAc(β 1,3)[GlcNAc(β 1,6)]GalNAc-ol) OG14 at m/z 1239 were tentatively identified in gill and ovary, respectively. The unusual core-4 OG14 was clearly identified in ovary owing to multiple cleavages of HexNAc-HexNAc linkages that established the nature of the two branches as HexNAc and Fuc(α 1,3)HexNAc-HexNAc (Supplementary Fig. 17). Although we have no formal proof for the identification of HexNAc residues, the wide distribution of LacDiNAc epitopes in both vertebrates and invertebrates coupled to the known biosynthetic pathways of mucin-types O-glycans strongly suggest that the structure OG14 is GlcNAc(β 1,6)[Fuc(α 1,3)GalNAc(β 1,3/4)GlcNAc(β 1,3)]GalNAc-ol. The identification of OG6 was deduced from the MS/MS losses of terminal HexNAc residues as Y/Z-ions at m/z 659/677 and Neu5Ac as Y-ion at m/z 561 which established the HexNAc-[Neu5Ac]GalNAc-ol sequence. It was tentatively assigned to GlcNAc(β 1,3)[Neu5Ac(α 2,6)]GalNAc-ol, but could similarly be core-5, (GalNAc(α 1,3)GalNAc-ol), core-6 (GlcNAc(β 1,6)GalNAc-ol) or core-7 (GalNAc(α 1,3)GalNAc-ol), although these are very rarely observed.

As for NGs, the most common sialic acid observed on OGs was Neu5Ac. Indeed, among the 33 identified OGs, 24 were substituted by at least one sialic acid, including 17 substituted by Neu5Ac, 9 by Neu5Gc and 4 by Kdn (Fig. 5, Supplementary Data 2). Substitution by Kdn was established owing to Y-ions at M-334 in all MS/MS spectra. Signal at m/z 854 was assigned as Gal(β 1,3)[Kdn(α 2,6)]GalNAc-ol based on the observation of M-Kdn at m/z 520, Hex-HexNAc-ol signal at m/z 520 and $[M+H]^+$ and $[M+Na]^+$ Kdn signals at m/z 335 and 357. Absence of Kdn-Hex B-ion signal at m/z 579 makes it very unlikely that Kdn substituted Gal residue in OG1 (Supplementary Fig. 18). Signal at m/z 1174 was shown to contain two isomers OG10 and OG11 with Kdn₂Hex₁HexNAc-ol compositions that were attributed to Kdn(α 2,8)Kdn(α 2,6)[Gal(β 1,3)]GalNAc-ol (OG10) and to Kdn(α 2,3)Gal(β 1,3)[Kdn(α 2,6)]GalNAc-ol (OG11). The Kdn-HexNAc-ol and Kdn-Hex fragments from OG11 were unambiguously identified owing to C/Z ions pair generated by the cleavage of the Gal(β 1,3)GalNAc-ol motif at m/z 579 and 618, whereas Kdn dimer was identified in OG10 owing to

the observation of the Kdn₂ fragment at m/z 677 and of the secondary cleavage of Kdn(α 2,6)GalNAc-ol motif at m/z 604 (Supplementary Fig. 18). The substitution position of Kdn could not be *de novo* established, but was tentatively attributed as (α 2,8) based on previous identification of Neu5Ac(α 2,8)Neu5Ac linkage in zebrafish OGs⁸. As commonly found in vertebrates, brain glycans are highly sialylated and present a higher proportion of oligosialylated OGs than any other organ. Irrespective of the nature of the sialic acid, the sialylation occurred either on the C6 of GalNAc-ol, on the C3 of Gal residue or on another sialic acid, presumably through an α 2-8 linkage as previously established, to generate a polysialic acid sequence⁸. Only Kdn was not identified within a polysialylated sequence (Fig. 5B). Another shared feature with NGs is the presence of Gal(β 1,4)Gal motif that was identified in core-2 glycans within the zebrafish specific Gal(β 1,4)Gal(β 1,4)[Fuc(α 1,3)]GlcNAc (OG23 at m/z 1361) and the Gal(β 1,4)[Neu5Ac/Gc(α 2,3)]Gal(α 1,4)[Fuc(α 1,3)]GlcNAc (OG30/31 at m/z 1722/1752) motifs (Supplementary Fig. 17). This strongly suggests that the set of enzymes involved in the synthesis of zebrafish epitopes can use both N- and OGs as substrates. The GalNAc(β 1,4)Gal(β 1,3)GalNAc-ol sequence which structure was previously described in the Fuc(α 1,3)GalNAc(β 1,4)[Neu5Ac/Gc(α 2,3)]Gal(β 1,3)GalNAc-ol OGs by a combination of NMR and MS/MS in zebrafish embryo was again identified in many OGs isolated from several organs⁹. It was associated to non-sialylated glycans (as in OG7 at m/z 953), non-fucosylated glycans (as in OG9 at m/z 1141) or fully sialylated and fucosylated (as in OG18 and OG20 at m/z 1315 and 1345) as previously described (Supplementary Fig. 17c and 17d)⁹. Ten OGs out of 33 were shown to be substituted by at least one fucose residue for more than 80% associated to Fuc(α 1,3)GalNAc(β 1,4)[Neu5Ac/Gc(α 2,3)]₀₋₁Gal and Gal(β 1,4)[Neu5Ac/Gc(α 2,3)]₀₋₁Gal(α 1,4)[Fuc(α 1,3)]GlcNAc motifs. Alternatively, it was associated to the terminal non-reducing residue of a HexNAc residue of LacdiNAc in two OGs observed in ovary (OG14) and gill (OG27).

Supplementary note 4

Structural analysis of glycosphingolipids - Monosaccharide composition analysis by GC/MS and LC/MS of total glycolipids established the presence of galactose (Gal), glucose (Glc), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), fucose (Fuc) and sialic acids (Neu5Ac, Neu5Gc and Kdn) (data not showed). The nature and of sphingoid bases was established by GC/MS analysis following methanolysis and TMS derivatization GSLs. In particular, d18:0 sphinganine, C18:1 sphingosine and t18:0 phytosphingosine could be distinguished owing to their specific retention times and EI/MS fragmentation patterns. Overall, the analysis of the lipid moiety by GC/MS showed that the ceramide moiety of GSLs was mainly constituted by a combination of sphingosine and phytosphingosine with C16:0, C18:0 and C24:1 fatty acyl chains.

CID MS/MS fragmentation of permethylated GSLs cleaves the Hex1-1'Cer linkage which permits to establish the composition of intact ceramides and thus evaluate the heterogeneity of lipid moiety associated to oligosaccharides. Overview of the 172 identified compounds established that the GSLs were substituted by a heterogeneous combination of sphingoid bases and fatty acids, with three major combinations identified as [d18:1, 16:0] at m/z 548, [d18:1, 24:1] at m/z 658 and [d18:1, 24:0] at m/z 660. However, the ceramide composition, as well as the carbohydrate composition, was to a certain extent species-specific.

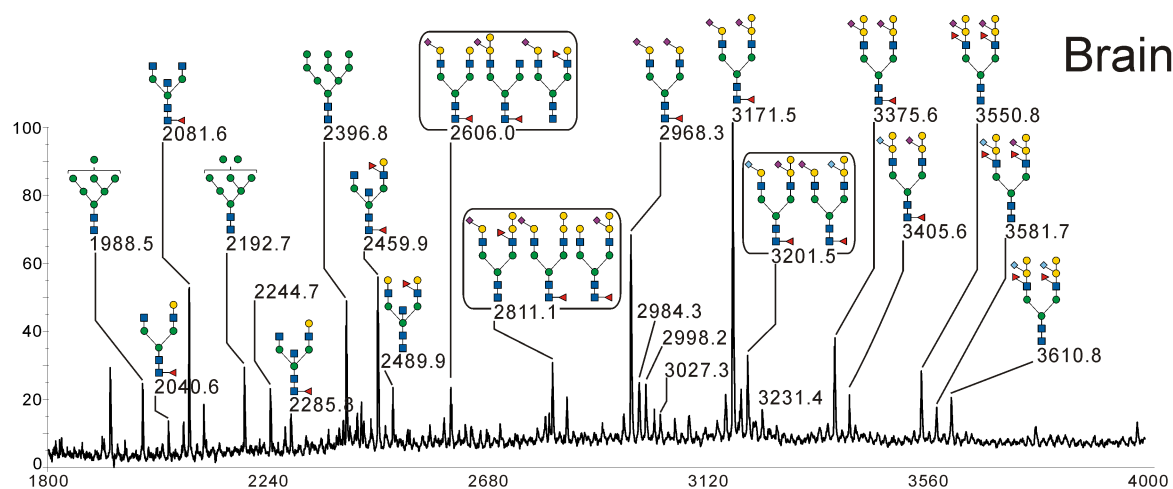
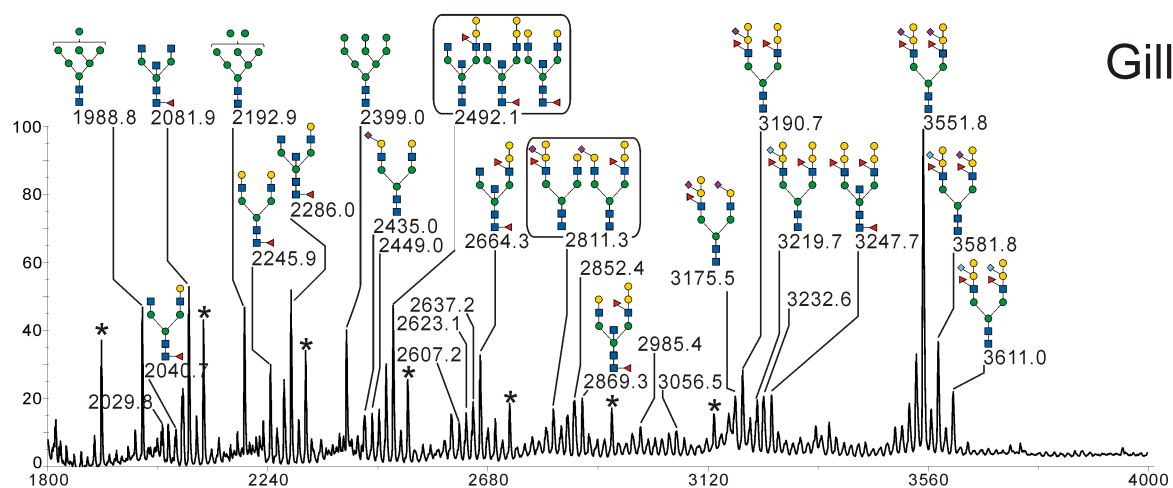
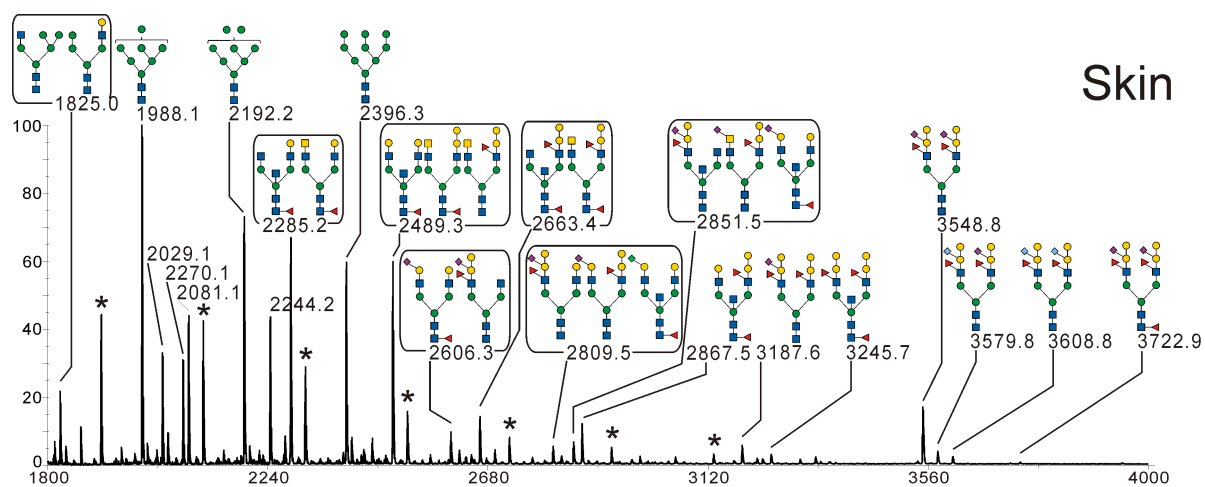
The Gala series was represented by four different glycan moieties Gal(β 1,1')Cer (G1), Neu5Ac(α 2,3)Gal(β 1,1')Cer (NAcGM4, G2), Neu5Gc(α 2,3)Gal(β 1,1')Cer (NGcGM4, G3) and Kdn(α 2,3)Gal(β 1,1')Cer (KdnGM4, G4) that exhibited a wide lipid heterogeneity. The Hemato series was composed of lactoceramide (LacCer) that could be further substituted by one, two or three sialic acids, presumably substituted in (α 2,8). Mono- and di-sialylated LacCer (GL55-59/G6 and GL60-62/G7) were exclusively substituted by Neu5Ac, but trisialylated LacCer (GT4, GL63-66, G8-11) could also incorporate a single Neu5Gc residue in all three positions of the oligosialylated chain. The presence of a homogenous Neu5Ac₃ linear chain was easily established by the observation of Y-type fragments at m/z 1024 [LacCer], m/z 1385 [Neu5Ac₁LacCer] and m/z 1746 [Neu5Ac₂LacCer] and of B-type fragments at m/z 376 [Neu5Ac₁], m/z 759 [Neu5Ac₂] and m/z 1120 [Neu5Ac₃], which permitted to identify GL63 at m/z 2123 as Neu5Ac(α 2,8)Neu5Ac(α 2,8)Neu5Ac(α 2,3)Gal(β 1,4)Glc(β 1,1')Cer⁵⁷⁶ (Supplementary Fig. 19a). By comparison, the combination of Y-type fragments at m/z at 1024 [LacCer], 1385 [Neu5Ac₁LacCer], 1415 [Neu5Gc₁LacCer], 1746 [Neu5Gc₂LacCer], 1776 [Neu5Gc₁cNeu5Gc₁LacCer] and of B-type fragments at m/z 376 [Neu5Ac₁], 406 [Neu5Gc₁], 759 [Neu5Ac₂], 789 [Neu5Gc₁Neu5Ac₁] and 1150 [Neu5Gc₁Neu5Ac₂] from the parent ion at 2153 established the presence of a mixture of Neu5Gc(α 2,8)Neu5Ac(α 2,8)Neu5Ac(α 2,3)Gal(β 1,4)Glc(β 1,1')Cer⁵⁷⁶,

$\text{Neu5Ac}(\alpha 2,8)\text{Neu5Gc}(\alpha 2,8)\text{Neu5Ac}(\alpha 2,3)\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}^{576}$ and
 $\text{Neu5Ac}(\alpha 2,8)\text{Neu5Ac}(\alpha 2,8)\text{Neu5Gc}(\alpha 2,3)\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}^{576}$ (Supplementary Fig. 19b). The extension of Hemato series to Ganglio series generated a large family of GSLs ranging from widely distributed mono-sialylated GM2 like compounds $\text{GalNAc}(\beta 1,4)[\text{Neu5Ac/Gc}(\alpha 2,3)]\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}$ to multisialylated GA1 substituted by up to five sialic acids. Fragmentation analyses of multisialylated compounds clearly established the presence of multiple isomers differing by the extension of the sialic acid chains and their respective positions. As an example, the MS/MS analysis of GT1 at m/z 2572 in brain showed that it was a mixture of $\text{Neu5Ac}(\alpha 2,3)\text{Gal}(\beta 1,3)\text{GalNAc}(\beta 1,4)[\text{Neu5Ac}(\alpha 2,8)\text{Neu5Ac}(\alpha 2,3)]\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}^{576}$ (GT1b, GL96) and $\text{GalNAc}(\beta 1,4)\text{Gal}(\beta 1,3)[\text{Neu5Ac}(\alpha 2,8)\text{Neu5Ac}(\alpha 2,8)\text{Neu5Ac}(\alpha 2,3)]\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}^{576}$ (GT1c, GL97) (Supplementary Fig. 12a). Surprisingly, we identified a so far un-described isomer of pentasialylated GP1 (GL106) characterized by the substitution of the internal Gal residue by a Neu5Ac_4 linear chain. The presence of this motif was unambiguously established owing to the fragmentation of signal at m/z 3294 that generated an intense series of Y/B ions at m/z 376/22919, 759/2558, 1120/2197 and 1482/1835 resulting from the recurrent cleavage of the Neu5Ac_4 chain (Supplementary Fig. 19c). The respective positions of the single Neu5Ac residue and the Neu5Ac_4 chain on the terminal non-reducing and the internal Gal residues were then established owing to the Y/B ions pair at 847/2469. This compound $\text{Neu5Ac}(\alpha 2,3)\text{Gal}(\beta 1,3)\text{GalNAc}(\beta 1,4)[[\text{Neu5Ac}(\alpha 2,8)]_3\text{Neu5Ac}(\alpha 2,3)]\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}$, which we named GP1d (GL106) in accordance with the nomenclature proposed by Svennerholm, was identified along the classical GP1c (GL105) $\text{Neu5Ac}(\alpha 2,8)\text{Neu5Ac}(\alpha 2,3)\text{GalNAc}(\beta 1,4)[[\text{Neu5Ac}(\alpha 2,8)]_2\text{Neu5Ac}(\alpha 2,3)]\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}$. Altogether, we identified members of GA1, GM2, GM1, GM1b, GD1a, GD1b, GT1b, GT1c, GQ1c, GP1c and GP1d within the Ganglio family. Finally, as observed for GT4, tri-sialylated GT1b (GL98-100) at m/z 2602 and tetra-sialylated GQ1c (GL103-104) at m/z 2964 appeared to be substituted by both Neu5Ac and Neu5Gc in different positions to generate multiple isomers (Supplementary Fig. 12). However, Neu5Gc containing isomers were always present in smaller quantities than their respective Neu5Ac equivalents. Along Hemato and Ganglio series, we identified by MS/MS a homogeneous family of GSLs that was characterized by a pentasaccharide linear sequence HexNAc-Hex-HexNAc-Gal-Glc-Cer. These compounds were in majority sialylated on the internal galactose residue, as established by the Y ion pair at m/z 1470 for GL118 at m/z 2353 (with cer^{660}) and fucosylated on the terminal non-reducing HexNAc residue as established by the recurrent B ions at m/z 434, 678 and 905 to generate the sequence deHex-HexNAc-Hex-HexNAc-[NeuAc/Gc]-Gal-Glc-Cer (GL113-121)

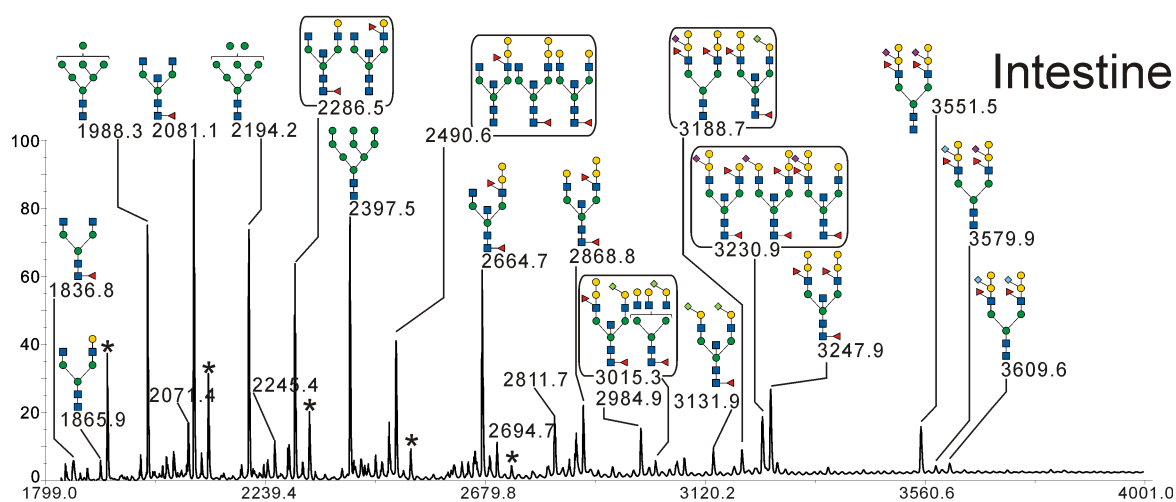
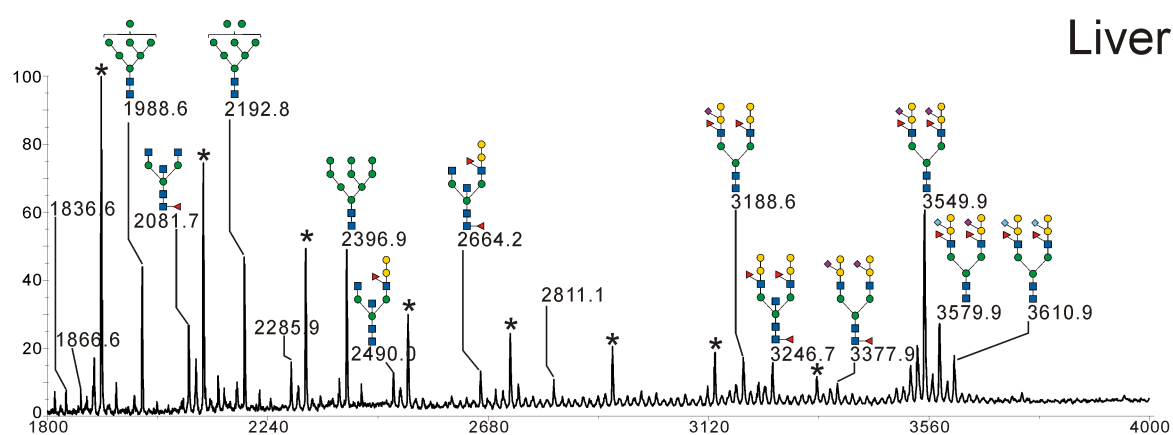
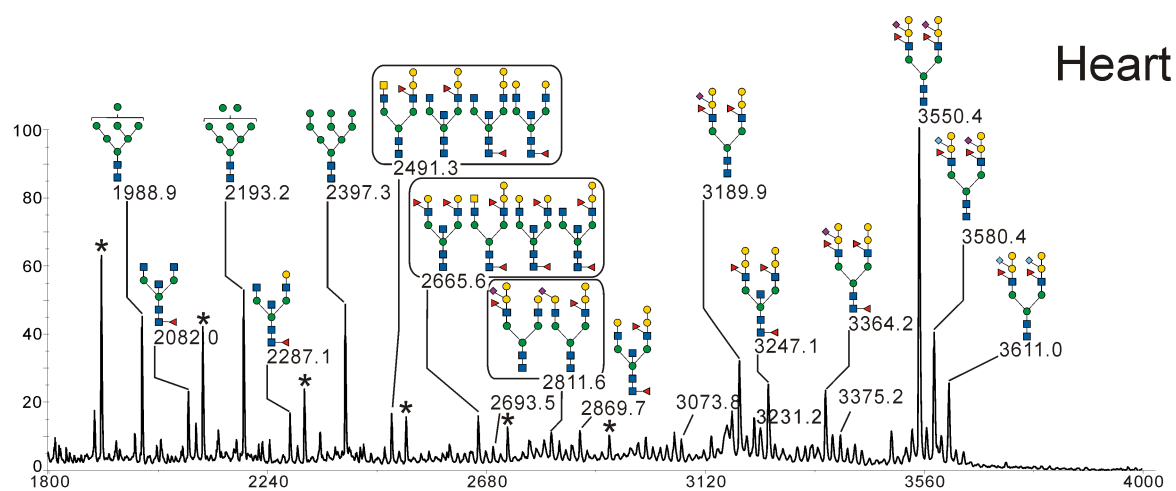
(Supplementary Fig. 20a). This particular sequence has been previously identified in salmon kidney by a combination of MS/MS and NMR as the extended Ganglio series GSLS $\text{Fuc}(\alpha 1,3)\text{GalNAc}(\beta 1,4)\text{Gal}(\beta 1,3)\text{GalNAc}(\beta 1,4)\text{Gal}(\beta 1,4)\text{Glc}(\beta 1,1')\text{Cer}^{10}$. Here, we have no formal proof for establishing the monosaccharides anomeries and linkages, however the MS/MS fragmentation pattern is in total agreement with the previously established sequence. In zebrafish, analogues of this compound were identified substituted either by Neu5Ac or by Neu5Gc (GL120-121) but not by Kdn, and also identified in non-fucosylated (GL109-110) m/z at 2066 and 2096) and non sialylated (GL107-108 at m/z 1879) forms (Supplementary Fig. 20b). Finally, another very heterogeneous family of GSLs with carbohydrate moieties ranging from four to thirteen monosaccharides was identified. They all have in common the HexNAc-[HexNAc]-Hex-Hex tetrasaccharide core as observed in the simple GSL GL 123 at m/z 1501. From the tetrasaccharide core a single HexNAc residue may be elongated by multiple hexoses, N-acetylhexosamine and fucose residues up to thirteen monosaccharides as observed in GL 171 at m/z 3545 in ovary. This peculiar HexNAc₂Hex₂ core has been independently described in two fishes, the English sole and the striped mullet, as the hybrid neolacto-ganglio series (LcGgCer), in which it appears to be the major type of GSLs^{11,12}. So far, its identification is restricted to teleost fishes as LcGgCer-based GSLs have never been observed in other animal branch to our best knowledge. All compounds exhibited a common pentasaccharide core structure $\text{Gal}(\beta 1,4)\text{GlcNAc}(\beta 1,3)[\text{GalNAc}(\beta 1,4)]\text{Gal}(\beta 1,4)\text{Glc}$ that is exclusively elongated from the GlcNAc, but never from the GalNAc residue. In the mullet, the external Gal residue could be substituted by both $\text{GalNAc}(\beta 1,4)$ and $\text{Neu5Ac}(\alpha 2,3)$ residues¹¹, whereas it could be substituted by $\text{Gal}(\alpha 1,3)$ in salmon¹³. Both features could be observed in zebrafish owing to the identification of GL128 at m/z 2066 in gill, testis and ovary and GL149 at m/z 2186 in ovary which MS/MS fragmentation patterns are totally compatible with previously identified $\text{Neu5Ac}(\alpha 2,3)\text{Gal}(\beta 1,4)\text{GlcNAc}(\beta 1,3)[\text{GalNAc}(\beta 1,4)]\text{Gal}(\beta 1,4)\text{Glc-Cer}$ and $\text{GalNAc}(\beta 1,3)\text{Gal}(\alpha 1,3)\text{Gal}(\beta 1,4)\text{GlcNAc}(\beta 1,3)[\text{GalNAc}(\beta 1,4)]\text{Gal}(\beta 1,4)\text{GlcCer}$ (Supplementary Fig. 21a). Other members of the LcGgCer showing a fucosylated HexNAc residue in terminal non reducing position was identified in several organs (GL130-148). As observed in the MS/MS fragmentation patterns of GL131, 134 and 135, the Y-type fragments at m/z 1274, 1330 and 1345 and the Y/Y-type secondary fragments at m/z 1014, 1070 and 1084 all associated to the B-type fragment at m/z 905 established the presence of the LcGg core in these GSLs (Supplementary Fig. 21b). The presence of the terminal deHex-HexNAc and the internal LacNAc disaccharide were unambiguously established by the observation of B-type fragment at m/z 434 and secondary fragment at m/z 472. The exact nature of the terminal deHex-HexNAc motif could not be directly deduced from the MS/MS data, but the apparent wide distribution of the terminal $\text{GalNAc}(\beta 1,3)$ and $\text{Fuc}(\alpha 1,3)\text{GalNAc}(\beta 1,3)$ epitopes in

GSLs from several fish species including English sole (*Parophrys vetulus*), Pacific salmon (*Oncorhynchus keta*) and striped mullet (*Mugil cephalus*) strongly suggests that these GSLs are made of the Fuc(α 1,3)GalNAc(β 1,3)Gal(β 1,4)GlcNAc(β 1,3)[GalNAc(β 1,4)]Gal(β 1,4)GlcCer sequence. In GL147 and 148 observed at m/z 2330 and 2386, this sequence could be further fucosylated on the internal GlcNAc residue, as demonstrated by the B-type fragments at 1079, 678 and 434 (Supplementary Fig. 21c). The secondary fragment at m/z 646 generated by the loss of the terminal Fuc(α 1,3)GalNAc disaccharide from the deHex₂HexNAc₂Hex₁ fragment at m/z 1079 further confirmed that the additional Fuc is not located on the terminal GalNAc residue, which strongly suggests GL147/148 are made of Fuc(α 1,3)GalNAc(β 1,3)Gal(β 1,4)[Fuc α 1,3)]GlcNAc(β 1,3)[GalNAc(β 1,4)]Gal(β 1,4)GlcCer^{580/636} sequences. The LcGgCer family comprised also 21 GSLs with 8 different carbohydrate moieties originating from the elongation of the Gal(β 1,4)GlcNAc(β 1,3) branch which sequences that have never been observed (GL150-171). Indeed, the Gal(β 1,4)GlcNAc(β 1,3) may be elongated a combination of one or two Hex, HexNAc and Fuc residues to form HexNAc-HexNAc-Hex-Hex, Hex-Hex-Hex, or HexNAc-Hex-Hex-Hex branches. Furthermore, five of the carbohydrate sequences (G44-45 and G49-51) were capped by Fuc-Fuc motifs as established by the observation of HexNAc₁Fuc₂ and Fuc₂ fragments at m/z 630 and 403 in GL (Supplementary Fig. 21d). The cleavage of Fuc-HexNAc linkage as a C-ion to generate the Fuc₂ fragment at m/z 403 strongly suggests that the Fuc residue is linked in C3-position on the HexNAc, in agreement with the previously described Fuc(α 1,3)GalNAc motif. However, the exact structural features of this complex series of unusual multifucosylated GSLs should be investigated through a combination of NMR, MS/MS and GC/MS analyses that is out of the scope of the present glycomics report as it requires the purification of individual GSLs prior to detailed analysis. This will be the focus of a coming publication.

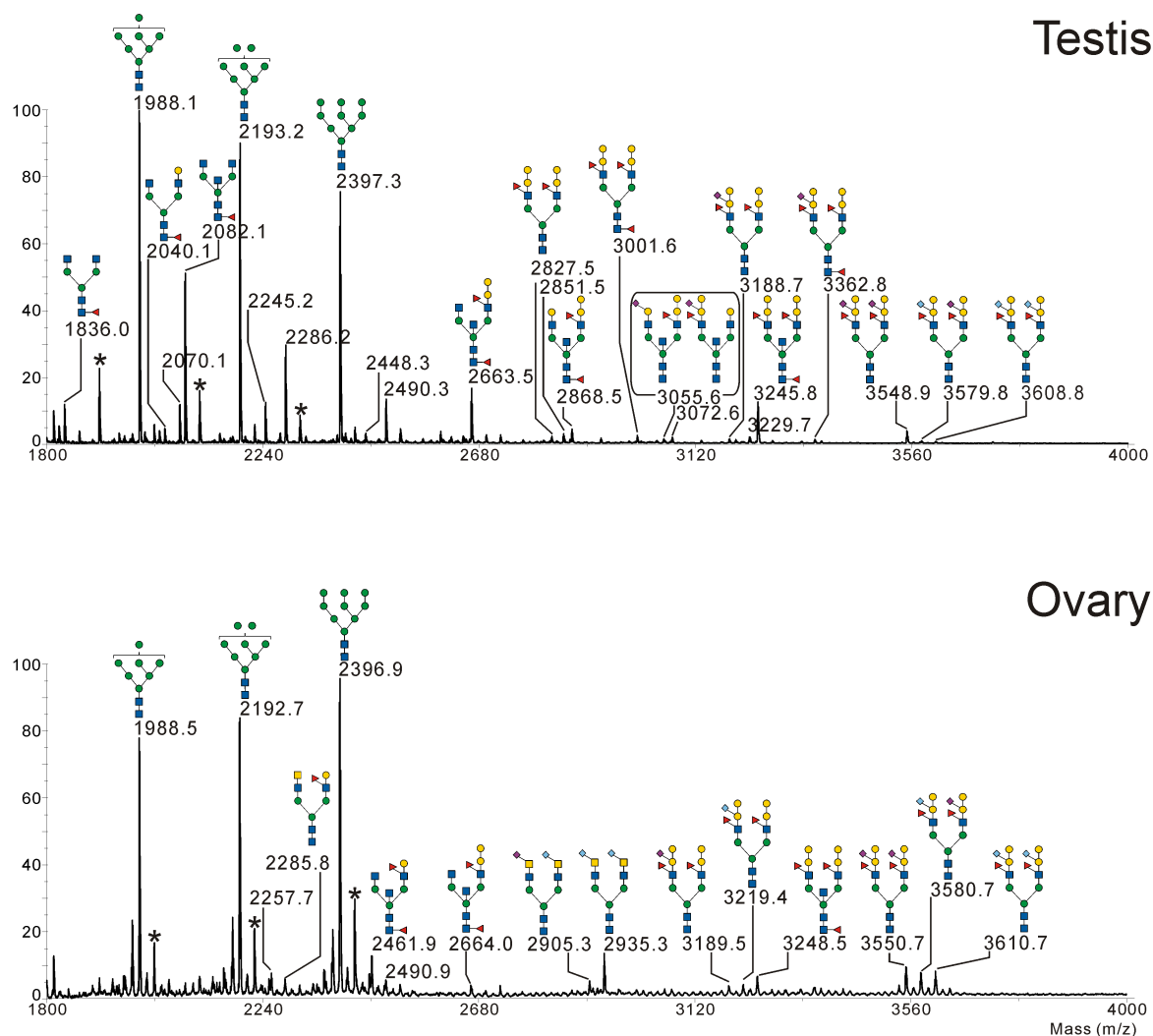
Supplementary Figure 1 (1/3)



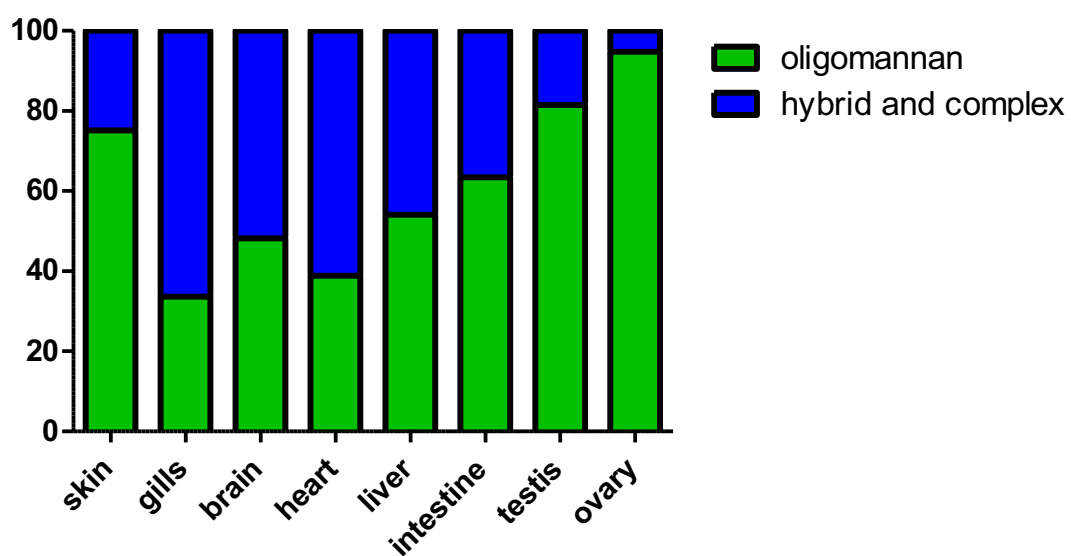
Supplementary Figure 1 (2/3)



Supplementary Figure 1 (3/3)

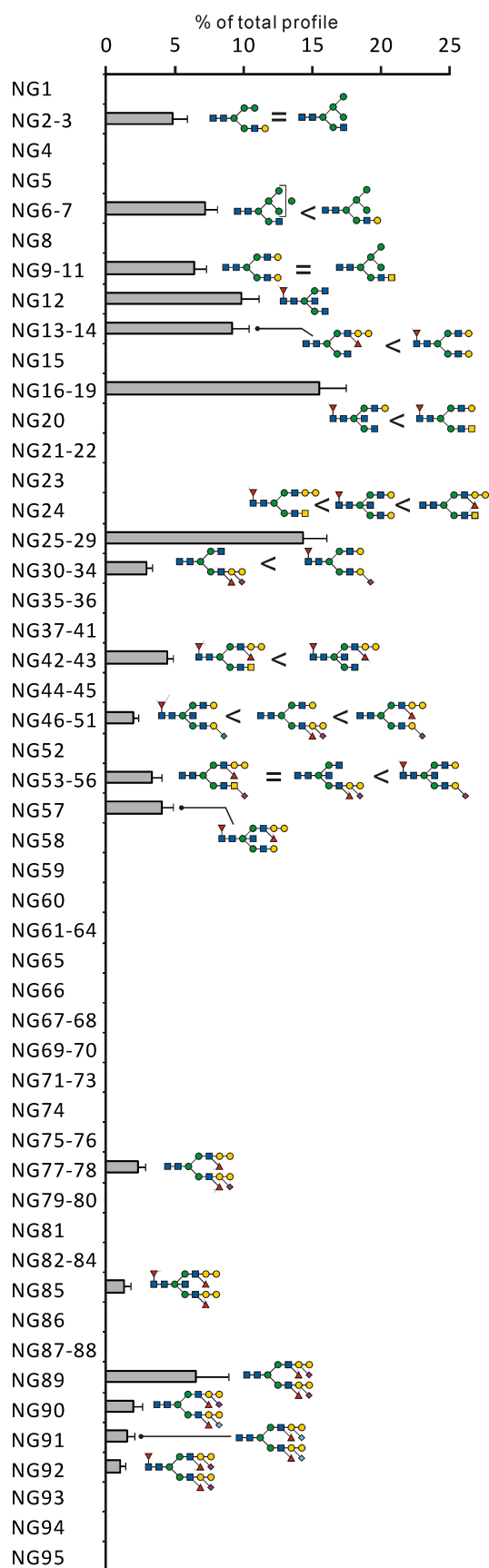


Supplementary Figure 1 - MALDI-TOF-MS spectra of permethylated NGs isolated from the eight organs. Graphical representation is based on accepted conventions for glycans and monosaccharide nomenclature as follows: yellow circle, Gal; yellow square, GalNAc; blue circle, Glc; blue square, GlcNAc; green circle, Man; red triangle, Fuc; purple diamond, Neu5Ac; light blue diamond, Neu5Gc; green diamond, Kdn.

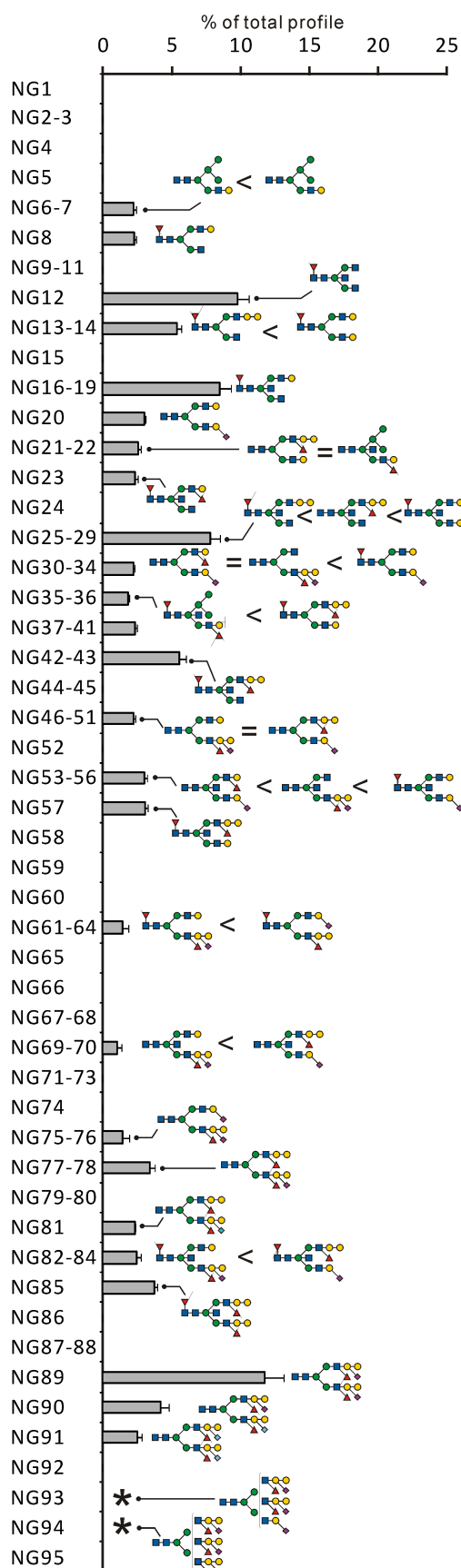


Supplementary Figure 2 - Ratios of oligomannosylated N-glycans (in green) versus hybrid and complex-type N-glycans. NGs were quantified using the integration value of identified m/z signals recorded by MALDI-MS analysis of permethylated glycans and summed according to their nature.

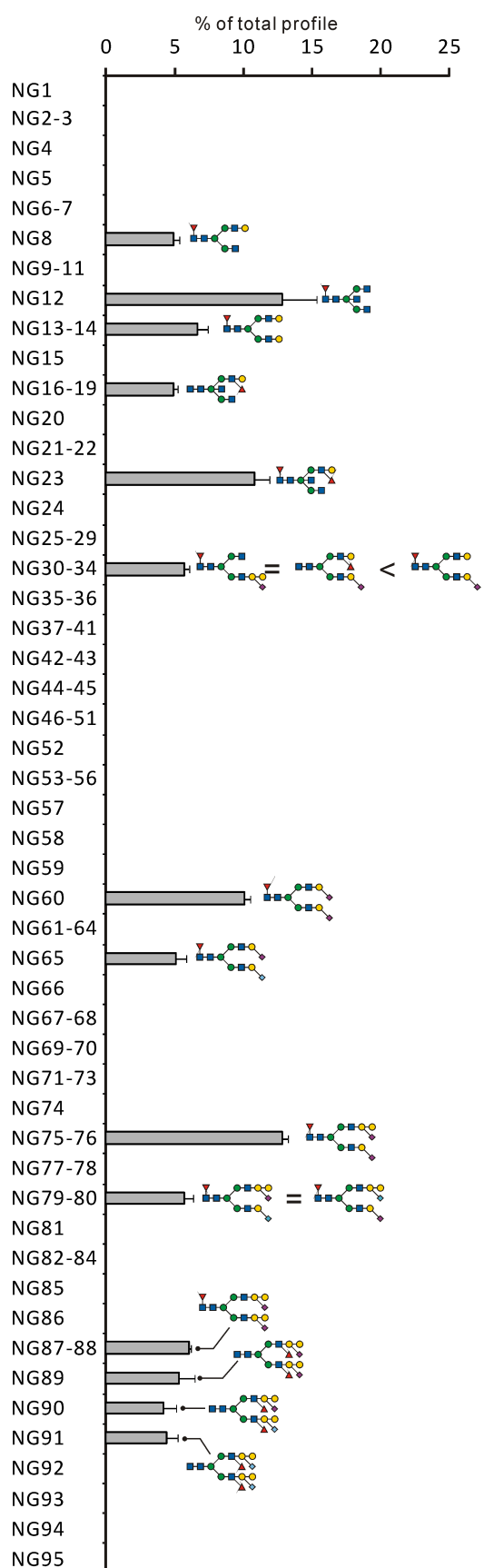
Supplementary Figure 3 (1/4)
Skin



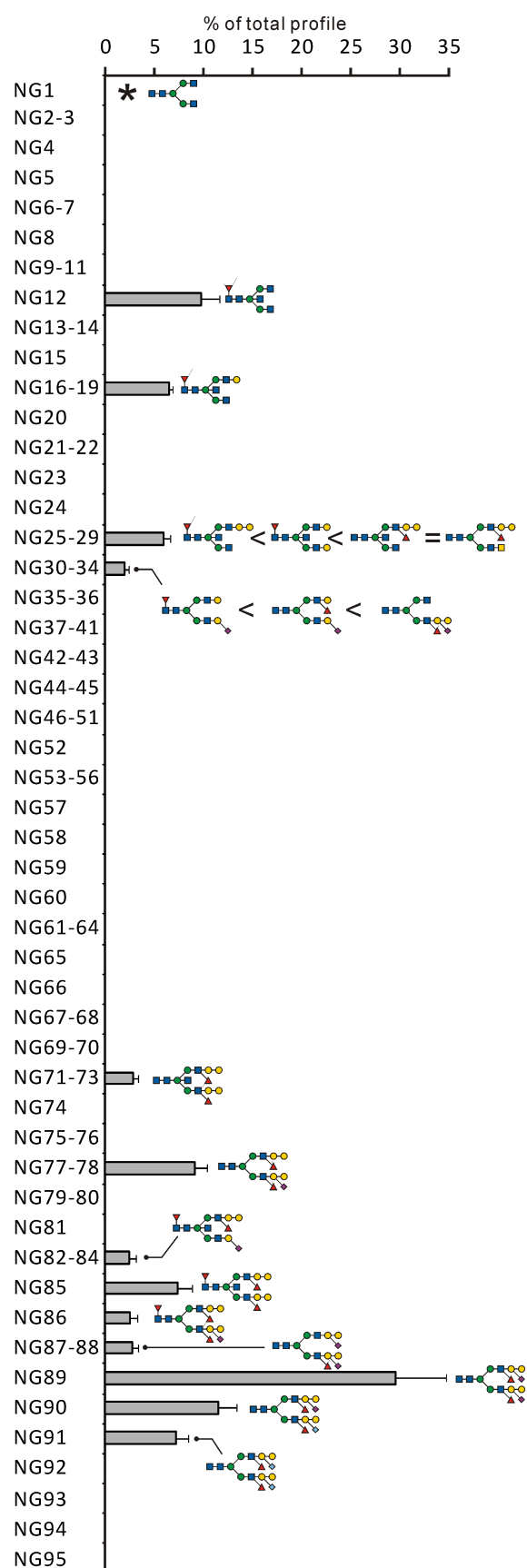
Gill



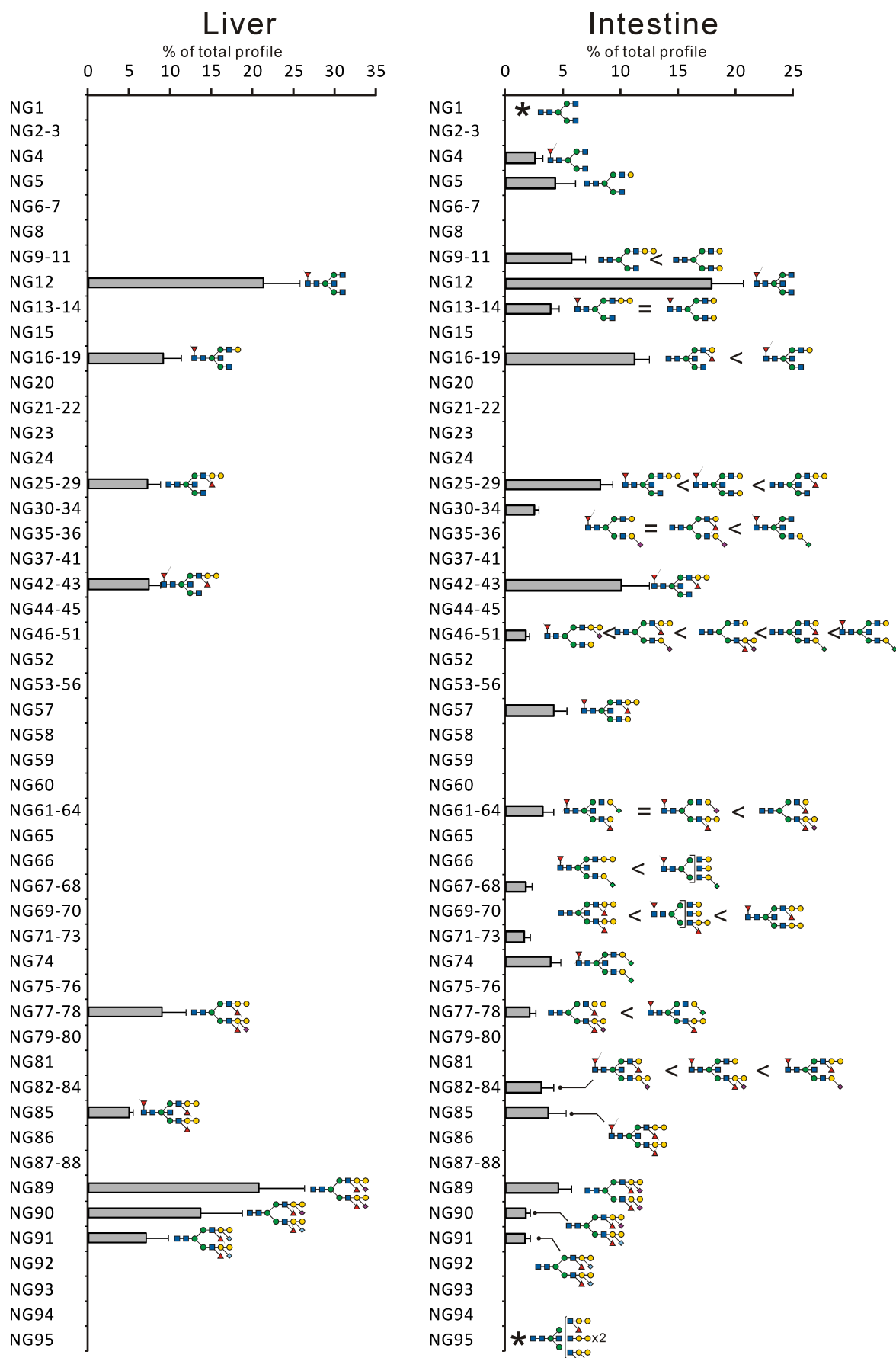
Supplementary Figure 3 (2/4)
Brain



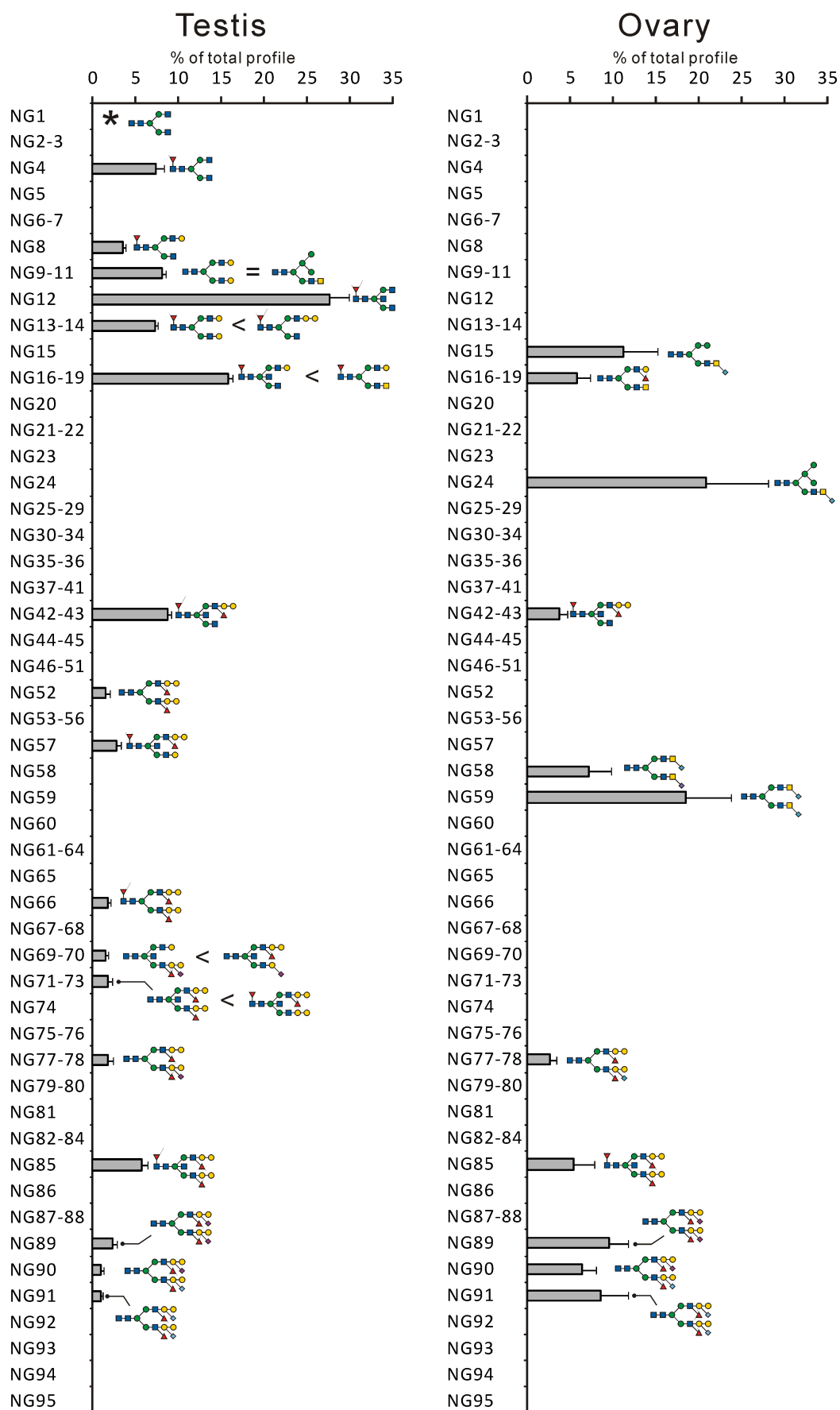
Heart



Supplementary Figure 3 (3/4)



Supplementary Figure 3 (4/4)

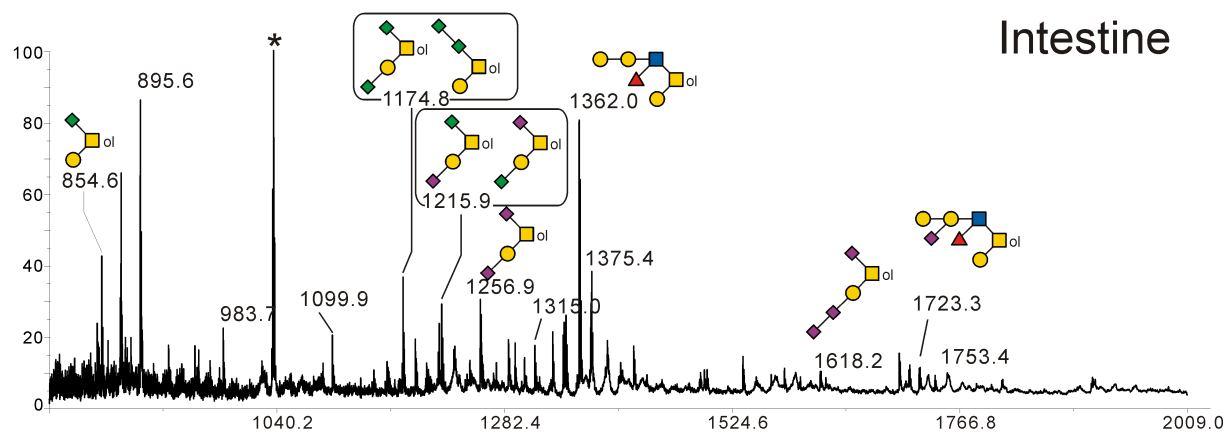
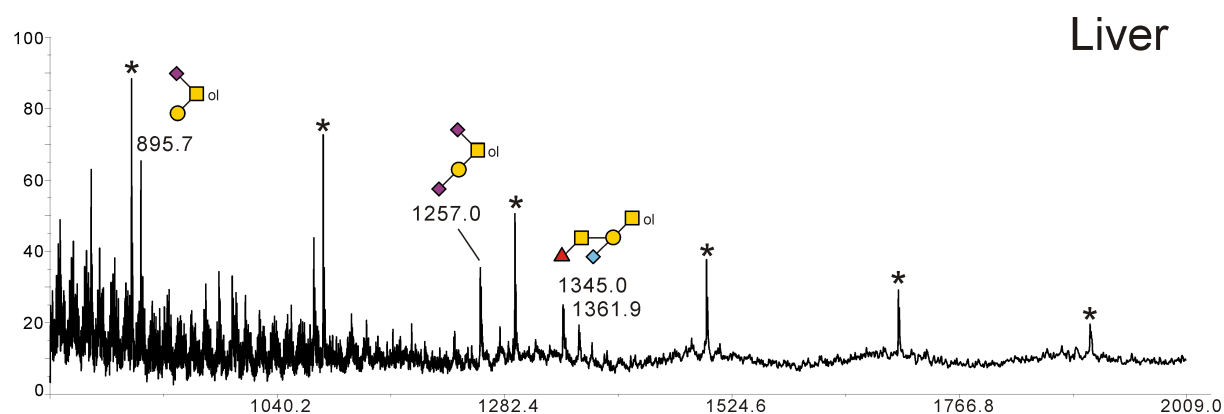
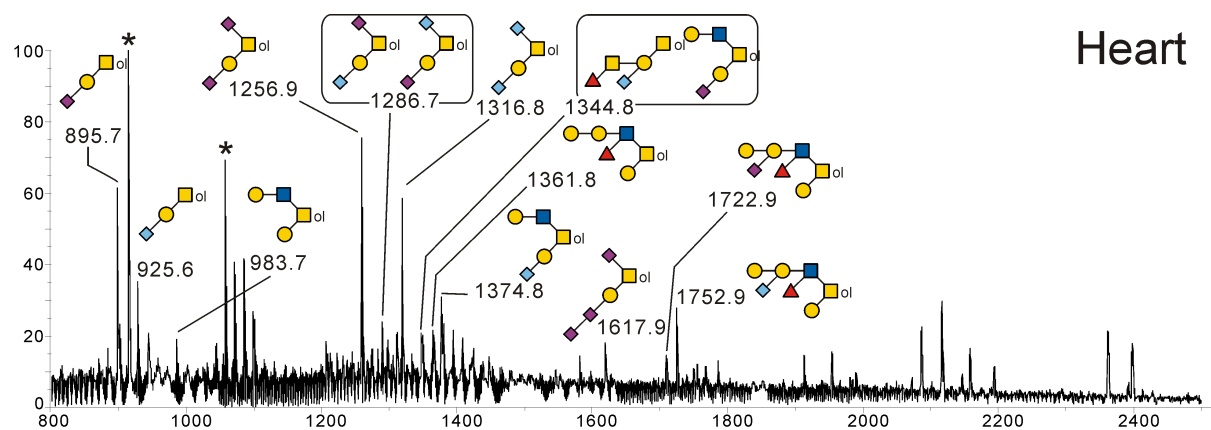


Supplementary Figure 3 - Relative quantification of complex NGs NG1-NG92 (as described in Data 1) in eight organs. NGs were quantified using the integration value of identified m/z signals recorded by MALDI-MS analysis of permethylated glycans as % of the total identified signals (see Fig. 3). Quantification values resulted from three independent experiments. The relative importance of isomeric structures in individual m/z signals was estimated by the intensity of MS/MS signals generated by MALDI-TOF/TOF fragmentation signals. Graphical representation is based on accepted conventions for glycans and monosaccharide nomenclature as follows: yellow circle, Gal; yellow square, GalNAc; blue circle, Glc; blue square, GlcNAc; green circle, Man; red triangle, Fuc; purple diamond, Neu5Ac; light blue diamond, Neu5Gc; green diamond, Kdn. Data are shown as mean \pm SE of three experiments.

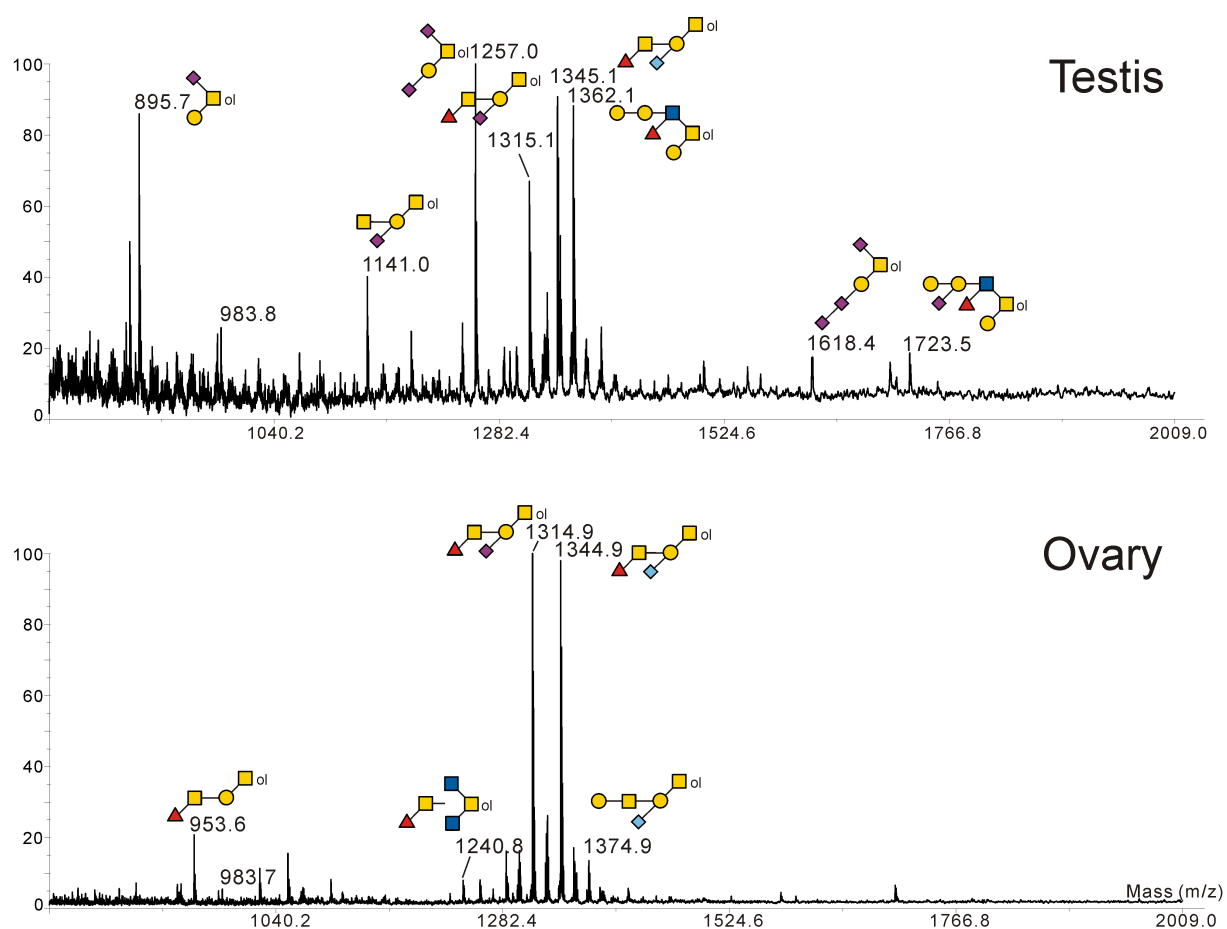
Supplementary Figure 4 (1/3)



Supplementary figure 4 (2/3)

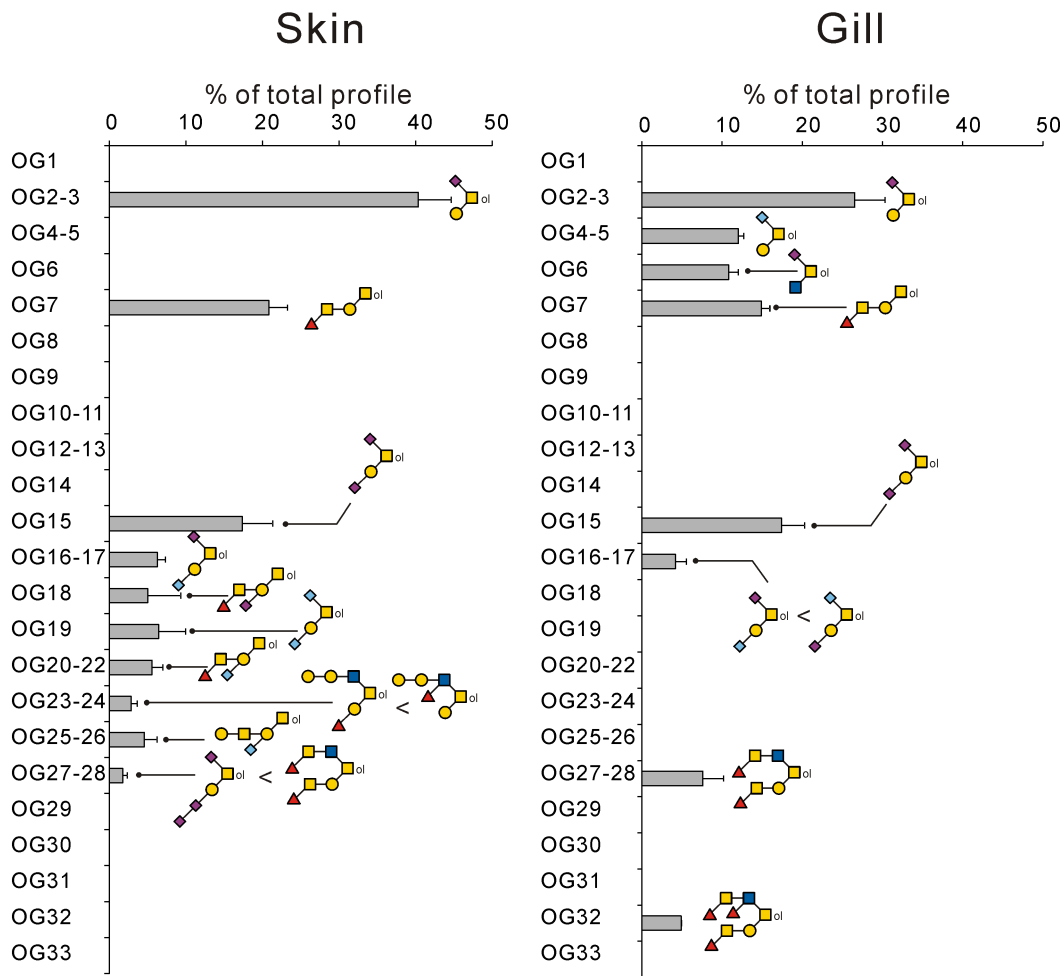


Supplementary figure 4 (3/3)

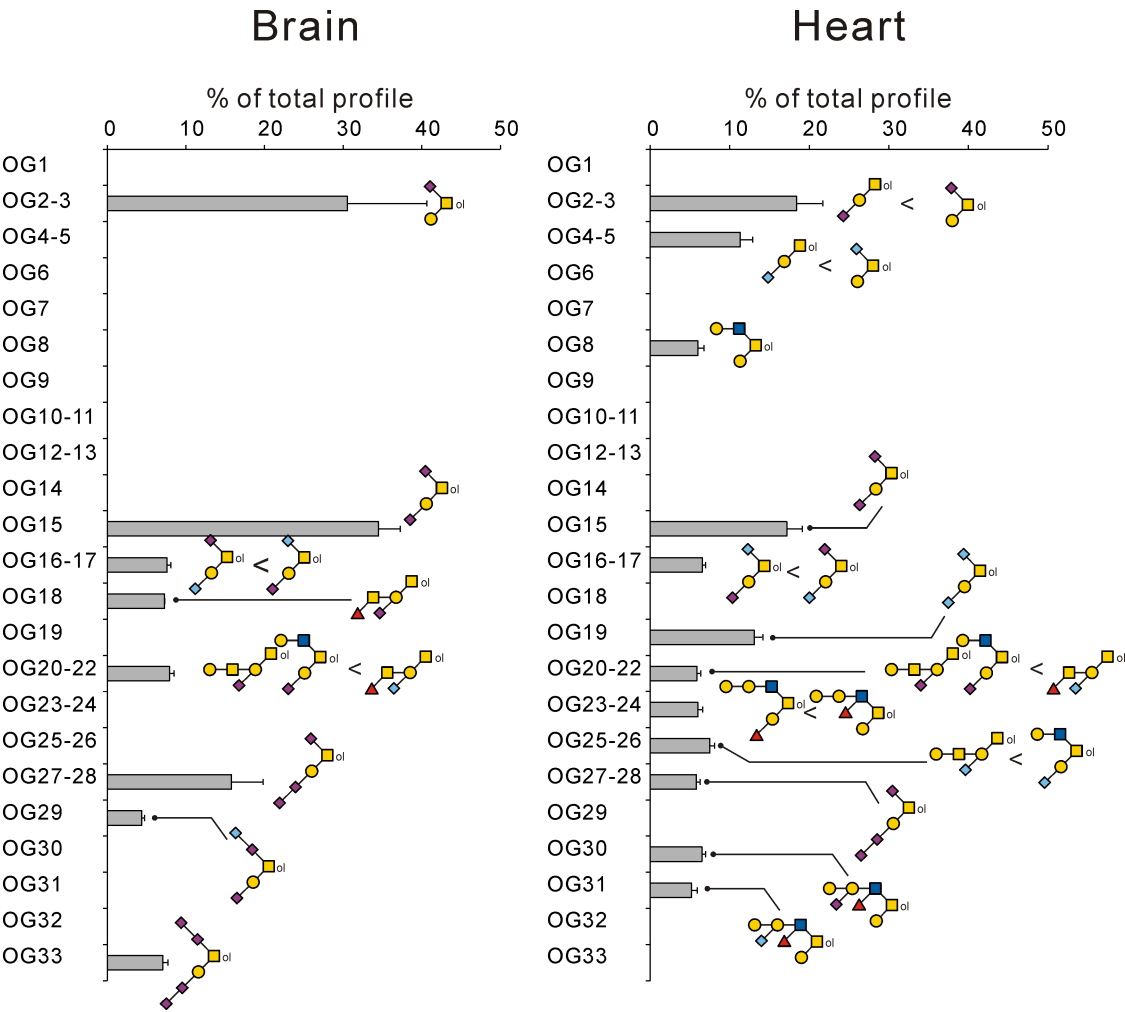


Supplementary Figure 4 - MALDI-TOF-MS spectra of permethylated OGs isolated from the eight organs. Graphical representation is based on accepted conventions for glycans and monosaccharide nomenclature.

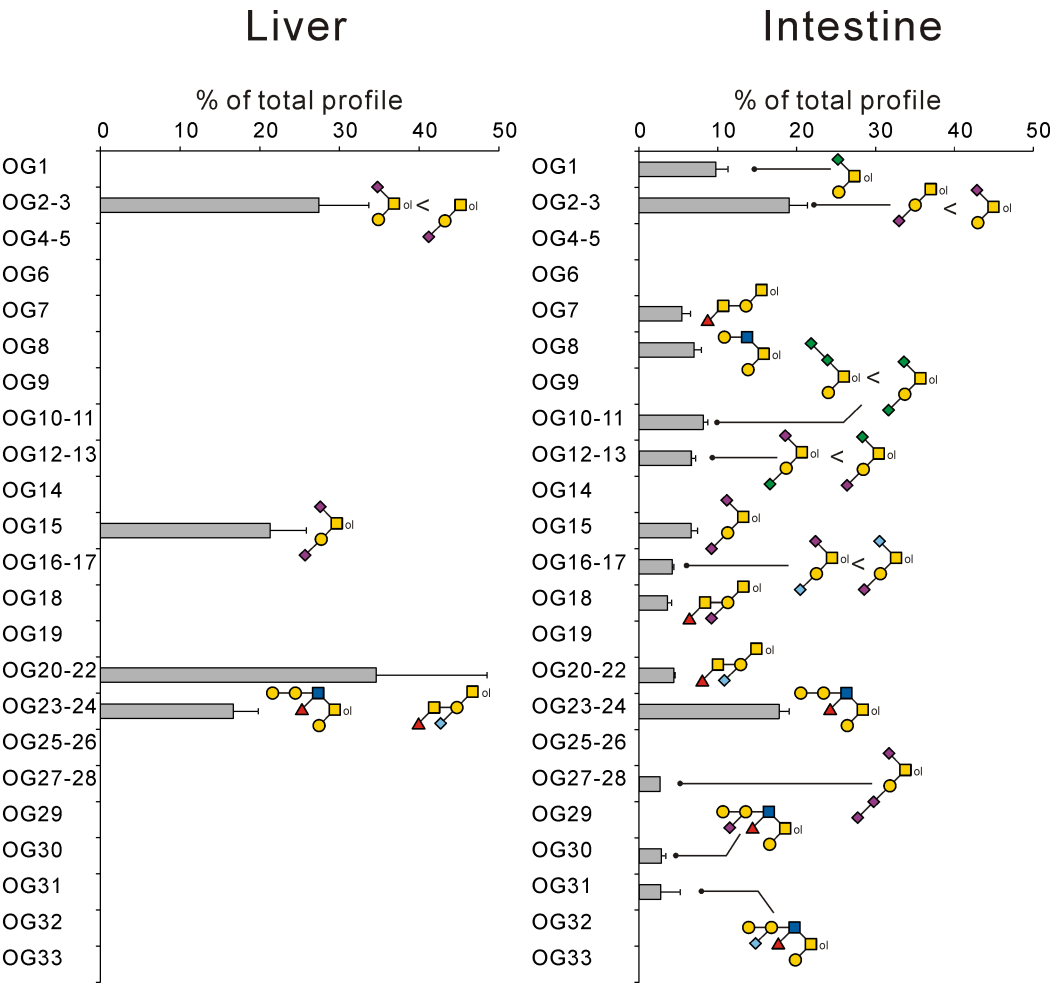
Supplementary Figure 5 (1/4)

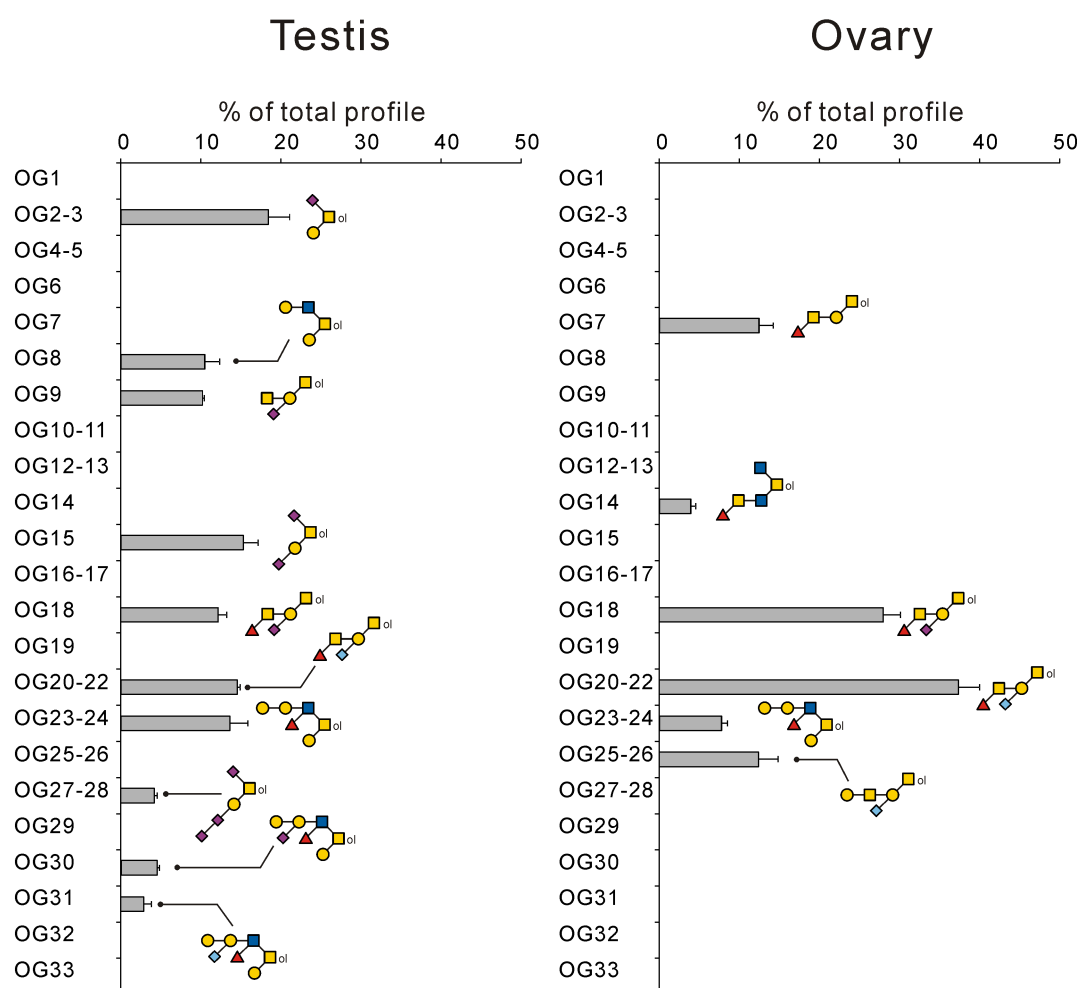


Supplementary Figure 5 (2/4)



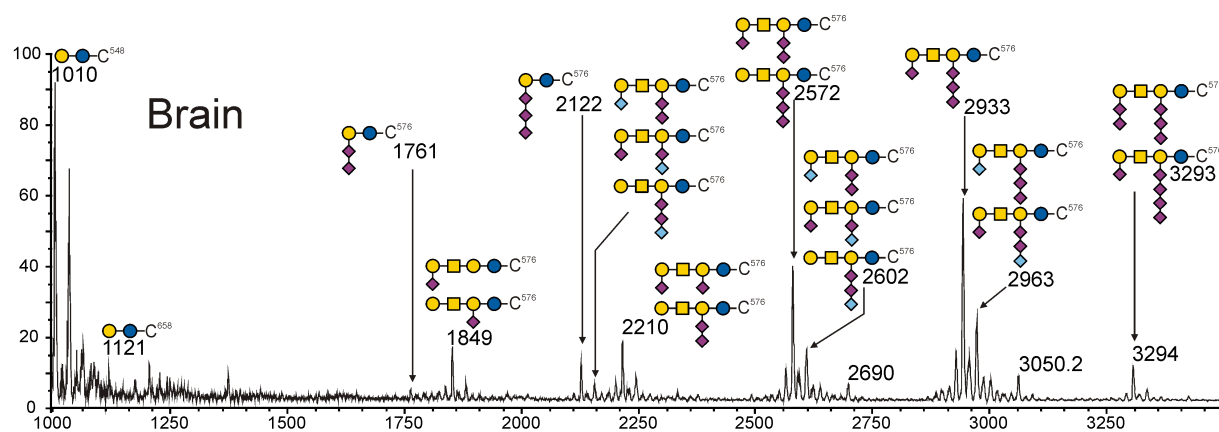
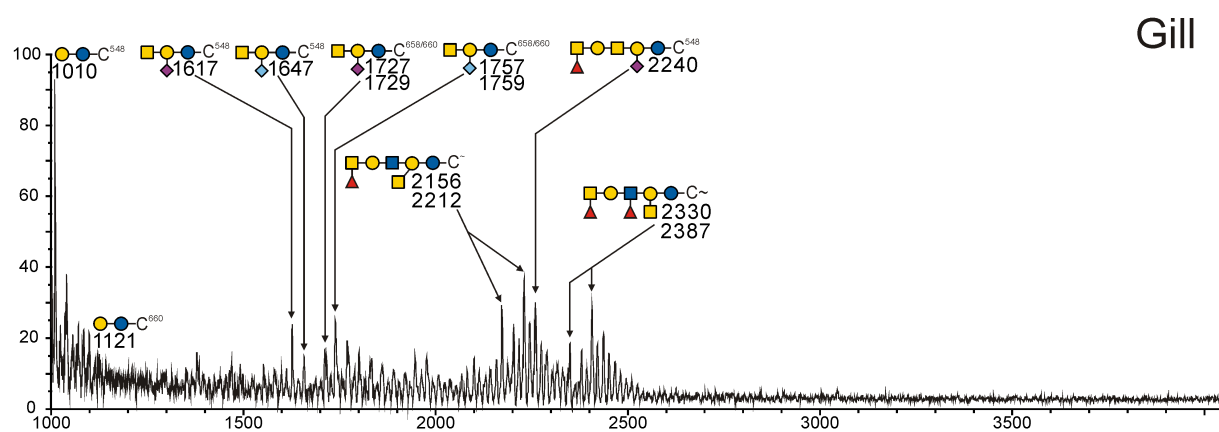
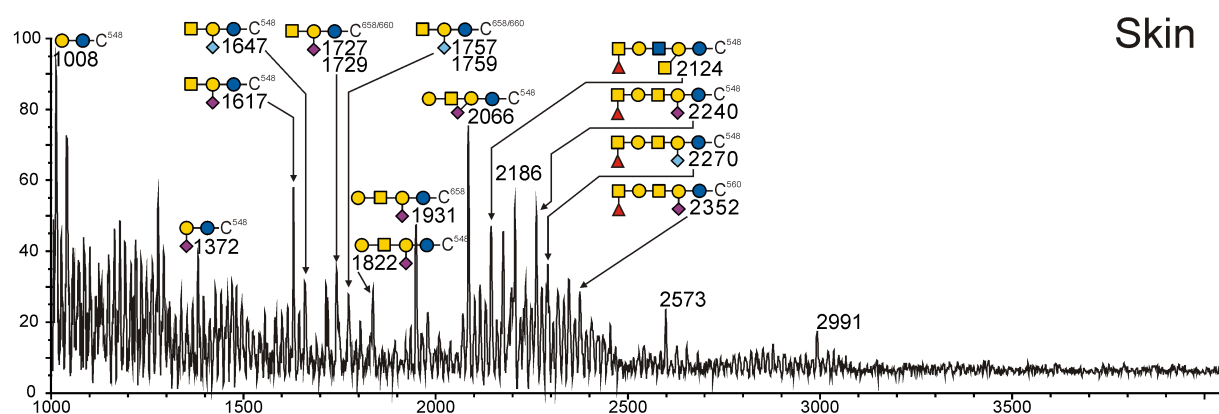
Supplementary Figure 5 (3/4)



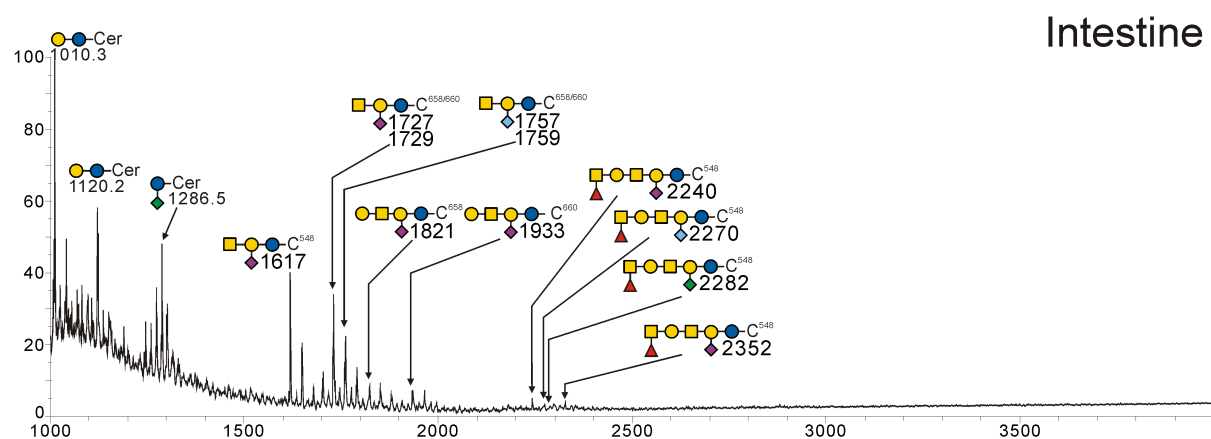
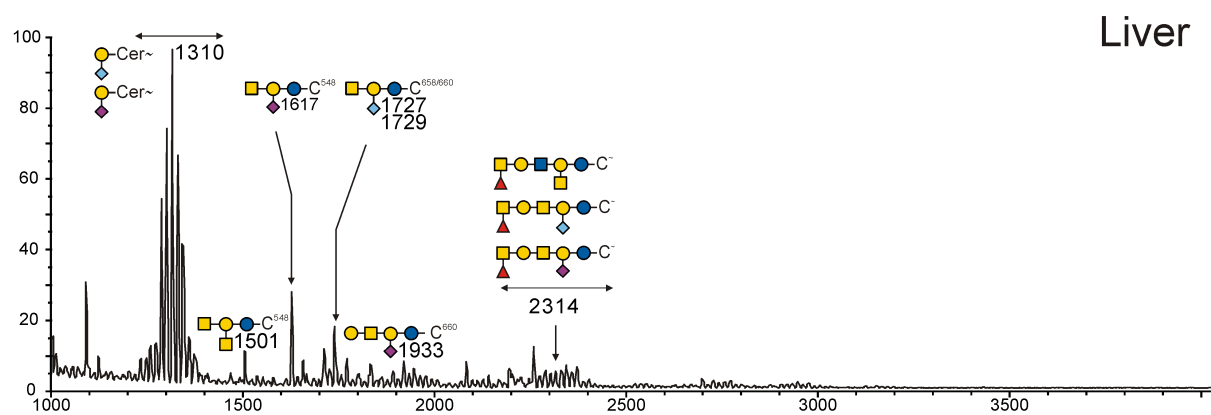
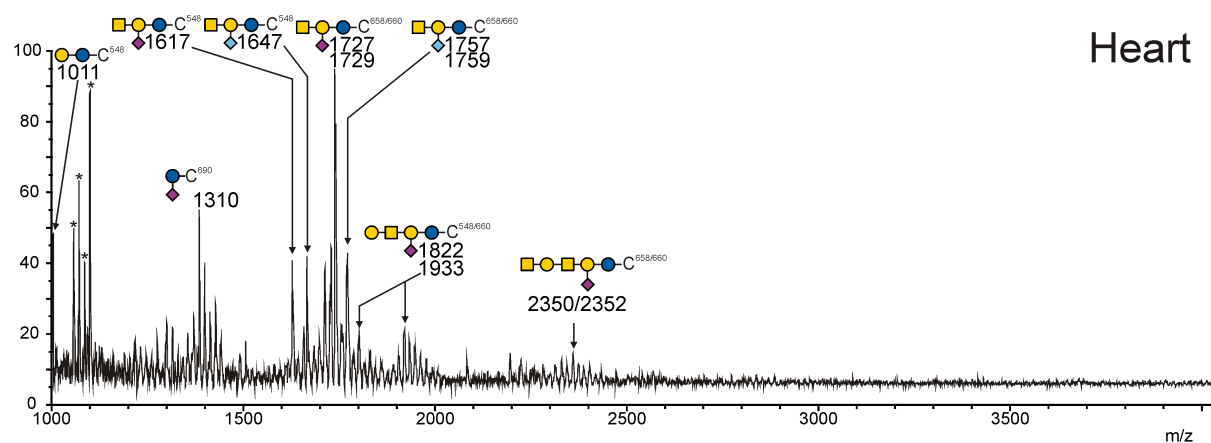


Supplementary Figure 5 - Relative quantification of OGs in eight organs. OGs were quantified using the integration value of identified m/z signals recorded by MALDI-MS analysis of permethylated glycans as % of the total identified signals (see Fig. 5). Quantification values resulted from three independent experiments. The relative importance of isomeric structures in individual m/z signals was estimated by the intensity of MS/MS signals generated by MALDI-TOF/TOF fragmentation signals. Data are shown as mean \pm SE of three experiments.

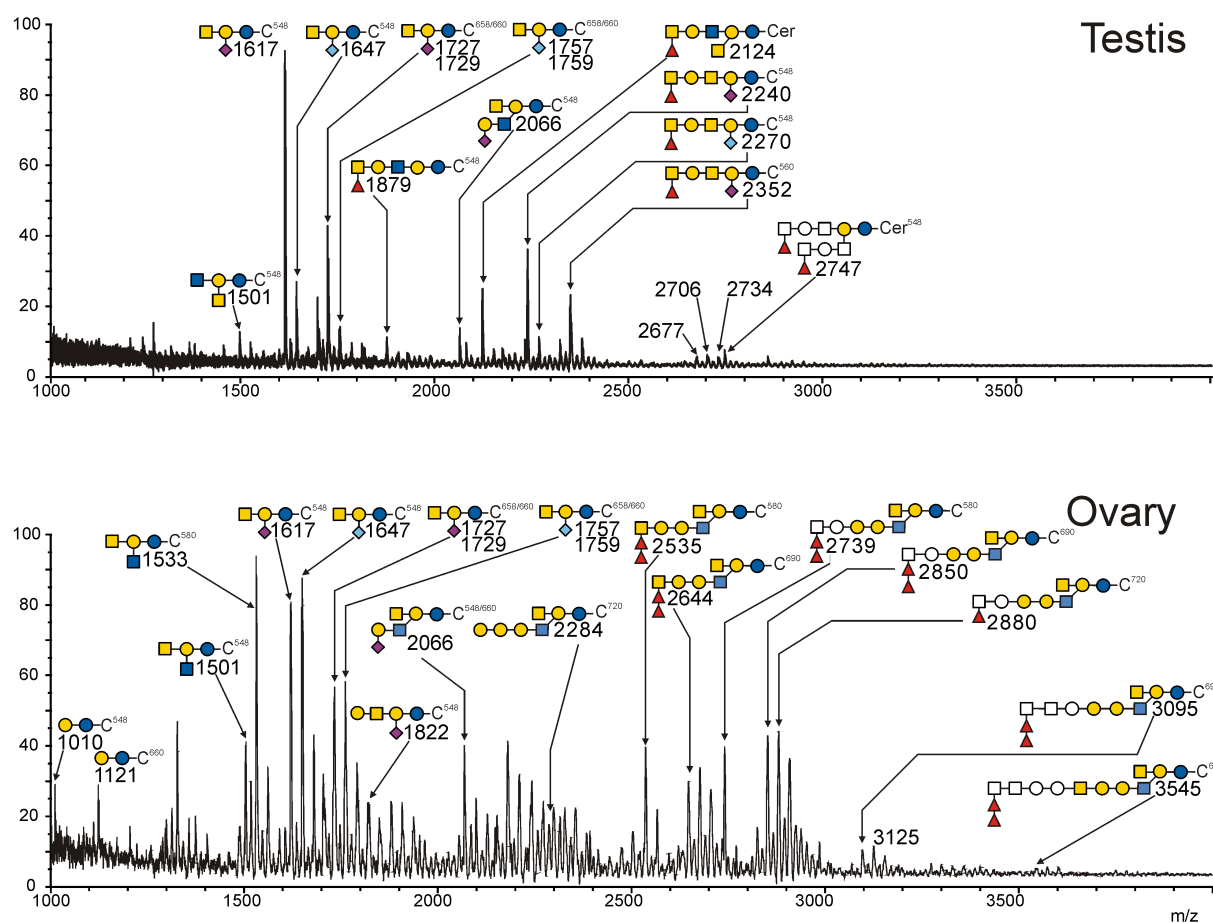
Supplementary figure 6 (1/3)



Supplementary figure 6 (2/3)

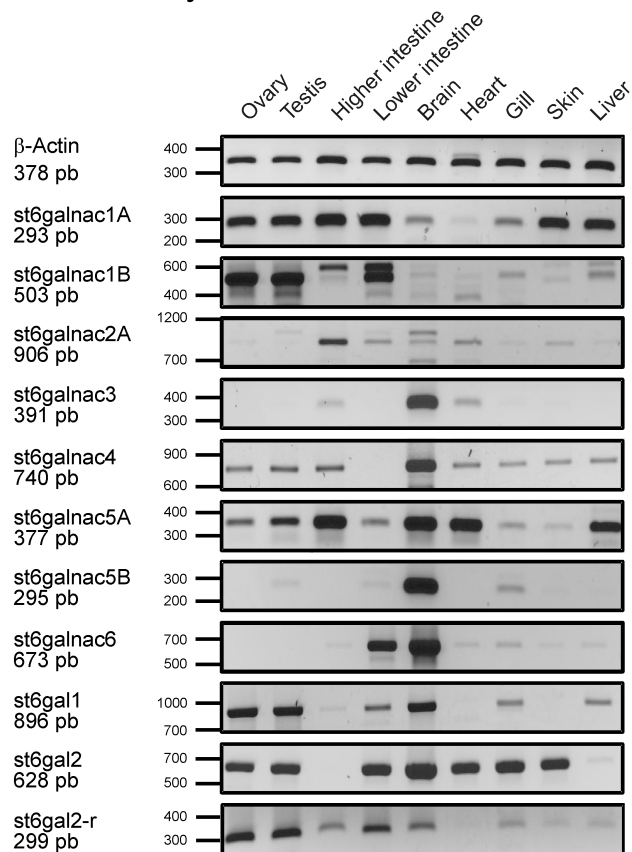


Supplementary figure 6 (3/3)

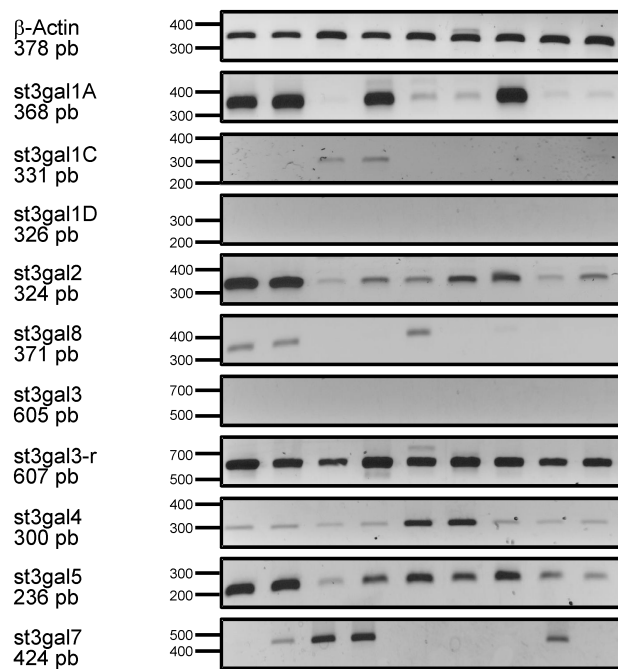


Supplementary Figure 6 - MALDI-TOF-MS spectra of permethylated GSLs isolated from the eight organs. Graphical representation is based on accepted conventions for glycans and monosaccharide nomenclature. White square, HexNAc; white circle, Hex.

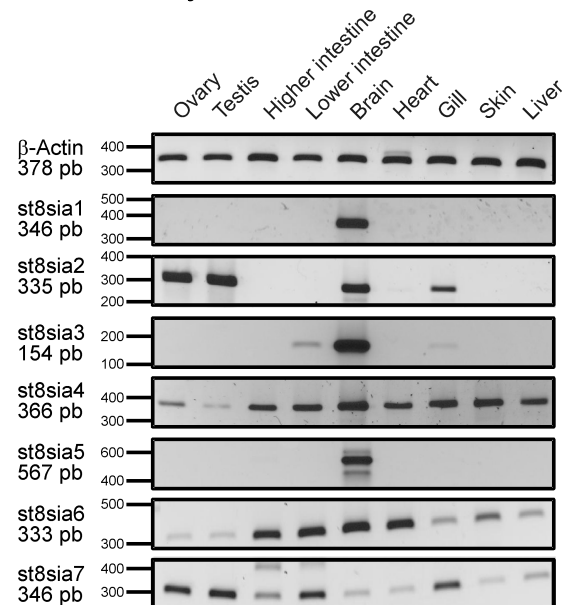
α 2,6-Sialyltransferases



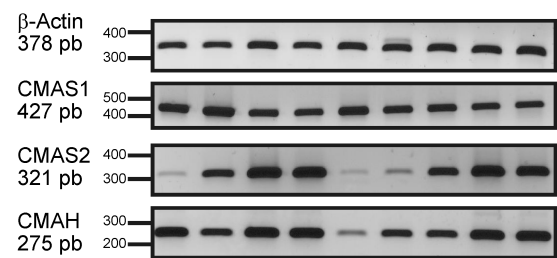
α 2,3-Sialyltransferases



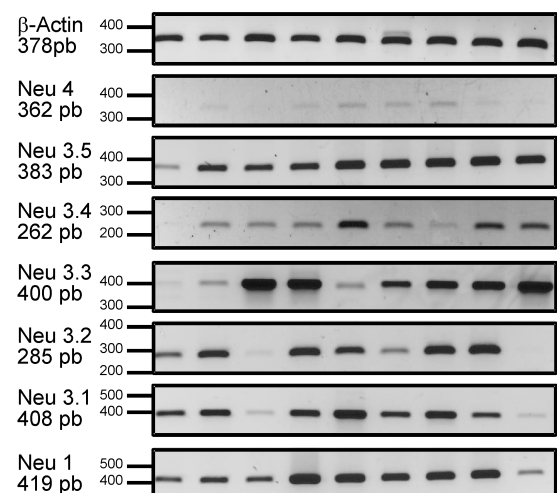
α 2,8-Sialyltransferases



Synthases



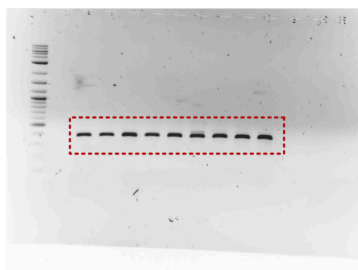
Sialidases



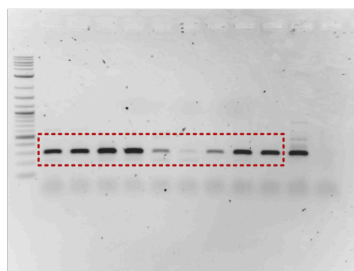
Supplementary Figure 7 - Expression patterns of sialyltransferases, CMP-sialic acid synthases (CMAS), sialidases (Neu) and CMP-sialic acid hydroxylases (CMAH) genes in various zebrafish tissues using RT-PCR. Relative expression levels of zebrafish genes and β -actin mRNA were evaluated as described in the material and methods. Oligonucleotide primer sequences specific of each zebrafish enzymes (sialidases, CMAH and CMAS) are given in Supplementary Data 5. The zebrafish β -actin (378 bp) was amplified as a control of cDNA synthesis and purity. The uncropped versions of agarose gel electrophoresis are provided in Supplementary Figure 8.

Supplementary Figure 8 (1/3)

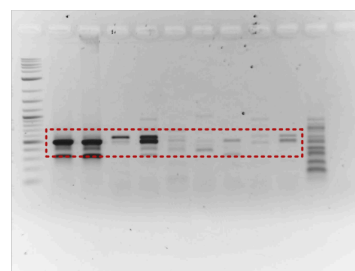
β -actin



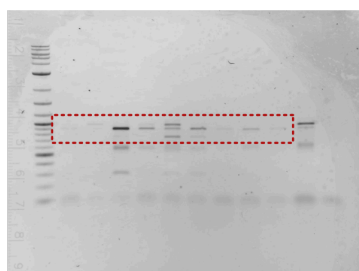
ST6GalNAc1A



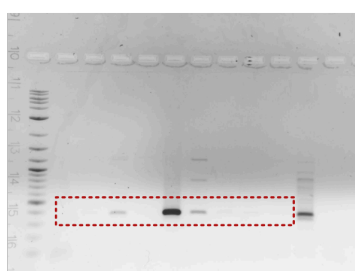
ST6GalNAc1B



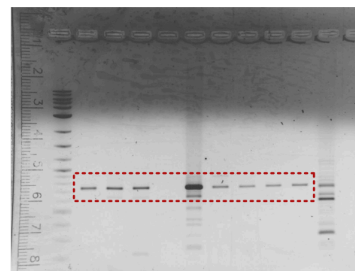
ST6GalNAc2A



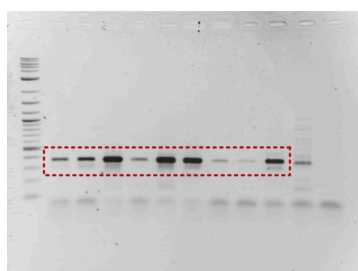
ST6GalNAc3



ST6GalNAc4



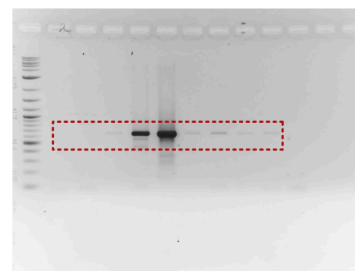
ST6GalNAc5A



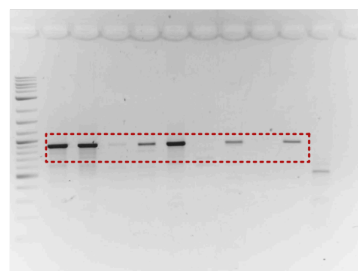
ST6GalNAc5B



ST6GalNAc6



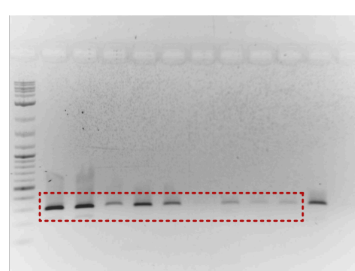
ST6Gal1



ST6Gal2



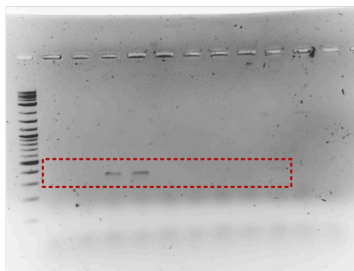
ST6Gal2-r



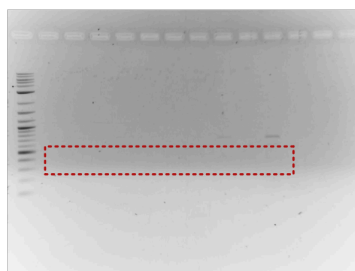
ST3Gal1A



ST3Gal1C

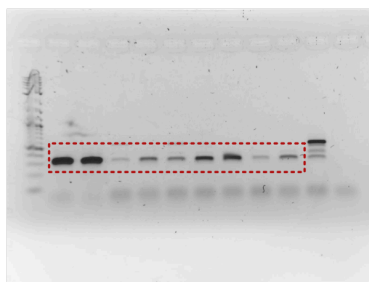


ST3Gal1D

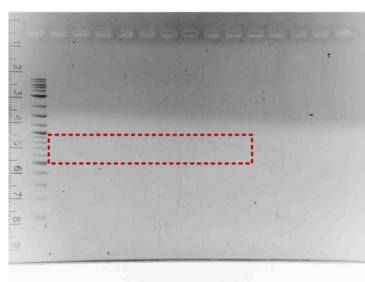


Supplementary Figure 8 (2/3)

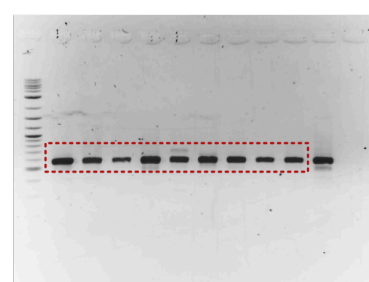
ST3Gal2



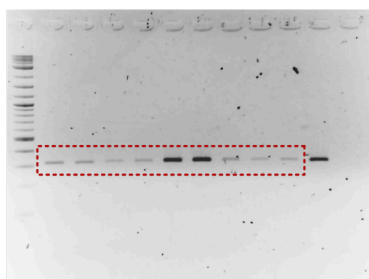
ST3Gal3



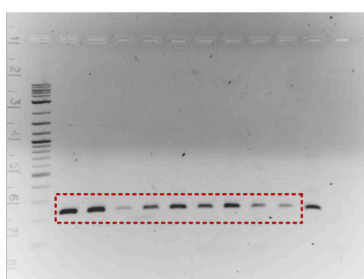
ST3Gal3-r



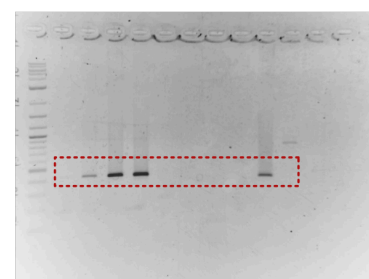
ST3Gal4



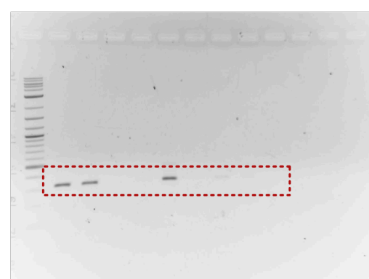
ST3Gal5



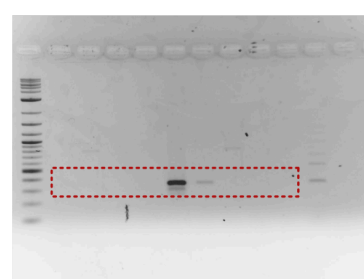
ST3Gal7



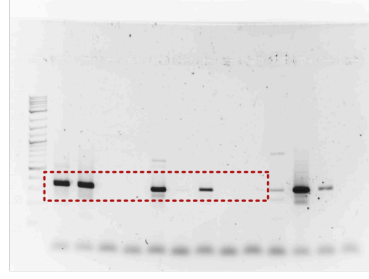
ST3Gal8



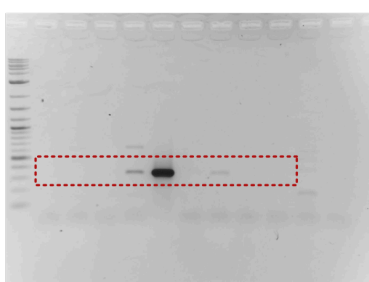
ST8Sia1



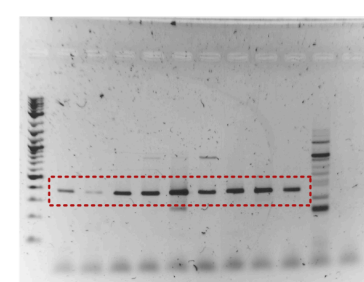
ST8Sia2



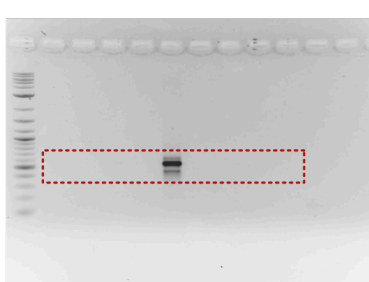
ST8Sia3



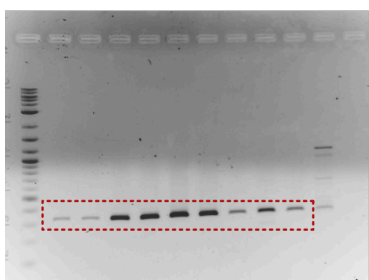
ST8Sia4



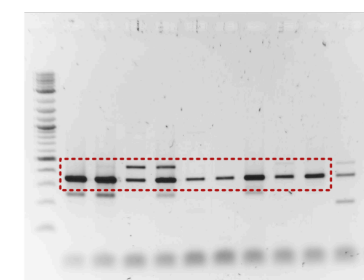
ST8Sia5



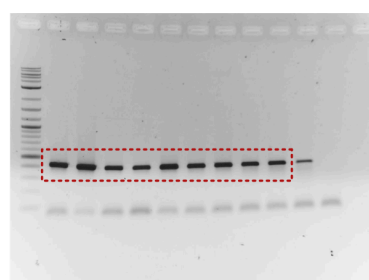
ST8Sia6



ST8Sia7

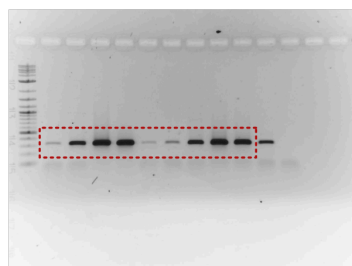


CMAS1



Supplementary Figure 8 (3/3)

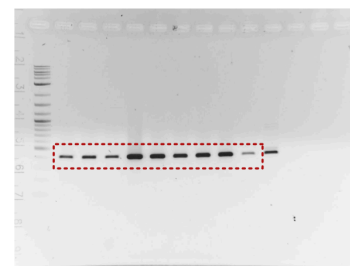
CMAH



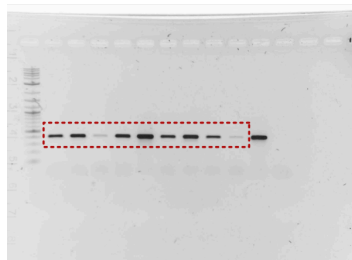
Neu1



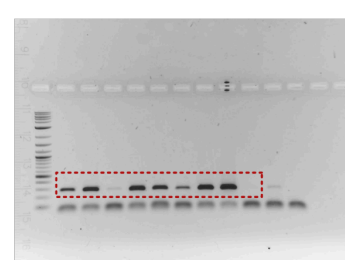
Neu31



Neu32



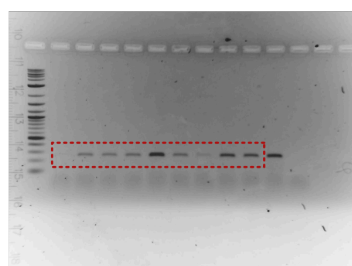
Neu33



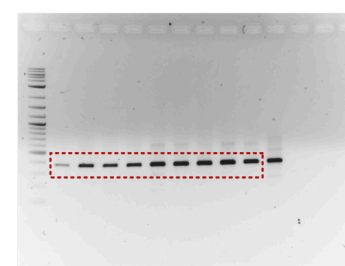
Neu34



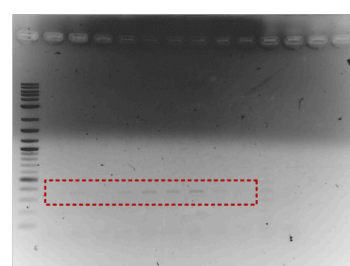
Neu35



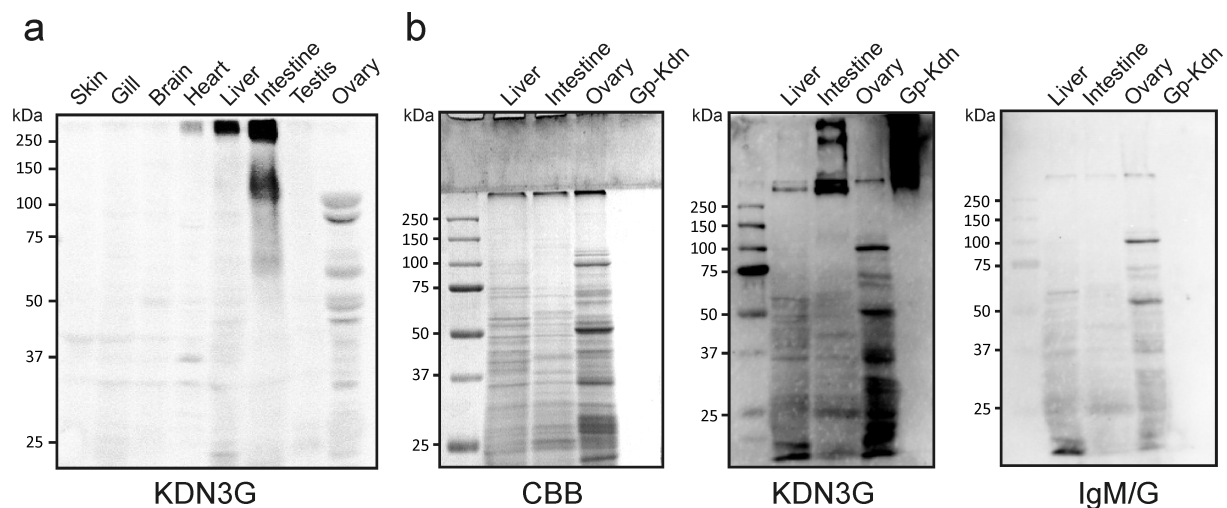
Neu4



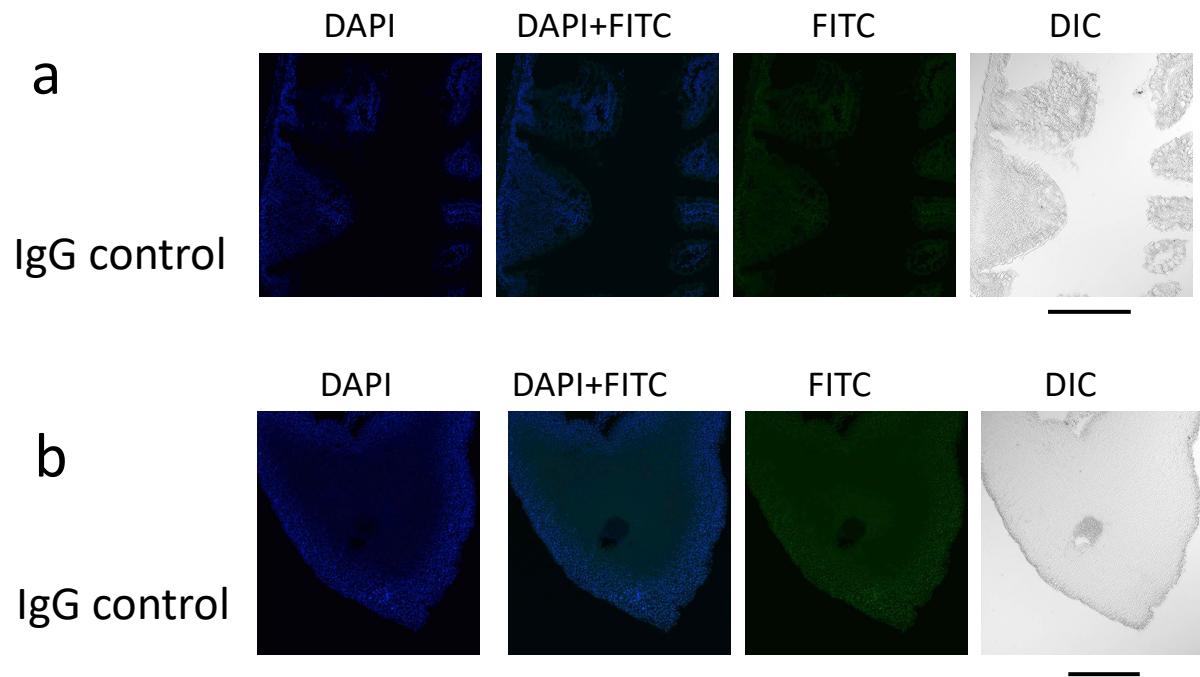
Neu4



Supplementary Figure 8 - 2% agarose gel electrophoresis of amplified fragments, visualized by ethidium bromide.

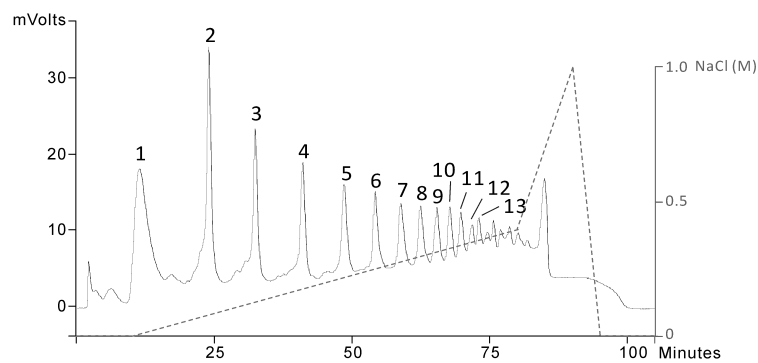


Supplementary Figure 9 - (a) Monoclonal antibody raised against Kdn(α 2,3)Gal epitope (KDN3G) showed a strong response for two high molecular weight broad bands around 125 kDa and above 250 kDa. (b) WB analysis of high molecular weight proteins from SDS-PAGE with concentration gels (CBB, coomassie blue). In addition to the high molecular weight component previously observed, intestine specifically contains KDN3G reactive components that did not reach the separating gel, probably because of their very high molecular weight. Their chromatographic behaviour, as well as the absence of detection by Coomassie Brilliant Blue staining, strongly suggest that these high molecular components are mucin-type highly glycosylated glycoproteins. IgM/G control showed that signals observed in ovary were non-specific.

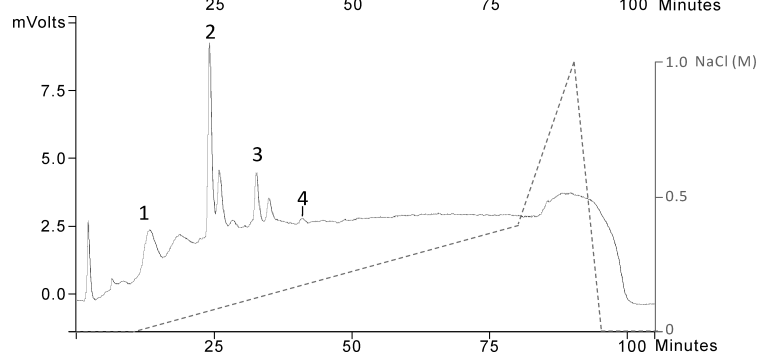


Supplementary Figure 10 - Control staining of (a) intestine (scale corresponds to 100 μm) and (b) liver (scale corresponds to 50 μm) using IgG only showed background staining.

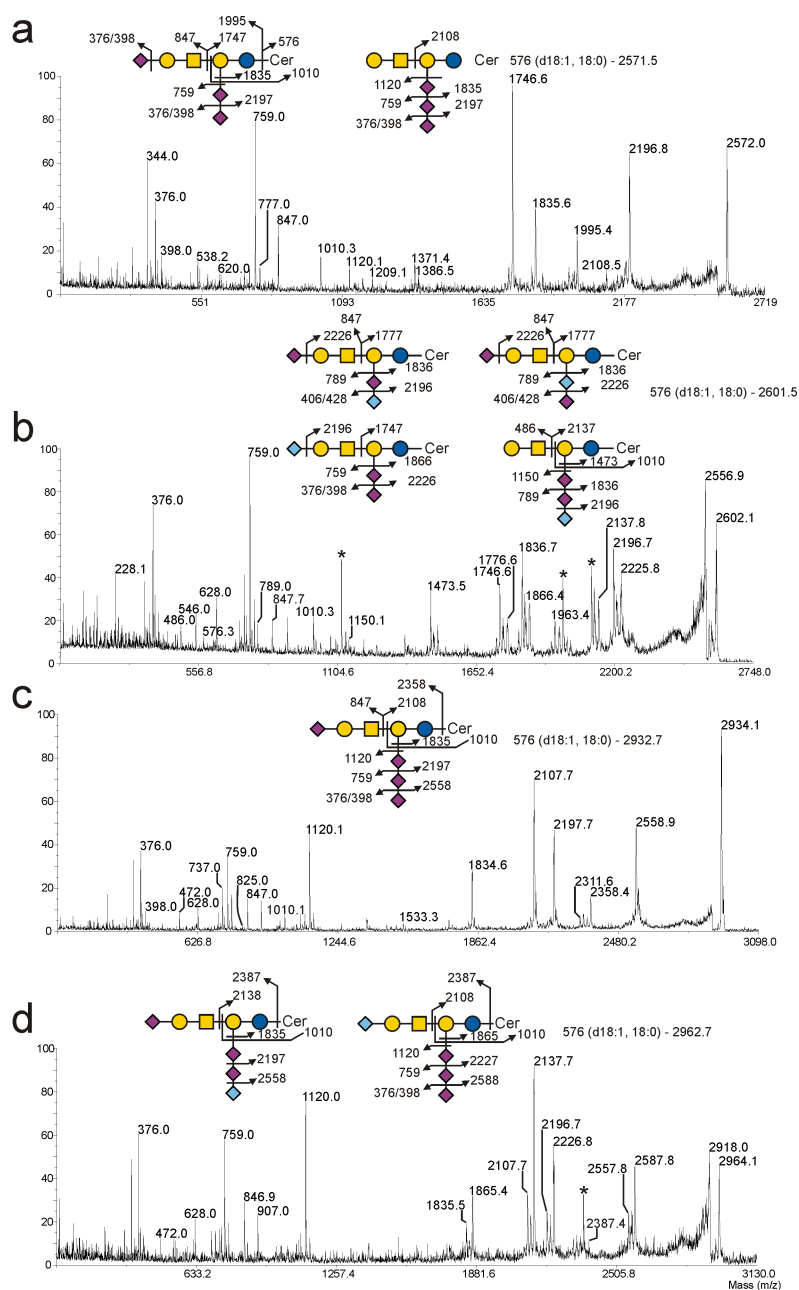
a



b

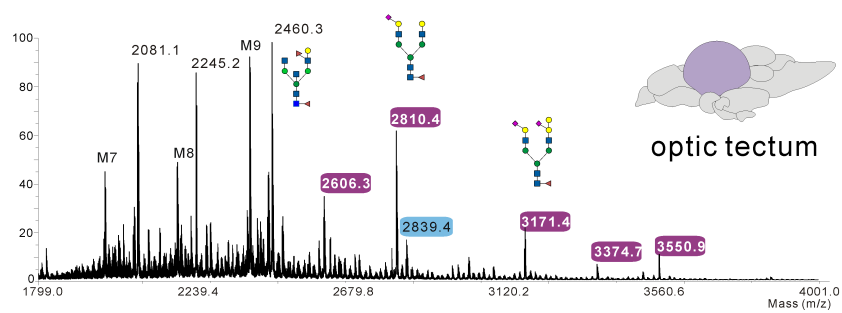
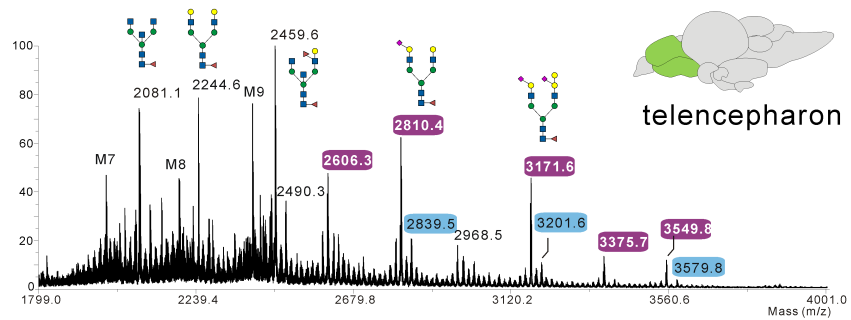
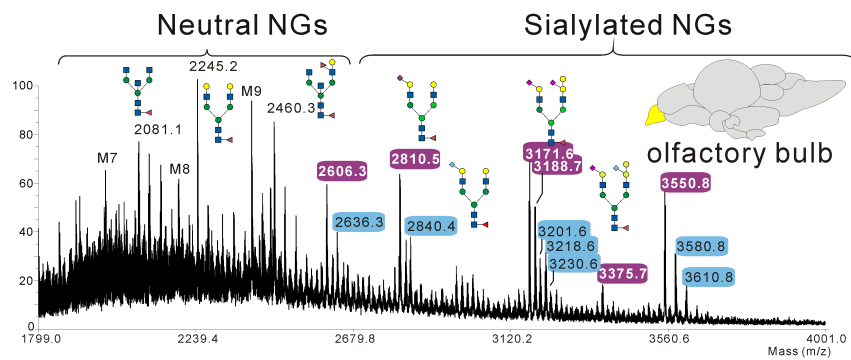
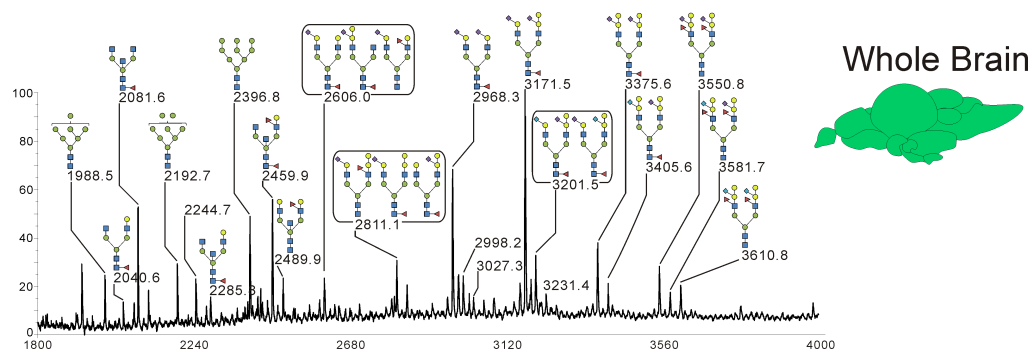


Supplementary Figure 11 - Profiles of oligosialylated sequences on brain GSLs. OligoSia sequences were released from GSLs isolated from whole brains, tagged with DMB and separated by HPLC-FD on an anion exchange column. OligoSia profiles from (a) polysialic colominic acid used as authentic standard and (b) GSLs. Peaks are labelled according to the DP values as established by comparison with colominic acid. The multiple peaks observed for DP2 and DP3 probably originate from different NeuAc/NeuGc content, as previously shown in embryonic glycoproteins⁸.

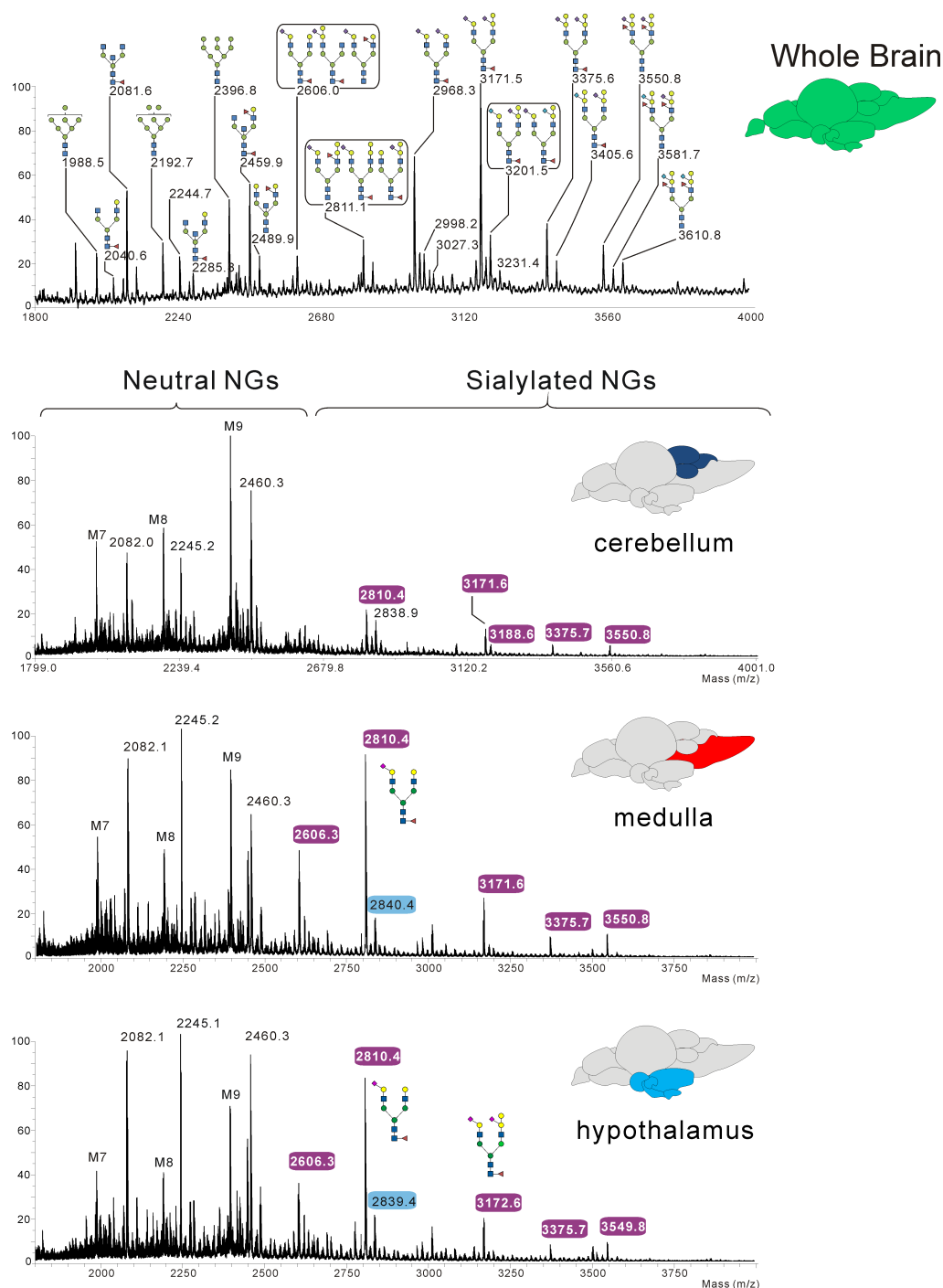


Supplementary Fig. 12 - MALDI-TOF/TOF MS/MS spectra of permethylated GSLs from ganglio-series GL96-97 (a), GL98-99-100-101 (b), GL102 (c) and GL103-104 (d)

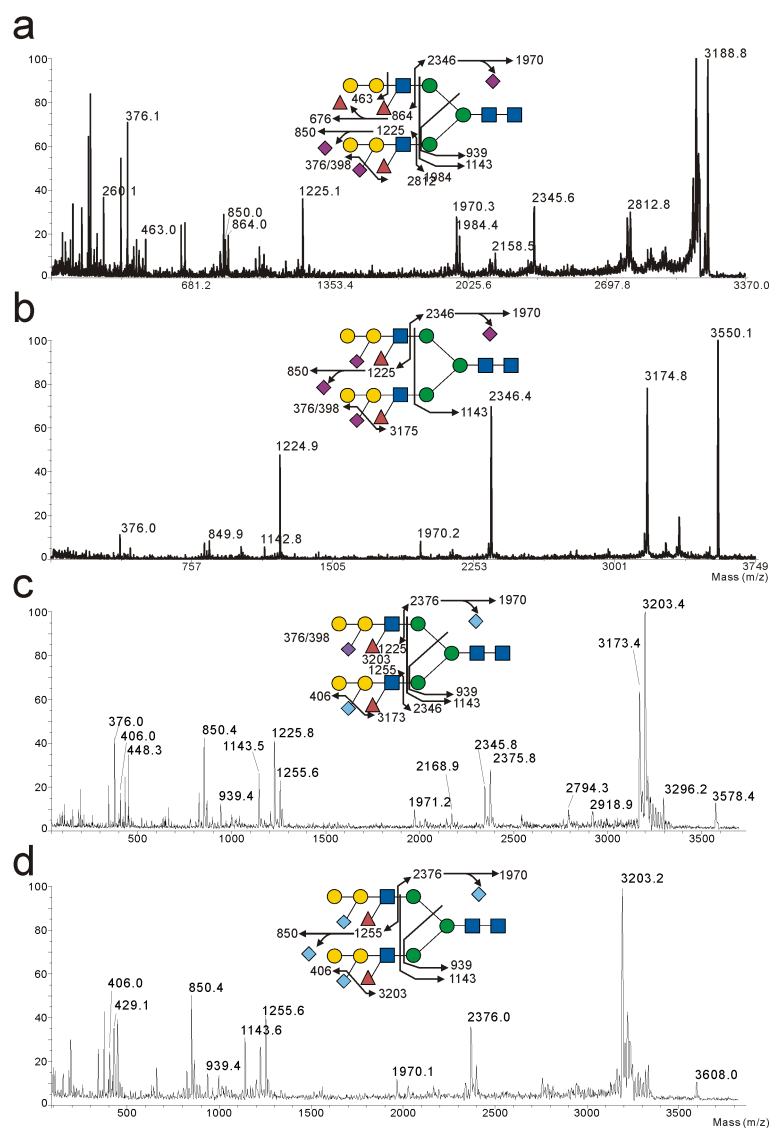
Supplementary Fig. 12 (1/2)



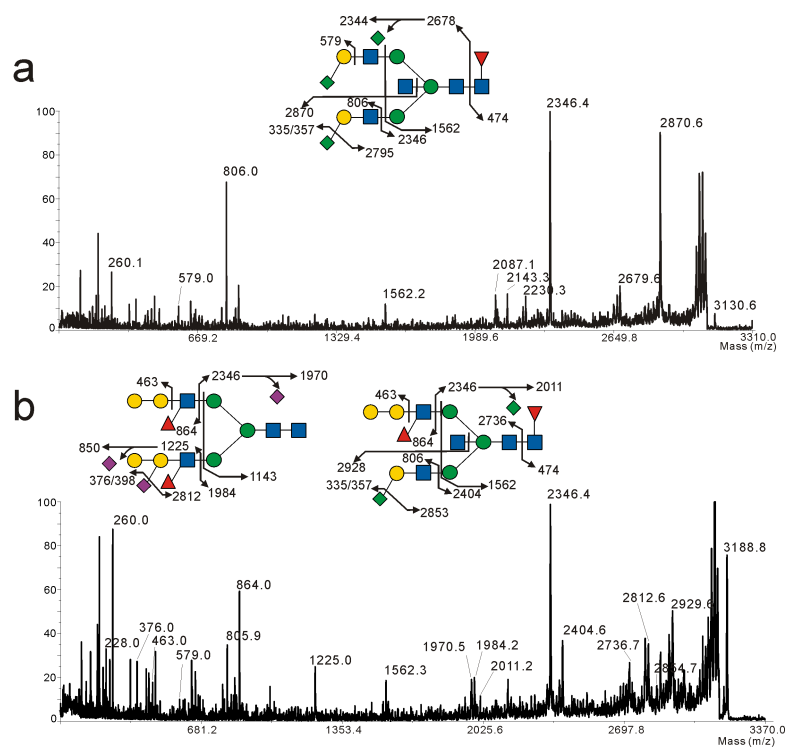
Supplementary Fig. 12 (2/2)



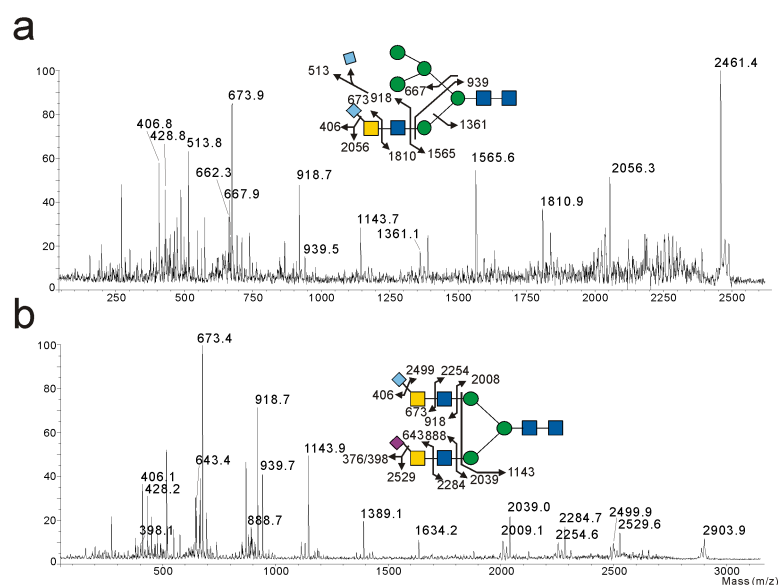
Supplementary Fig. 13 - MALDI-TOF MS spectra of permethylated NGs isolated from whole brain (top) and different regions of brain. Considering the small quantities of starting material, an alternative extraction protocol based on Triton X-100 was used, as described in Methods section of the manuscript. Values highlighted in purple exclusively contain Neu5Ac, those highlighted in light blue contain Neu5Gc. Olfactory bulb is the only region that show intense Neu5Gc associated signals. Analyses were done on three independent samples with similar results.



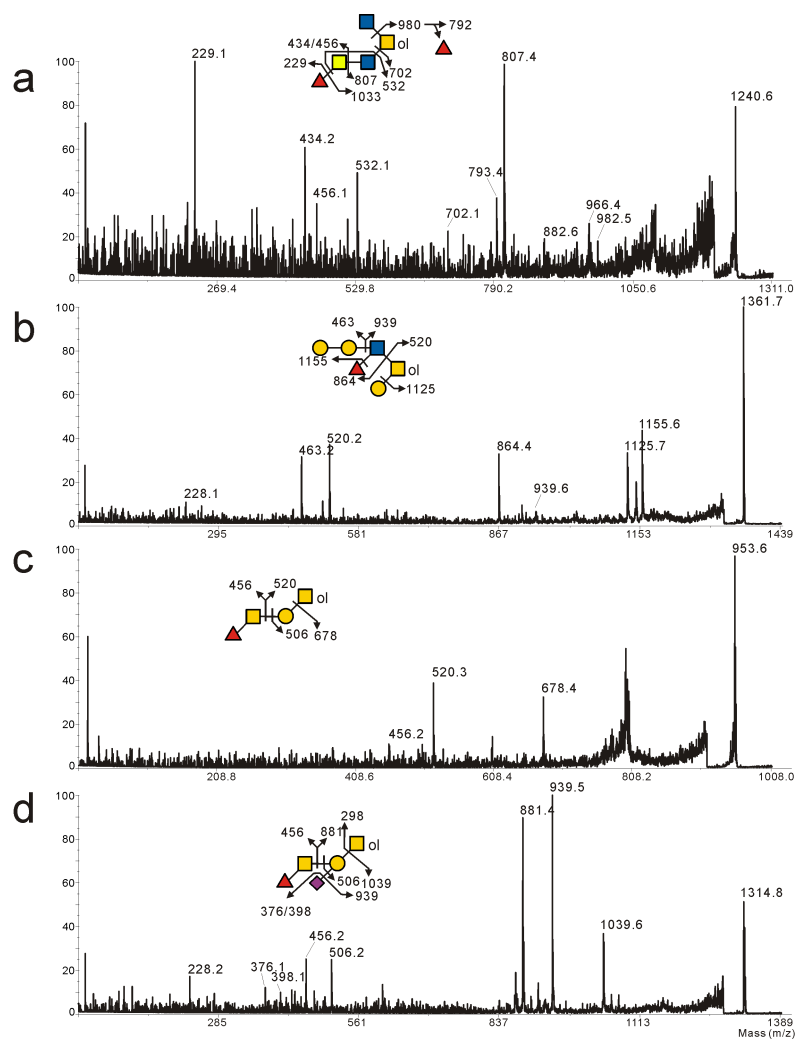
Supplementary Figure 14 - MALDI-TOF/TOF MS/MS spectra of permethylated zebrafish epitope Gal(β 1,4)(NeuAc/Gc(α 2,3))Gal(β 1,4)[Fuc(α 1,3)]GlcNAc containing NGs NG77 (a), NG89 (b), NG90 (c) and NG91 (d).



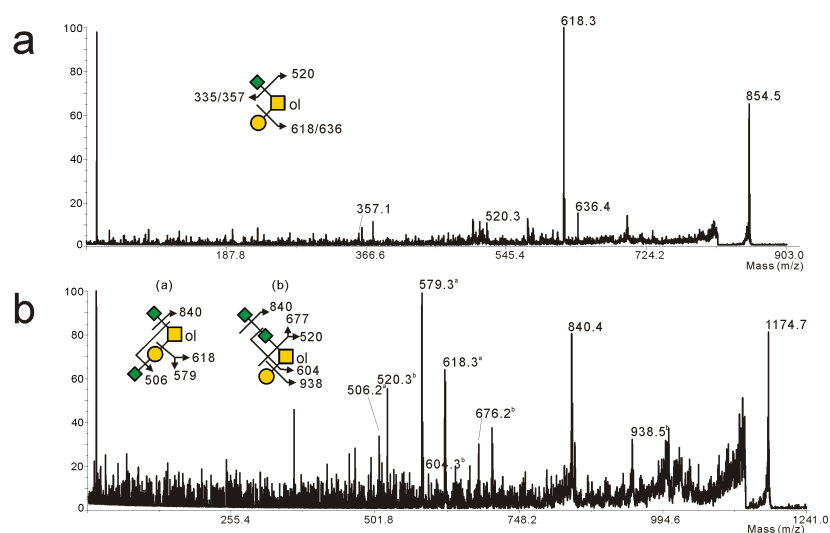
Supplementary Figure 15 - MALDI-TOF/TOF MS/MS spectra of permethylated Kdn containing NGs NG74 (a) and NG78 (b).



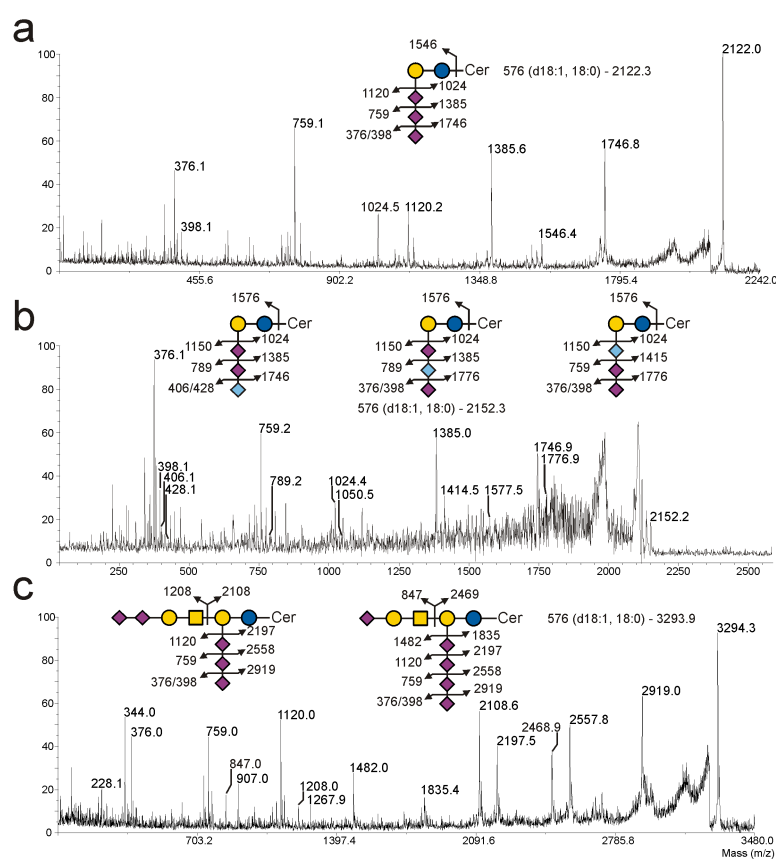
Supplementary Figure 16 - MALDI-TOF/TOF MS/MS spectra of permethylated LacdiNAc-containing NGs NG24 (a) and NG58 (b).



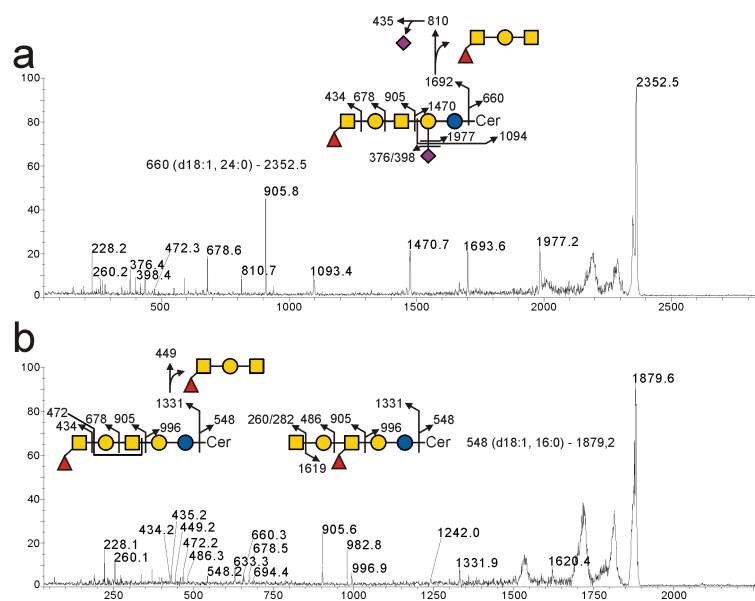
Supplementary Figure 17 - MALDI-TOF/TOF MS/MS spectra of permethylated OGs OG14 (a), OG23 (b), OG7 (c) and OG18(d)



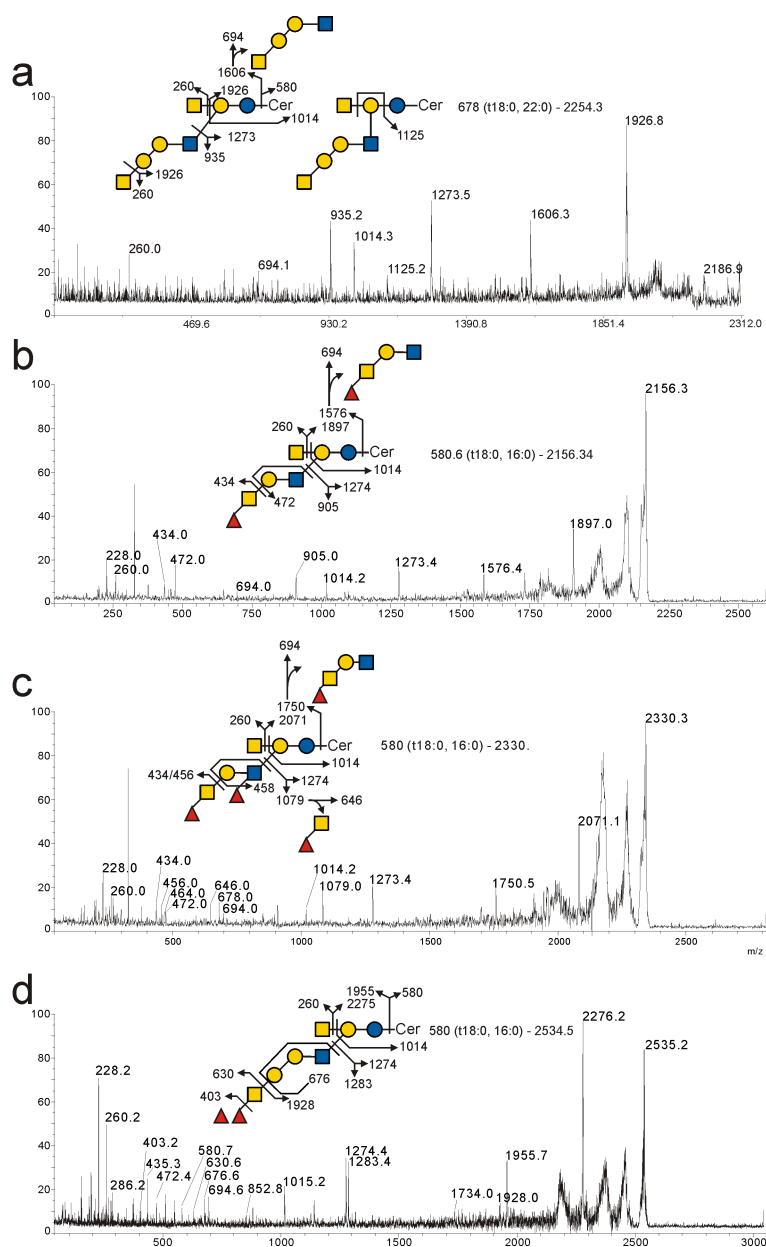
Supplementary Figure 18 - MALDI-TOF/TOF MS/MS spectra of permethylated Kdn-substituted OGs OG1 (a) and OG10 & OG11 (b)



Supplementary Figure 19 - MALDI-TOF/TOF MS/MS spectra of permethylated GSLs from hemato-series GL63 (a) and GL64-65-66 (b) and from ganglio-series GL105-106 (c)



Supplementary Figure 20 - MALDI-TOF/TOF MS/MS spectra of permethylated GSLs from the extended ganglio-series GL118 (a) and GL107-108 (b)



Supplementary Figure 21 - MALDI-TOF/TOF MS/MS spectra of permethylated GSLs from the lacto-ganglio-series GL149 (a), GL131 (b), GL147 (c) and GL150 (d)

Supplementary References

1. Yu, S.-Y., Wu, S.-W. & Khoo, K.-H. Distinctive characteristics of MALDI-Q/TOF and TOF/TOF tandem mass spectrometry for sequencing of permethylated complex type N-glycans. *Glycoconj. J.* **23**, 355–369 (2006).
2. Wada, Y. *et al.* Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17**, 411–422 (2007).
3. Wada, Y. *et al.* Comparison of methods for profiling O-glycosylation: Human Proteome

Organisation Human Disease Glycomics/Proteome Initiative multi-institutional study of IgA1. *Mol. Cell. Proteomics MCP* **9**, 719–727 (2010).

4. Aoki, K. *et al.* Dynamic Developmental Elaboration of N-Linked Glycan Complexity in the *Drosophila melanogaster* Embryo. *J. Biol. Chem.* **282**, 9127–9142 (2007).

5. Guérardel, Y., Chang, L.-Y., Maes, E., Huang, C.-J. & Khoo, K.-H. Glycomic survey mapping of zebrafish identifies unique sialylation pattern. *Glycobiology* **16**, 244–257 (2006).

6. Vanbeselaere, J. *et al.* Mapping the expressed glycome and glycosyltransferases of zebrafish liver cells as a relevant model system for glycosylation studies. *J. Proteome Res.* **11**, 2164–2177 (2012).

7. Hanzawa, K., Suzuki, N. & Natsuka, S. Structures and developmental alterations of N-glycans of zebrafish embryos. *Glycobiology* **27**, 228–245 (2017).

8. Chang, L.-Y. *et al.* Developmental regulation of oligosialylation in zebrafish. *Glycoconj. J.* **26**, 247–261 (2009).

9. Guérardel, Y., Chang, L.-Y., Maes, E., Huang, C.-J. & Khoo, K.-H. Glycomic survey mapping of zebrafish identifies unique sialylation pattern. *Glycobiology* **16**, 244–257 (2006).

10. Niimura, Y. & Ishizuka, I. Unique disialosyl gangliosides from salmon kidney: Characterization of V3 α Fuc, IV3 β GalNAc, II3(α NeuAc)2-Gg4Cer and its analogue with 4-O-acetyl-N-acetylneuraminic acid. *Glycoconj. J.* **23**, 489–499 (2006).

11. DeGasperi, R. *et al.* Isolation and characterization of gangliosides with hybrid neolacto-ganglio-type sugar chains. *J. Biol. Chem.* **262**, 17149–17155 (1987).

12. Ostrander, G. K., Levery, S. B., Hakomori, S. & Holmes, E. H. Isolation and characterization of the major acidic glycosphingolipids from the liver of the English sole (*Parophrys vetulus*). Presence of a novel ganglioside with a Forssman antigen determinant. *J. Biol. Chem.* **263**, 3103–3110 (1988).

13. Ostrander, G. K. *et al.* Isolation and characterization of four major neutral glycosphingolipids from the liver of the English sole (*Parophrys vetulus*). Presence of a novel branched lacto-ganglio-iso-globo hybrid structure. *J. Biol. Chem.* **263**, 18716–18725 (1988).