HOKKAIDO UNIVERSITY

Title	Human serum N-glycans as highly sensitive cancer biomarkers : Potential benefits and the risks
Author(s)	Gebrehiwot, Abrha Gebreselema
Citation	.　（　　　）　13604
Issue Date	2019-03-25
DOI	10.14943/doctoral.k13604
Doc URL	http://hdl.handle.net/2115/91646
Type	theses (doctoral)
File Information	Abrha_Gebreselema_Gebrehiwot.pdf

Instructions for use

博士学位論文

Doctoral Dissertation

論文題目

Title

Human serum *N*-glycans as highly sensitive cancer biomarkers: Potential

benefits and the risks

(高感度がんマーカーとしてのヒト血清 *N*-グリカン：その潜在的な恩恵と危険性)

氏名

Name

Abrha Gebreselema Gebrehiwot

北海道大学大学院生命科学院

Graduate School of Life Science, Hokkaido University

March 2019

Abstract

In cancer progression, aberrant serum *N*-glycan alterations have been associated with proliferation, invasion, metastasis, aggressiveness, angiogenesis, oncogenic signaling, and immune regulation of tumor cells. Deciphering cancer-associated alterations in protein glycosylation has been of critical interest not only to understand the oncogenic mechanisms, but also to suggest distinct glycan signatures of cancer that can be considered for diagnostic biomarker and therapeutic opportunities. In this context, more than half of current FDA-approved protein cancer markers are glycoproteins (including CA15-3 and CA27-29 for breast cancer; CA19-9 for pancreatic, gastric, and colonic cancers; CA-125 for ovarian cancer; CEA for Colorectal Cancer; PSA for prostate cancer; and α-fetoprotein for hepatocellular cancer). Breast cancer (BC) has become the most common malignancy and the leading cause of cancer death among women globally. Although there have been improvements in patient diagnosis and survival for breast cancer, there is no clinically validated serum biomarker for its early diagnosis. This study primarily aimed to explore the potentials of *N*-glycans directly released from patient serum glycoproteins and purified immunoglobulin G (IgG) fraction in distinguishing breast cancer patients from matched normal controls (NC) towards the discovery of non-invasive markers.

I used a comprehensive glycomics approach by integrating glycoblotting-based glycan purification with MALDI-TOF/MS-based quantitative analysis using sera of BC patients belonging to stages I-IV and normal controls collected from Ethiopian women during 2015-2016. IgG was purified from serum of all subjects by affinity chromatography using protein G spin plate and further subjected to glycoblotting for glycan release. Mass spectral data were further processed and evaluated rigorously, using various bioinformatics and statistical tools. From total of 46 detected serum *N*-glycan structures, 35 were significantly up-regulated in the sera of all BC patients within which 17 *N*-glycans (comprising core-fucosylated, multiply branched and sialylated structures) showed strong diagnostic potential (AUC, 0.8-1) for early stage (I and II). High abundance of these glycans has been strongly associated with greater invasive and metastatic potential of cancer. In a more detail analysis, *N*-glycans from IgG confirmed consistent abundance in BC patients, particularly two core-fucosylated and agalactosylated glycans specifically distinguished (AUC, 0.944 and 0.921, p≤0.001) stage II patients from NC. Our findings of increased IgG core-fucosylation and agalactosylation were reported to be associated with a decrease in immunosuppressive potential of IgG towards tumor

cells, which in part may correlate with the aggressive nature of BC commonly noticed in black population. Altogether, these findings reveal characteristic tumor-associated glycoforms which demonstrate significant distinguishing potential not only between the whole BC patients and the healthy controls but also in a cancer stage dependent manner.

Most glycomics studies have previously focused on understanding disease mechanisms and proposing serum markers for various cancers, yet the influence of confounding factors like ethnic variation on the identified glyco-biomarker remains poorly addressed. Hence, my second study aimed to investigate the inter-ethnic variation in serum *N*-glycan and free sialic acid patterns among US origin control, Japanese, Indian, and Ethiopian healthy volunteers in association with the identified glycan biomarkers for hepatocellular (HCC) and other cancers. Ethnic specific detections and differential expression levels, particularly highest abundance (p < 0.001) of high mannose, core-fucosylated, hyperbranched/hypersialylated *N*-glycans were demonstrated in Ethiopians. Glycan abundance trend of healthy Ethiopians was nearly close to that of Japanese HCC patients. Surprisingly, some of the glycoforms greatly elevated in the Ethiopian subjects have been identified as sensitive serum biomarkers of various cancers. Fluorescence HPLC-based quantification of free sialic acid demonstrated significant up-regulation primarily in the Ethiopians, compared to the other ethnicities.

In conclusion, the patient based comprehensive study has addressed for the first time both whole serum and IgG *N*-glycosylation signatures of native black women suffering from BC and revealed novel glyco-biomarkers with marked overexpression and distinguishing ability at early stage patients. The informative findings from the inter-ethnic serum *N*-glycomic study emphasize the substantial influence of ethnic difference in human serum *N*-glycome variation, the ignorance of which may provide unclear and imprecise conclusion of the diagnosis by using glycan-related disease biomarkers.

Finally, I have clearly shown the potential of serum glycomics profiling as non-invasive approach to discover early stage specific cancer biomarkers, as well as the need to strictly consider ethnic matching in population based glyco-biomarker discovery researches to avoid inaccurate diagnosis.

Contents

List of Figures   Page

List of Tables

Page

Abbreviations

ADCC	Antibody-dependent cell-mediated cytotoxicity
AFP	Alpha Fetoprotein
AGP	Alpha-acid glycoprotein
ANOVA	Analysis of variance
AUC	Area under the curve
BMI	Body mass index
BRCA	Breast cancer susceptibility gene
CA	Carbohydrate antigen
CDC	Complement-dependent cytotoxicity
CEA	Carcino-embryonic antigen
CSF	Cerebrospinal fluid
EGFR	Endothelial growth factor receptor
Fc	Fragment crystallizable
FDA	Food and Drug Administration
FUT8	Fucosyltransferase 8
GFAT	Glutamine: fructose-6-P amidotransferase
GlcNAc	N-acetylglucosamine
GLUT	Glucose transporter
GnT	N-acetylglucosaminyltransferase
HBP	Hexosamine biosynthesis pathway
HER2	Human epidermal growth factor receptor2
HR	Hormone receptor
IgG	Immunoglobulin G
MALDI-TOF/MS	Matrix-assisted laser desorption/ionization mass spectrometry
PNGase F	Peptide:N-Glycosidase F

Neu5Ac	*N*-Acetylneuraminic acid
Neu5GC	*N*-Glycolylneuraminic acid
PSA	Prostate specific antigen
ROC	Receiver operating characteristic
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TNBC	Triple negative breast cancer
TNM	Tumor, Node, Metastasis

Chapter 1

General Introduction

1.1.   Breast Cancer

Breast cancer is a malignant disease caused by an uncontrolled growth and division of breast cells that typically begins either in the ducts that carry milk to the nipple (ductal cancers) or in the glands that make breast milk (lobular cancers), Figure 1.1. Both types of breast cancer have subtypes of in situ (the cancer remains in the area it originated) or invasive (the breast cancer that has spread from where it began in the breast ducts or lobules to surrounding normal tissue and eventually throughout the body) [1, 2]. Based on a recent report, breast cancer has become the most common malignancy and the leading cause of cancer deaths among women worldwide with an estimated 2.4 million incident cases and 523,000 deaths in 2015 [3]. It continues to be a disease with tremendous public health significance worldwide. In United States, where incidence and mortality rates of most cancer types are frequently studied, breast cancer currently accounts for highest incidence rate (30%) and second highest death rate (14%) in females among all cancers [4]. Although breast cancer is thought to be a disease of the developed world, the highest burden of breast cancer mortality, particularly in women younger than age 50 years, poses in less developed regions like Africa where the majority of cases are diagnosed in late stages [5, 6]. Apart from the lack of early detection, such disparities in breast cancer mortality can be explained by inadequate diagnosis and treatment facilities in the low-income countries.
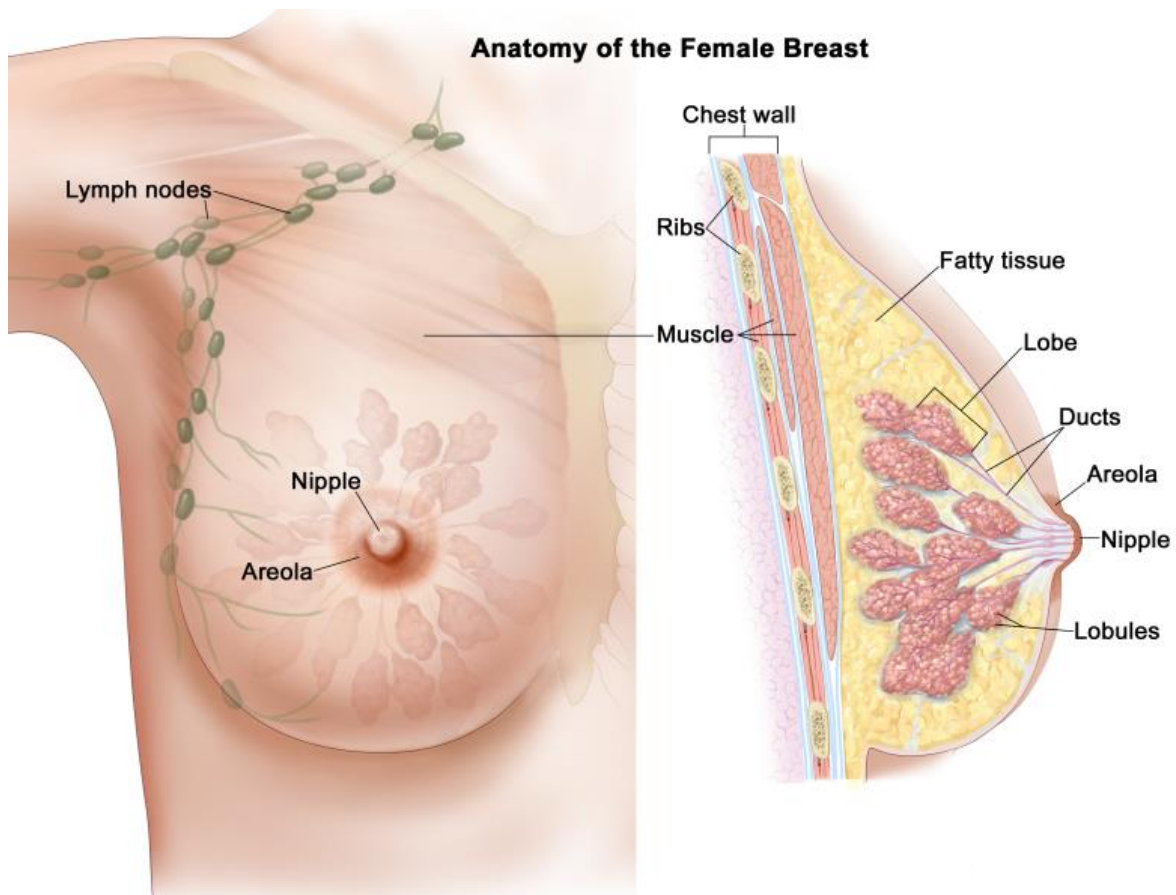
Figure 1.1. Anatomy of the female breast [2].

1.1.1. Risk Factors of Breast Cancer

Cancer is the result of several genetic changes occurring in a cell, altering the balance between proliferation and apoptosis mechanisms [7]. Although the cause of breast cancer is not yet clearly known, several factors pertaining to environment, lifestyle, inheritance, host, and infectious agents have been associated with its oncogenesis process. Among these, being female gender, increase in age, early menarche, late menopause, family history of breast or ovarian cancer, prolonged hormone replacement therapy, oral contraceptive hormones, exposure to chest radiation, smoking, alcohol consumption, race, obesity, and genetic mutations of **br**east **ca**ncer susceptibility genes (BRCA1 and BRCA2: tumor suppressor genes involved mainly in DNA repair*)* have been identified [2, 8]. However, the known risk factors

for the development of breast cancer account for only less than half of all cases [8]. Several studies have reported a black-white disparity in breast cancer incidence and mortality patterns in which incidence rates are higher during younger age (< 60), but lower during older age (≥ 60) in black women compared to white women. Similarly, breast cancer death rates are reported to be 40% higher among black women than their white counterparts [9]. These differences are the result of the fact that young black women are more likely to face an aggressive type of breast cancer, as it tends to have a poorer prognosis and present at a later stage, a higher proliferative rate, a triple negative hormone receptor status, and a higher chance of BRCA1 or BRCA2 mutations [10-12].

1.1.2. Classification of Breast Cancer

Because of its highly heterogeneous nature with variable histopathological and biological features, a suitable classification of breast cancer has not been developed. Thus, breast cancer follows multiple classification systems depending on various parameters such as its type, histological grade, stage, and gene expression profiling [1]. For instance, types of breast cancer are usually classified based on the location and aggressiveness of the disease. The main types are ductal carcinoma in situ, invasive ductal carcinoma, lobular carcinoma in situ, and invasive lobular carcinoma. Invasive ductal breast cancer is the most common type of breast cancer, accounting for about 55% of all breast cancers [13, 14]. Based on its histological grade, breast cancer can be poorly differentiated, moderately differentiated, or well differentiated. Breast cancer staging is performed by combining pathological assessments on tumour size (T), spread to nearby lymph nodes (N) and distant metastases (M), commonly named as TNM staging system. TNM staging is widely used in the clinical practice with major stages range from stage I through IV while the numbers refer to the extent and evolution of the breast cancer [1]. The prognosis of invasive breast cancer is strongly influenced by the stage

4

of the disease with stage IV is always a predictor of an incurable disease as the cancer has spread to other organs of the body. Distant metastases account for over 90% of breast cancer deaths with Bone, lung, liver, and brain (Figure 1.2) ranked as the vital target organs for breast cancer metastasis [15, 16].
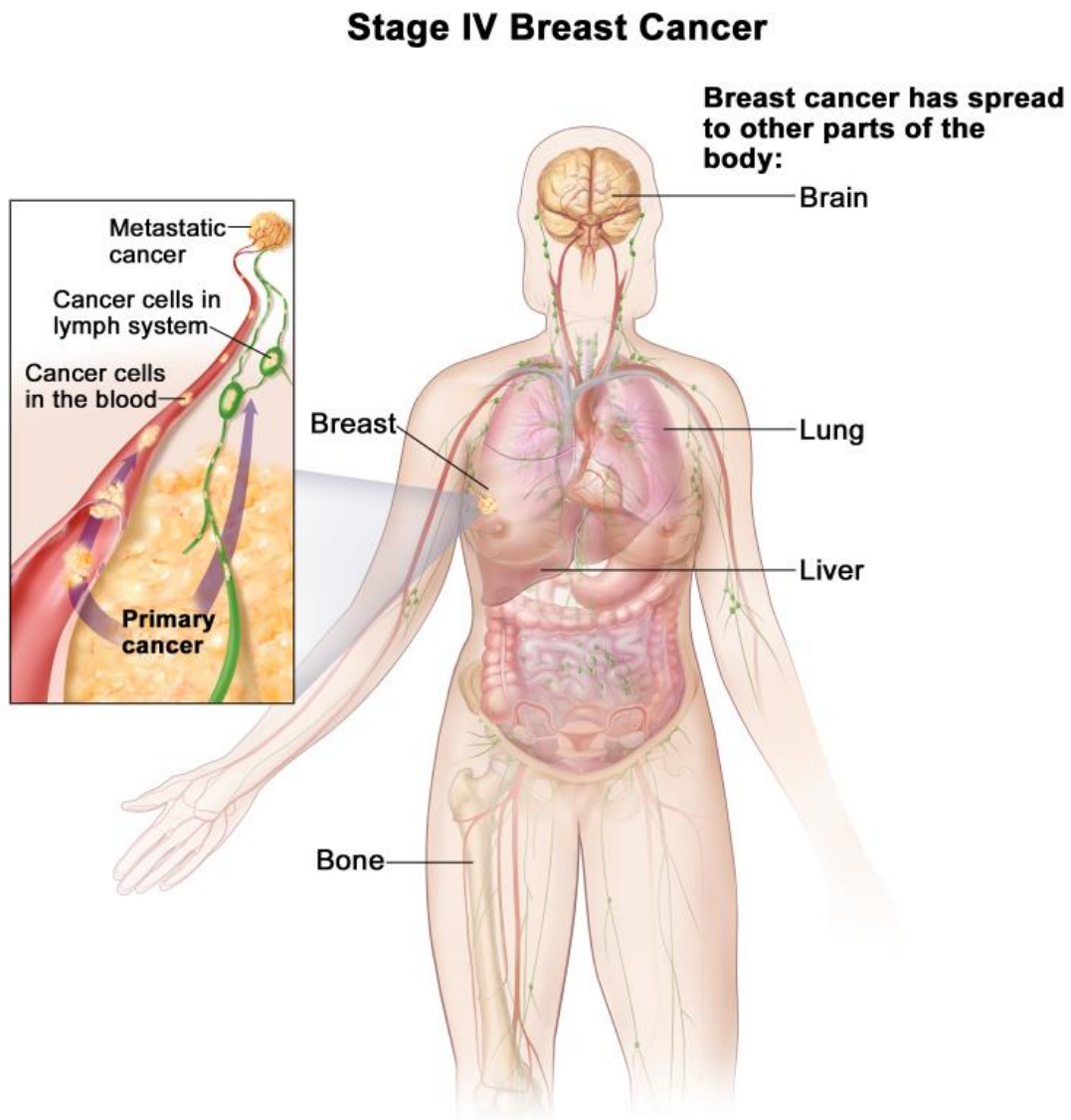
Figure 1.2. Breast cancer metastasis to other parts of the body [2].

In clinical settings, breast cancer classifications based on histopathology have been well established as prognostic and predictive factors for breast cancers. However, these traditional classifications do not fully reflect the heterogeneity of breast cancer. Hence, there has been investigation into better classifications to predict outcomes for breast cancer patients. As a

result, reshuffling of breast cancer classifications has been underway from the histopathologic type to the molecular subtype determined by microarray-based gene expression profiling [17]. Extensive gene profiling studies resulted in the identification of four molecular subtypes of breast cancer, namely luminal A, luminal B, HER2-enriched, and basal-like or triple negative breast cancers (TNBC). These molecular subtypes have been identified based on presence or absence of biological markers, including hormone (estrogen or progesterone) receptors (HR+/HR-), excess levels of human epidermal growth factor receptor 2 (HER2, a growth-promoting protein) and/or extra copies of the HER2 gene (HER2+/HER2-) [18, 19]. Each subtype has shown different incidence, prognosis, response to treatment, preferential metastatic organs, and recurrence or disease-free survival outcomes [20]. Luminal A (HR+/HER2-) cancers are slow-growing, more prevalent and less aggressive than other subtypes and associated with the most favorable prognosis, in part because they are more responsive to anti-hormone therapy. Unlike luminal A cancers, luminal B (HR+/HER2+) cancers are defined by being highly positive for Ki67 (indicator of a large proportion of actively dividing cells) or HER2. Luminal B breast cancers show higher grade and are associated with poorer survival than luminal A cancers [21]. HER2-enriched (HR-/HER2+) cancers are found in 15% to 25% of invasive breast cancers. They overexpress HER2 enabling them grow and spread more aggressively than other subtypes and are associated with poorer short-term prognosis compared to HR+ breast cancers [21, 22]. However, HER2+ cancers respond well to the recently introduced HER2-targeted therapies like trastuzumab. Triple negative (HR-/HER2-) subtypes are negative for estrogen receptor, progesterone receptor, and HER2 (ER-, PR-, HER2-). These cancers make up 10%-20% of all breast cancers [22]. They are more common among African-American black women, younger (premenopausal) women, and those with a BRCA1 gene mutation. More than 75% of TNBC fall in to the basal-like subtype defined by gene expression profiling. Triple negative breast cancers are the most aggressive with worst prognosis of all

subtypes and are associated with high mortality in women., in part because there are currently no targeted therapies for these tumors [23, 24].

1.1.3. Diagnosis of Breast Cancer

Breast cancer diagnosis at its earliest possible stage is important for its successful treatment and is therefore an important public health strategy in all settings. The 5-year survival rate after diagnosis is estimated to be 99% when the tumor is still localized, but 27% for distant-stage disease) [1]. Current methods for early-stage diagnosis of breast cancer mainly rely on mammography and magnetic resonance imaging tools. Nevertheless, in addition to their inaccurate diagnosis, many patients do not use them due to lack of access to the facility, high cost, procedural discomfort, and perceived risks of radiation exposure [25-27].

1.2.   Protein Glycosylation

Glycosylation, the process in which sugars are attached to proteins or lipids, is the most abundant and complex post-translational process, accounting for immense structural and functional variabilities in majority of eukaryotic cell proteins [28, 29]. Increasing reports described that glycosylation modulates the polarity and solubility of proteins (due to the hydrophilic nature of the glycan), controls protein tertiary structure, directs protein localization and recognition by other proteins, and prevents their cleavage by proteolytic enzymes [30]. Glycan parts of glycoconjugates are known to facilitate essential roles in almost all physio-pathological processes including fertilization, embryogenesis, cell differentiation, cell adhesion, cell recognition, molecular trafficking, signal transduction, protein folding, immunological regulation, host-pathogen interactions, aging, and even malignant alterations [28, 29, 31]. The indispensable role of glycosylation in complex organisms like humans is apparent from the fact that many eukaryotic cells can function and survive without nucleus; however, none of them can function normally without glycans at least on their surface [32-34], and hence total absence of glycans is embryologically lethal [35]. Protein glycosylation can be resulted when the glycan binds to the protein through asparagine residues (*N*-linked), or through hydroxyl group of serine or threonine residues of the proteins (*O*-linked). With highly heterogenous structural composition, *N*-linked glycans are of three types all sharing a conserved pentasaccharide core of two GlcNAc and trimannosyl (Manα1-6(Manα1-3)Manβ1-4GlcNacβ1-4GlcNacβ1-Asn) sequences. The first *N*-glycan type are called high-mannose, in which only mannose residues are attached to the core; the second type are complex *N*-glycans, which do not contain mannose residues in addition to the trimannosyl core and often carry extended and branched oligosaccharide sequences. The third type are hybrid *N*-glycans, which contain structural features of both high-mannose and complex glycans [36]. The consensus

sequence of Asn-X-Ser/Thr in a protein terminal is required for *N*-glycosylation to occur, where X can be any amino acid except proline.

1.2.1. Hexosamine Biosynthesis Pathway and *N*-glycosylation of Proteins

Following the hexoseamine biosynthetic pathway (HBP), *N*-linked glycosylation takes place in the endoplasmic reticulum (ER) and glolgi apparatus of cells, where various enzymes take part in catalyzing the growth and processing of the glycoprotein [36, 37]. After glucose entry into the cell via the glucose transporter (GLUT), it is immediately phosphorylated by hexokinase (HK) to give glucose-6-phosphate (Glc-6-P) which is further isomerized into fructose-6-phosphate (Fru-6-P), whose main metabolic fate is entering glycolysis for energy production. Alternatively, Glu-6-P is used as a substrate for the pentose phosphate pathway (PPP), the main product of which (ribose-5-phosphate) is highly important for nucleotide metabolism. Given that majority of Fru-6-P is acted by phosphofructokinase (PFK) and enters glycolysis, up to 5% of glucose influx is shunted to the HBP [38]. The HBP utilizes glucose (Glc), glutamine (Gln), acetyl-CoA, and UTP to produce uridine diphosphate-*N*-acetylglucosamine (UDP-GlcNAc) [Figure 1.3] that serves as a key substrate for the biosynthesis of various glycoconjugates including *N*-glycans, *O*-glycans, glycolipids, proteoglycans, and mucopolysaccharides. Since synthesis of UDP-GlcNAc through the HBP depends on the products of glucose, amino acid, fatty acid, and nucleotide metabolisms, levels of UDP-GlcNAc and its subsequent glycoconjugates can be metabolic sensors. In line with this, recent reports in the context of cancer revealed increased levels of the above-mentioned cellular nutrients as characteristic features of cancer cells [39]. *N*-glycan biosynthesis is a multi-step reaction that begins at the cytoplasmic surface of the ER, where the UDP-GlcNAc serves as GlcNAc-1-phosphate donor to a membrane bound lipid dolichol-phosphate, forming

a disaccharide GlcNAc-PP-dolichol, followed by addition of five mannose residues from GDP-mannose [40]. The assembled lipid-linked heptasaccharide precursor is then translocated into the lumen of ER, within which it receives four mannose and three glucose residues to produce Glc_3Man_9 $GlcNAc_2$-PP-dolichol glycan precursor. Oligosaccharyl transferases (OST) then catalyze the transfer of $Glc_3Man_9GlcNAc_2$ part from the dolichol pyrophosphate carrier to the Asn residues of nascent growing peptide having consensus sequence Asn-X-Ser/Thr [Figure 1.3]. After trimming of the two glucoses by α-glucosidases and subsequent protein folding, the glycoprotein with $Man_{8/9}GlcNAc_2$ glycan is moved out of the ER and enters the Golgi. Here, mannose residues are further trimmed by α-mannosidase I until forming $Man_5GlcNAc_2$, which is a substrate for $Man_5GlcNAc_3$ formation by *N*-acetylglucosaminyltransferase I (GnT-I) enzyme. The $Man_5GlcNAc_3$ linked glycoprotein undergoes further enzyme catalyzed reactions to form various hybrid type or complex type *N*-linked glycoproteins, whose specific glycan structure and abundance depends on the fate of the cell [40].
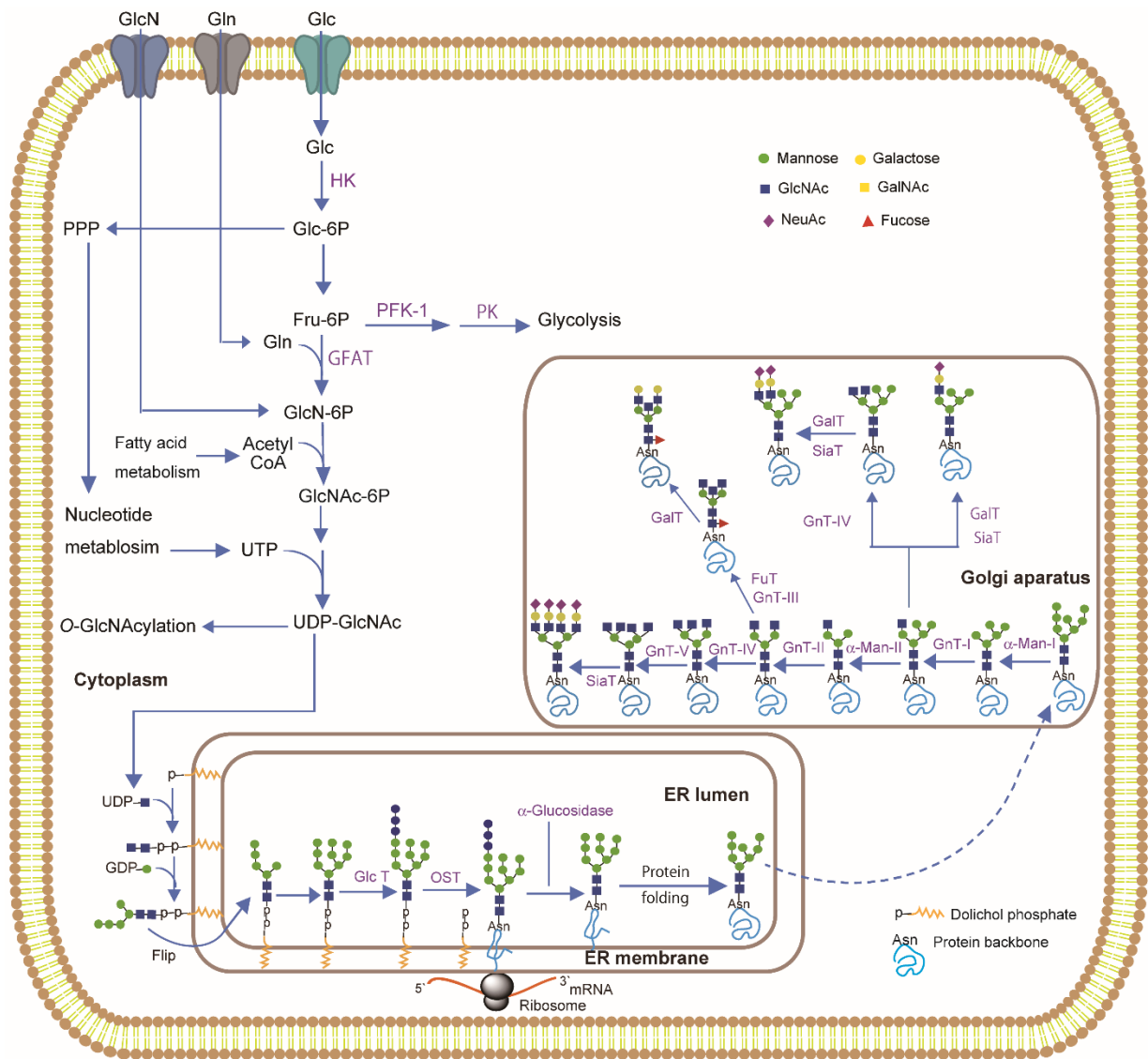
Figure 1.3. Interplay among carbohydrate, protein, fatty acid, and nucleotide metabolism for the synthesis of *N*-linked glycans through the hexosamine biosynthetic pathway. Purple color text indicates the enzymes involved in the respective steps. GFAT: glutamine: fructose-6-P amidotransferase.

1.2.2. Glycosylation and Cancer Progression

The glycosylation machinery of cells is greatly affected by malignant transformation not only due to altered expression of glycogenes [30], but also the availability of metabolic nutrients that feed the HBP [39]. Subsequent changes in glycan biosynthesis and diversity provide distinct expression or quantitative alterations in tumor-associated glycans as common hallmarks of tumor progression [41]. In cancer progression, aberrant serum *N*-glycan alterations have been associated with proliferation, invasion, metastasis, aggressiveness, angiogenesis, oncogenic signaling, and immune regulation of tumor cells [42]. The fact that most glycans and glycoconjugates are found in the outer leaflet of the cellular membrane as a dense sugar coat surrounding the cells (glycocalyx) indicates their key role in recognition and interaction processes between cells and their extracellular environment, cell-matrix interactions, mediating cell adhesion, cellular signalling, and in host-pathogen interactions. During cancer progression, the most-widely occurring cancer-associated changes in protein glycosylation are generally reported as increased sialylation (both α2,6- and α2,3-linked sialylation), increased glycan branching, and increased core-fucosylation [43]. For example, these *N*-glycosylation alterations have particularly been highlighted as oncogenic features of cell adhesion and signaling molecules such as cadherins, integrins, and receptor tyrosine kinases (RTKs) [44]. These aberrant *N*-glycosylation changes lead to impaired cell-cell adhesion (favoring cell detachment) and activation of oncogenic signaling pathways which in turn help the tumor cells to migrate and invade the surrounding tissue, blood, or lymphatic vessels, and finally endure extravasation to metastasize into different organs. Furthermore, presence of branching *N*-glycans on vascular endothelial growth factor receptor (VEGFR) modifies its interaction with galectins and is associated with tumour angiogenesis [45], whereas increased core fucosylation of EGFR enhances its dimerization and phosphorylation, leading to increased downstream signaling associated with tumour cell

growth and malignancy [46]. In several molecular studies, overexpressions of *N*-acetylglucosaminyltransferase-V (GnT-V), beta-galactoside alpha-2,6-sialyltransferase 1 (ST6GAL1), and fucosyltransferase 8 (FucT8) have been shown in a wide range of cancers, and their subsequent increased β1-6 GlcNAc branching, terminal sialylation, and core-fucosylation features were associated with increased invasiveness and metastatic potential [43, 47].

Given the complexity of glycosylation and its versatile impact on various biological and pathological processes, the diverse chemical and biological information stored in glycoconjugates is yet to be exploited well in the clinical area. In this context, profiling aberrant alterations in protein glycosylation during cancer progression not only helps in understanding their oncogenic mechanisms, but also suggests distinct glycan signatures of cancer that can be translated to diagnostic and therapeutic opportunities [44]

1.2.3. Glycomics and Cancer Biomarker Discovery

Analogous to genomics and proteomics, glycomics is a rapidly growing field that focuses on identification of the structures and functional roles of glycans and glycoconjugates produced by an organism, tissue or cell [48]. Although deciphering altered glycosylation in cancer has been of critical interest in the field of biomarker research, glycomics research still lags behind that of genomics and proteomics due to the fact that biosynthesis of glycans is template independent (genes do not encode them directly), their highly diverse structural composition and linkages, and subsequent difficulties in analytical methods [49, 49]. Tremendous advancements in analytical techniques and bioinformatics platforms have recently revolutionized the area, enabling comprehensive profiling of glycans released from various biological samples [50, 51]. In this regard, mass spectrometry and microarray-based profiling

platforms are among the crucial tools that have been providing exciting insights in the cancer biomarker discovery [52-56]. Since most glycans in an organism are found conjugated with proteins or lipids and their structure as well as abundance is regulated by metabolism which in turn is orchestrated by various enzymes coded at the gene level, cancer biomarker development using glycomics approach also requires comprehensive tools of genomics, proteomics, lipidomics, and metabolomics [57], (Figure 1.4).
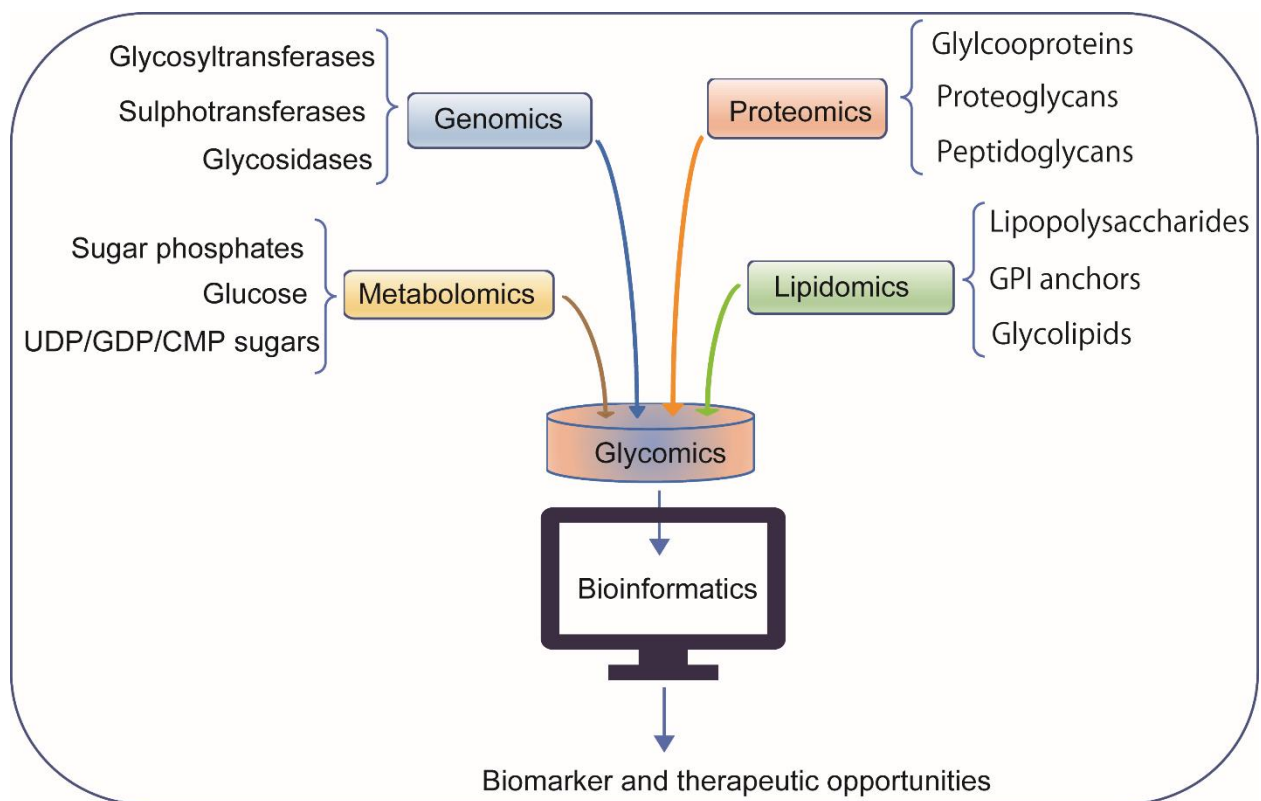
Figure 1.4. Glycomics is interconnected with other Omics fields.

A cancer biomarker is a biological molecule including DNA, transcription factor, cell surface receptor, metabolite, antibody, secreted protein or glycoprotein produced by a tumor tissue or a specific response of the body to signals from the cancer or associated inflammation. It can also be a process like angiogenesis, proliferation, or apoptosis [30]. These tumor markers are secreted, released, or leaked (as in necrotic and apoptotic processes) into the interstitial fluids and then into the circulation, where they become detectable in serum samples. From

clinical and diagnostic perspective, serum-based cancer markers are attractive option for disease prognosis, staging and monitoring due to their ease of accessibility and minimally invasive nature [58]. Given the immense challenges exist in translating cancer biomarker research into the clinical practice, a number of biomarkers have already been used in patient care mostly in combination with imaging, biopsy, and related clinicopathological features. Notably, many of the Food and Drug Administration (FDA) approved and currently used serological cancer biomarkers are glycoproteins, attesting the vital role of aberrant glycosylation in oncogenesis. The main ones include CA15-3 or CA27-29 (breast cancer), CA19-9 (pancreatic, gastric, or colonic cancer), CA-125 (ovarian cancer), CEA (Colorectal Cancer), PSA (prostate cancer), and α-Fetoprotein: (hepatocellular carcinoma) [58-62]. Apart from quantitative elevations in serum levels of the intact glycoproteins associated with cancer, glycan structure specific changes mainly increased core fucosylation of α-fetoprotein and increased core fucosylation as well as terminal sialylation of PSA *N*-glycans have been reported as improved serum markers for HCC and prostate cancers, respectively [63, 64]. However, the aforementioned cancer biomarkers have significant application limitations as their tests lack the required specificity (the percentage of healthy control samples who test negative for the biomarker) and sensitivity (the percentage of individuals with confirmed disease who test positive for the biomarker) for use in early stage detection, leading to false positive and negative diagnosis [65-67]. In addition to the high need for searching new biomarkers, this drawback strongly suggests the need for rigorous analytic and clinical validation of the cancer marker before used in the clinic [67, 68].

1.3.   Aim of the Thesis

Taking all the current pressing needs for early diagnosis and thus better treatment of breast cancer into account, the work presented in this doctoral thesis primarily aimed to comprehensively profile whole serum and purified IgG-*N*-glycome towards the discovery of glycosylation-based non-invasive markers of breast cancer. Using a high-throughput glycoblotting-assisted sample enrichment coupled with MALDI-TOF/MS-based quantitative analysis, the author performed glycomic study for the first time in native Ethiopian breast cancer patients and matched-controls, considering that glycome profile of African populations, from physiological or pathological perspective, had not previously been studied elsewhere.

The second chapter focuses on investigating total serum *N*-glcome profile by comparing the whole breast cancer patients as well as their clinical stages with the healthy control subjects. The third chapter gives further attention for IgG-focused *N*-glycomic analysis, after specifically purifying IgG fraction from serum of the study participants. In the fourth chapter, I emphasized the influence of ethnic variation on healthy human serum *N*-glycosylation patterns and identified cancer-relevant glycan biomarkers. The final chapter presents the concluding remarks of the whole thesis work.

1.4.   References

1. American Cancer Society. Breast Cancer Facts and Figures 2017-2018. Atlanta, Ga: American Cancer Society; 2017.

2. PDQ Adult Treatment Editorial Board. Breast Cancer Treatment During Pregnancy (PDQ®): Patient Version. 2018 Apr 9. In: PDQ Cancer Information Summaries [Internet]. Bethesda (MD): National Cancer Institute (US); 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK65716/ (Accessed on 11/10/2018).

3. Global Burden of Disease Cancer Collaboration. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: A systematic analysis for the global burden of disease study. JAMA Oncol. 2017; 3:524-548. doi: 10.1001/jamaoncol.2016.5688

4. Siegel R.L., Miller K.D., Jemal A. Cancer Statistics, 2018. CA Cancer J Clin. 2018; 68(1):7-30. doi.org/10.3322/caac.21442

5. Ferlay J., Soerjomataram I., Ervik M., et al. GLOBOCAN 2012 v1.0: Cancer incidence and mortality worldwide-IARC CancerBase No. 11. Lyon, France,International Agency for Research on Cancer, 2013.

6. Bellanger M., Zeinomar N., Tehranifar P., Terry M.B. Are Global Breast Cancer Incidence and Mortality Patterns Related to Country-Specific Economic Development and Prevention Strategies? Journal of Global Oncology 2018; 4:1-16.

7. Vogelstein B., Papadopoulos N., Velculescu V.E., Zhou S., Diaz L.A. Jr, Kinzler K.W. Cancer Genome Landscapes, Science 2013; 339(6127):1546-1558.

8. L. Jardines, S. Goyal, P. Fisher et al. Breast Cancer Overview: Risk Factors, Screening, Genetic Testing, and Prevention, Cancer Management Handbook, 14th edition, 2011.

9. Richardson L.C, Henley J., Miller J., Massetti G., Thomas C.C. Patterns and trends in black-white differences in breast cancer incidence and mortality – United States, 1999-2013. MMWR. 2016; 65:1093-1098.

10. Swanson GM, Haslam SZ, Azzouz F. Breast Cancer among Young African-American Women. Cancer. 2002; 97:273-279.

11. Assi H.A., Khoury K.E., Dbouk H., Khalil L.E, Mouhieddine T.H., El Saghir N.S. Epidemiology and prognosis of breast cancer in young women. J Thorac Dis. 2013; 5(Suppl 1):S2-S8.

12. Morris G.Y., Mitchell E.P. Higher Incidence of Aggressive Breast Cancers in African-American Women: A Review. Journal of the National Medical Association 2008; 100:698-702.

13. Allred D.C. Ductal carcinoma in situ: terminology, classification, and natural history. J Natl Cancer Inst Monogr. 2010; 2010:134-138.

14. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol. 2015; 8:23-31.

15. Xin J., Ping M. Targeting Breast Cancer Metastasis, Breast Cancer (Auckl). 2015; 9:23-34.

16. Kimbung S., Lomana N., Hedenfalk I. Clinical and molecular complexity of breast cancer metastases. Seminars in Cancer Biology 2015, 35:85-95.

17. Ng C.K., Schultheis A.M., Bidard F.C., Weigelt B., Reis-Filho J.S. Breast cancer genomics from microarrays to massively parallel sequencing: paradigms and new insights. J Natl Cancer Inst. 2015; 107 (5). pii: djv015. doi: 10.1093/jnci/djv015.

18. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012; 490:61-70.

19. Cheang M.C., Martin M., Nielsen T.O., et al. Defining breast cancer intrinsic subtypes by quantitative receptor expression. Oncologist 2015; 20: 474-482.

20. Sorlie T., Perou C.M., Tibshirani R., Aas T., Geisler S., Johnsen H., et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001; 98:10869-10874.

21. Haque R., Ahmed S.A., Inzhakova G., et al. Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. Cancer Epidemiol Biomarkers Prev. 2012; 21:1848-1855.

22. Cho N. Molecular subtypes and imaging phenotypes of breast cancer, Ultrasonography 2016; 35:281-288.

23. Dietze E.C., Sistrunk C., Miranda-Carboni G., O'Regan R., Seewaldt V.L. Triple-negative breast cancer in African-American women: disparities versus biology. Nat Rev Cancer 2015; 15(4): 248-254.

24. Bianchini G., Balko J.M., Mayer I.A., Sanders M.E., Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nature Rev Clin Oncol. 2016; 13:674-690.

25. Unger-Saldaña K. Challenges to the early diagnosis and treatment of breast cancer in developing countries. World J Clin Oncol. 2014; 5(3):465-477.

26. Wang L. Early Diagnosis of Breast Cancer. Sensors (Basel). 2017; 17(7):1572.

27. Schneble E.J., Graham L.J., Shupe M.P., Flynt F.L., Banks K.P., Kirkpatrick A.D., Nissan A., Henry L., Stojadinovic A., Shumway N.M., et al. Current Approaches and Challenges in Early Detection of Breast Cancer Recurrence. J Cancer. 2014; 5(4):281-290.

28. Ohtsubo K., Marth J.D. Glycosylation in cellular mechanisms of health and disease. Cell 2006; 126(5):855-867. https://doi.org/10.1016/j.cell.2006.08.019

29. Pilobello K.T., Mahal L.K. Deciphering the glycocode: the complexity and analytical challenge of glycomics. Curr Opin Chem Biol. 2007; 11(3):300-305. https://doi.org/10.1016/j.cbpa.2007.05.002

30. Silva M.L. Cancer serum biomarkers based on aberrant post-translational modifications of glycoproteins: Clinical value and discovery strategies. Biochim Biophys Acta. 2015; 1856(2):165-177.

31. Contessa J.N., Bhojani M.S., Freeze H.H., Rehemtulla A., Lawrence T.S. Inhibition of N-linked glycosylation disrupts receptor tyrosine kinase signaling in tumor cells. Cancer Res. 2008; 68(10):3803-3809. doi: 10.1158/0008-5472.CAN-07-6389

32. Ajit V. Evolutionary Forces Shaping the Golgi Glycosylation Machinery: Why Cell Surface Glycans Are Universal to Living Cells. Cold Spring Harb Perspect Biol. 2011; 3(6): a005462. doi:10.1101/cshperspect. a005462

33. Ajit V., John B.L. Biological roles of glycans: Essentials of Glycobiology. 2nd ed. Cold Spring Harbor Laboratory Press; 2009.

34. Lauc G., Krištić J., Zoldoš V. Glycans–the third revolution in evolution. Front Genet. 2014; 5:145. https://doi.org/10.3389/fgene.2014.00145

35. Marek K.W., Vijay I.K., Marth J.D. A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. Glycobiology 1999; 9(11):1263-1271. https://doi.org/10.1093/glycob/9.11.1263

36. Ajit V., Cummings R.D., Esko J.D., Freeze H.H., Stanley P., Bertozzi C.R., et al. Essentials of Glycobiology. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 2009.

37. Vasconcelos-Dos-Santos A., Oliveira I.A., Lucena M.C., Mantuano N.R., Whelan S.A., Dias W.B., Todeschini A.R. Biosynthetic Machinery Involved in Aberrant Glycosylation: Promising Targets for Developing of Drugs Against Cancer. Front Oncol. 2015; 5:138.

38. Marshall S., Bacote V., Traxinger R.R. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system: role of hexosamine biosynthesis in the induction of insulin resistance. J Biol Chem. 1991; 266:4706-4712.

39. Chiaradonna F., Ricciardiello F., Palorini R. The Nutrient-Sensing Hexosamine Biosynthetic Pathway as the Hub of Cancer Metabolic Rewiring. Cells 2018; 7(6):53. doi:10.3390/cells7060053

40. Banerjee D.K. N-glycans in cell survival and death: cross-talk between glycosyltransferases. Biochim Biophys Acta 2012; 1820(9):1338-1346. doi: 10.1016/j.bbagen.2012.01.013

41. Stowell S.R., Ju T., Cummings R.D. Protein glycosylation in cancer. Annu Rev Pathol. 2015; 10:473-510.

42. Pinho S.S., Carvalho S., Marcos-Pinto R., Magalhães A., Oliveira C., Gu J., Dinis-Ribeiro M., Carneiro F., Seruca R., Reis C.A. Gastric cancer: adding glycosylation to the equation, Trends Mol Med. 2013; 19:664-676.

43. Rodrigues J.G., Balmaña M., Macedo J.A., Poças J., Fernandes Â., de-Freitas-Junior J.C.M., Pinho S.S., Gomes J., Magalhães A., Gomes C., Mereiter S., Reis C.A. Glycosylation in cancer: Selected roles in tumour progression, immune modulation and metastasis. Cell Immunol. 2018; https://doi.org/10.1016/j.cellimm.2018.03.007

44. Pinho S.S., Reis C.A. Glycosylation in cancer: mechanisms and clinical implications, Nat Rev Cancer 2015; 15:540-555.

45. Croci, D. O., Cerliani J.P., Dalotto-Moreno T., Méndez-Huergo S.P., Mascanfroni I.D., Dergan-Dylon S., Toscano M.A., et al. Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti- VEGF refractory tumors. Cell 2014; 156:744-758.

46. Takahashi M., Kuroki Y., Ohtsubo K., Taniguchi N. Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. Carbohydr. Res. 2009; 344:1387-1390.

47. Oliveira-Ferrer L., Legler K., Milde-Langosch K. Role of protein glycosylation in cancer metastasis. Semin Cancer Biol. 2017; 44:141-152. https://doi.org/10.1016/j.semcancer.2017.03.002

48. Cummings R.D., Pierce J.M. The challenge and promise of glycomics. Chem Biol. 2014; 21(1):1-15. https://doi.org/10.1016/j.chembiol.2013.12.010

49. Hart G.W., Copeland R.J. Glycomics Hits the Big Time. Cell 2010; 143(5):672-676. https://doi.org/10.1016/j.cell.2010.11.008 PMID: 21111227.

50. Nishimura S., Niikura K., Kurogochi M., Matsushita T., Fumoto M, Hinou H., et al. High-throughput protein glycomics: combined use of chemoselective glycoblotting and MALDI-TOF/TOF mass spectroscopy. Angew Chem Int Ed Engl. 2004; 44(1):91-96. https://doi.org/10.1002/anie.200461685.

51. Nishimura S. Toward automated glycan analysis. Adv Carbohydr Chem Biochem. 2011; 65:219–271. https://doi.org/10.1016/B978-0-12-385520-6.00005-4

52. Crutchfield C.A., Thomas S.N., Sokoll L.J., Chan D.W. Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics. 2016; Jan 7;13:1. doi: 10.1186/s12014-015-9102-9.

53. de Oliveira R.M., Ornelas Ricart C.A., Araujo Martins A.M. Use of Mass Spectrometry to Screen Glycan Early Markers in Hepatocellular Carcinoma. Front Oncol. 2018; 7:328. doi: 10.3389/fonc.2017.00328.

54. Macdonald I.K., Parsy-Kowalska C.B., Chapman C.J. Autoantibodies: Opportunities for Early Cancer Detection. Trends Cancer. 2017; (3):198-213. doi: 10.1016/j.trecan.2017.02.003.

55. Hyun J.Y., Pai J., Shin I. The Glycan Microarray Story from Construction to Applications. Acc. Chem. Res.*,* 2017; 50(4):1069-1078.

56. Song X., Lasanajak Y., Xia B., Heimburg-Molinaro J., Rhea J.M., Ju H., Zhao C., Molinaro R.J., Cummings R.D., Smith D.F. Shotgun glycomics: a microarray strategy for functional glycomics. Nat Methods. 2011; 8(1):85-90. doi: 10.1038/nmeth.1540.

57. Hart G.W., Copeland R.J. Glycomics Hits the Big Time. Cell 2010; 143(5):672-676. https://doi.org/10.1016/j.cell.2010.11.008

58. Kirwan A., Utratna M., O'Dwyer M.E., Joshi L., Kilcoyne M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. Biomed Res Int. 2015; 2015: 490531. doi:10.1155/2015/490531.

59. Sharma S. Tumor markers in clinical practice: General principles and guidelines. Indian J Med Paediatr Oncol. 2009; 30(1):1-8. doi: 10.4103/0971-5851.56328.

60. Zhang D., Yu M., Xu T., Xiong B. Predictive value of serum CEA, CA19-9 and CA125 in diagnosis of colorectal liver metastasis in Chinese population. Hepatogastroenterology 2013; 60(126):1297-1301. doi: 10.5754/hge121125.

61. Kailemia M.J., Park D., Lebrilla C.B. Glycans and glycoproteins as specific biomarkers for cancer. Anal Bioanal Chem. 2017; 409(2):395-410. doi: 10.1007/s00216-016-9880-6.

62. Ueda Y., Enomoto T., Kimura T., Miyatake T., Yoshino K., Fujita M., Kimura T. Serum biomarkers for early detection of gynecologic cancers. Cancers (Basel). 2010; 2(2):1312-1327. doi: 10.3390/cancers2021312

63. Comunale M.A., Wang M., Hafner J., Krakover J., Rodemich L., Kopenhaver B., Long R.E., Junaidi O., Bisceglie A.M., Block T.M., Mehta A.S. Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. J Proteome Res. 2009; 8(2):595-602. doi: 10.1021/pr800752c.

64. Llop E., Ferrer-Batallé M., Barrabés S., Guerrero P.E., Ramírez M., Saldova R., Rudd P.M., Aleixandre R.N., Comet J., de Llorens R., Peracaula R. Improvement of Prostate Cancer Diagnosis by Detecting PSA Glycosylation-Specific Changes. Theranostics 2016; 6(8):1190-1204. doi: 10.7150/thno.15226.

65. Tsuchiya N., Sawada Y., Endo I., Saito K., Uemura Y., Nakatsura T. Biomarkers for the early diagnosis of hepatocellular carcinoma. World J Gastroenterol. 2015; 21(37):10573-10583. doi: 10.3748/wjg.v21.i37.10573.

66. Duffy M.J. Serum tumor markers in breast cancer: are they of clinical value? Clin. Chem. 2006; 52:345-351, doi: 10.1373/clinchem.2005.059832.

67. Yotsukura S., Mamitsuka H. Evaluation of serum-based cancer biomarkers: a brief review from a clinical and computational viewpoint. Crit Rev Oncol Hematol. 2015; 93(2):103-115. doi: 10.1016/j.critrevonc.2014.10.002.

68. Hayes D.F. Biomarker validation and testing. Mol Oncol. 2015; 9(5): 960-966. doi: 10.1016/j.molonc.2014.10.004.

Chapter 2

Total Serum *N*-Glycomics for Breast Cancer Biomarker Discovery

2.1. Introduction

Despite improvements in patient diagnosis and survival for most cancers, including breast cancer, to date, there is no clinically validated plasma/serum biomarker for early diagnosis of breast cancer. Consequently, there has been much interest in the development and validation of serum-based noninvasive biomarkers whose alteration in the serum precedes the appearance of a malignancy, of which there are currently none [1]. Post-translational protein glycosylation has recently received attention as a key component of cancer progression and thus unique expression or quantitative alterations in tumor-associated glycans can provide novel diagnostic and therapeutic targets [2, 3]. The fact that majority of the human serum proteins are heavily glycosylated makes them potential reservoir of glycans released from tissues and cells reflecting their physiological and pathological states. [4, 5]. Therefore, patient-based serological glycomic profiling can be an attractive approach for discovering non-invasive molecular biomarkers, allowing increased treatment options. With recent remarkable improvements in analytical tools, mass spectrometry is a widely used method to analyze glycans, proteins, or glycoproteins for clinical biomarker discovery [6]. In this regard, several studies have suggested *N*-glycan biomarkers for a wide range of cancer types including ovarian, colorectal, breast, gastric, pancreatic, prostate, and hepatocellular [7-13] cancers. However, most glycomic studies performed in the context of breast cancer were focusing on animal models, cultured cell lines, a small number of human patients, or didn't appreciate/employ the current state of methods for glycan purification and quantitative analysis from multiple types of biological samples [13-18]. Our group has recently developed a novel glycan purification and enrichment method operated in an automated machine named as "SweetBlot 7" (System Instruments, Tokyo, Japan) and designed with 96-well format workstations [19]. Using this glycoblotting method, a high-throughput and versatile approach, our laboratory previously identified serum *N*-glycan biomarkers for different pathological states including cancer,

neurodegenerative diseases, and other eye, kidney, and arthritis related disorders [10-13, 20-25].

Driven by our advanced glycoblotting technology and the unmet needs for better molecular biomarkers and drug targets, this study aimed to explore the potentials of *N*-glycans directly released from patient serum glycoproteins in predicting whole cancer patients and specific cancer stages. For this purpose, our study participants were recruited in Ethiopia where breast cancer incidence has been increasing drastically in the younger women, while inadequate and ineffective control measures are increasing the mortality rate. Our results illustrate cancer specific up-regulation of several glycoforms and suggest candidate biomarkers that reliably distinguish early stage breast cancer cases from healthy controls.

2.2. Materials and Methods

2.2.1. Study Population and Sample Collection

As summarized in Table 2.1, serum samples were collected from 115 women BC patients clinically belonged to stages I-IV and as normal controls (NC), serum samples from 33 gender, ethnic, and age matched healthy volunteers were obtained. All study subjects were Ethiopians and serum collection was performed during 2015-2016 in the Oncology Center of Black Lion Specialized Teaching Hospital (the largest and the only hospital with cancer treatment center in the country) of Addis Ababa University, Ethiopia. Informed consent had been signed and obtained from all study participants. The study was performed in accordance with the ethical standards of the Declaration of Helsinki on approval by the ethical review boards of Addis Ababa University, School of Medicine, Ethiopia and Hokkaido University, Faculty of Advanced Life Sciences, Japan. After freezed in plastic vials at -80 °C and carefully packed with dried CO_2 in foam box, serum samples were immediately transported by airplane and reached Japan within 24 hours.

2.2.2. Reagents and Materials

Ammonium bicarbonate 99% (ABC), 1-propanesulfonic acid, 2-hydroxyl-3-myristamido (PHM), 3-Methyl-1-*p*-tolyltriazene (MTT), and disialyloctasaccharide were purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). Peptide *N*-glycosidase F (PNGase F) was purchased from New England Biolabs[R] Inc. (Ipswich, Massachusetts, USA), whereas BlotGlyco H beads were purchased from Sumitomo Bakelite, Co. Ltd. (Tokyo, Japan). Dithiothreitol (DTT), iodoacetamide (IAA), *O*-benzylhydroxylamine hydrochloride (BOA), α-cyano-4-hydroxycin-namic acid (CHCA) and trypsin were from Sigma-Aldrich, Inc. (St.

Louis, MO, USA). Other reagents and solvents were obtained from Wako Pure Chemical Industries, Ltd. (Tokyo, Japan), unless otherwise stated. Protein G spin plate was purchased from Thermo Scientific. SweetBlot[TM] (automated glycan processing and incubating machine) was from System Instruments Co., Ltd. (Hachioji, Japan). Multi Screen Slvinert[R] filter plates were purchased from Millipore Co., Inc. (Tokyo, Japan). All mass quantifications were done using MALDI-TOF/MS (Ultraflex III, Brukers Daltonics, Germany).

2.2.3. Serum *N*-glycans Release and Selective Capturing by Glycoblotting

All the procedures for *N*-glycan release, purification, labeling, and spotting were performed in the SweetBlot 7 automated system (all-in-one approach) as previously described [19, 26-28]. Accordingly, whole serum or IgG fraction purified from each sample was pretreated enzymatically for producing oligosaccharides carrying reducing terminal that will enable them to chemically ligate with hydrazide-functionalized BlotGlyco[H] beads during the subsequent glycoblotting process. Hence, 10 µL of whole serum or IgG fraction was auto-transferred into a 96 well polymerase chain reaction (PCR) plate and was dissolved with freshly prepared 0.33M Ammonium bicarbonate (ABC) containing 120 mM 1, 4-dithiothreitol (DTT) and 0.4% of 1-propanesulfonic acid, 2-hydroxyl-3-myristamido (PHM) in 10mM ABC. As an internal standard, 12 µL of 60 µM disialyloctasaccharide was also added and mixed in each well, enabling us to perform eventual absolute quantification of individual glycan level. Solublized proteins were reduced by incubation at 60 °C for 30 minutes, followed by alkylation with 10 µL of 123 mM iodoacetamide in the dark at room temperature for 1 hour. The mixture was then digested with 400 U of trypsin in 1 mM HCl at 37 °C for 2 hours, and subsequently heated at 90 °C for 10 minutes to inactivate the enzyme. Soon after being cooled to room temperature, free *N*-glycans were released using 2 U of Peptide *N*-Glycosidase F (PNGase F) and incubation at 37 °C for 6 hours. In the glycoblotting-based strategy (diagrammed in Figure 2.1), 250 µL

of BlotGlyco H bead (Sumitomo Bakelite Co., Ltd., 10 mg/mL suspension with water) was placed into a well of a Multi Screen Solvinert filter plate (Millipore), with vacuuming to remove the water. 20 µL of PNGase F digested mixture containing released *N*-glycans was mixed with the bead in each well, followed by the addition of 180 µL of 2% acetic acid (AcOH) in acetonitrile (ACN). To capture the *N*-glycans in sample mixtures specifically onto beads via reversible hydrazone bonds, the plate was incubated at 80 °C for 45 minutes. Then, the beads were washed twice with each 200 µL of 2 M guanidine-HCl in ABC, water, and 1% triethyl amine in methanol. Unreacted hydrazide functional groups on beads were capped by incubation with 10% acetic anhydride in MeOH for 30 minutes at room temperature. After removing the solution by vacuum, the beads were successively washed twice with each 200 µL of 10 mM HCl, MeOH, and dioxane. To stabilize sialic acids, on-bead methyl esterification of carboxyl groups in sialic acids was carried out by incubation with fresh 100 mM 3-methyl-1-*p*-tolyltriazene (MTT) in dioxane at 60 °C for 90 minutes. Each well was then washed twice using 200 µL of dioxane, water, methanol, and water again. The glycans captured on beads were subjected to the trans-iminization reaction with 20 µL of 50 mM *O*-benzyloxyamine hydrochloride (BOA) and with mild acid hydrolysis of hydrazone bond using 180 µL of 2% AcOH in ACN at 80 °C 45 minutes. Finally, the released and BOA-tagged *N*-glycans were eluted with 75 µL of water under vaccum and applied for subsequent MALDI-TOF/MS analysis.
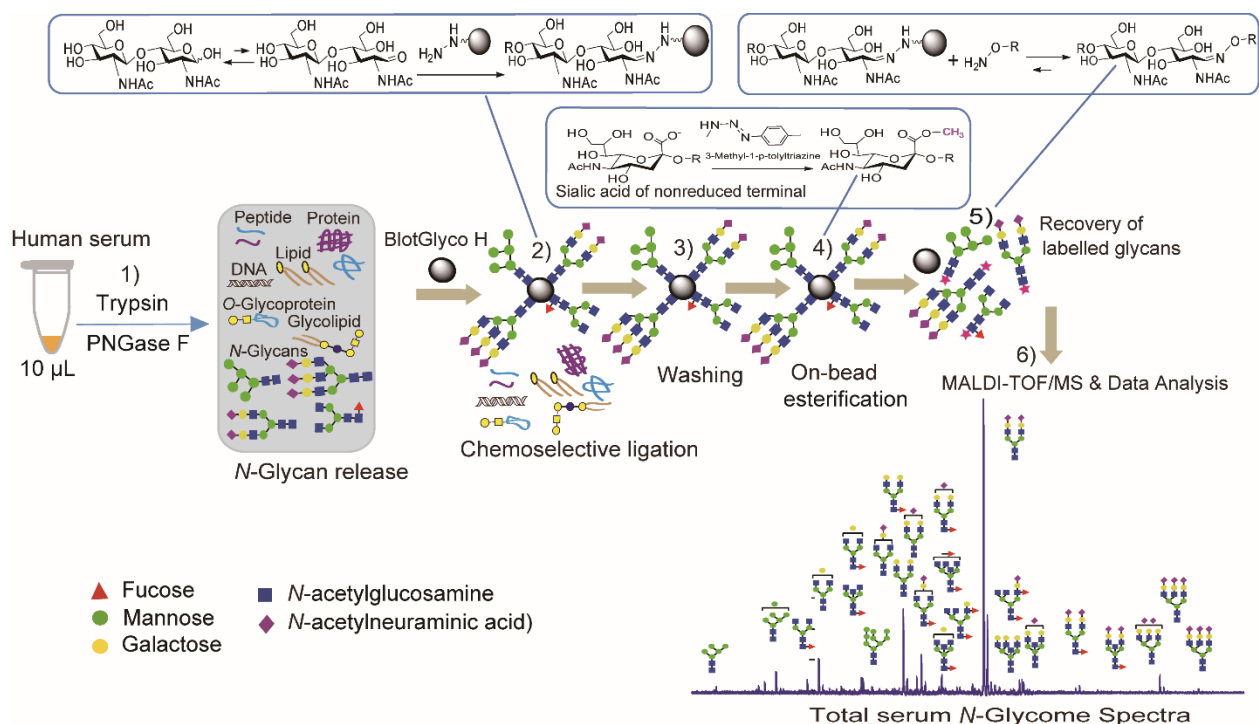
Figure 2.1. Schematic workflow of glycoblotting-based MALDI-TOF/MS analysis of *N*-glycomes derived from serum glycoproteins. **1**) Enzymatic release of glycans directly from serum **2**) Chemoselective capturing of reducing sugars onto a hydrazide-functionalized BlotGlyco H beads, **3**) Washing to remove all impurities, **4**) On-bead methyl esterification of sialic acid residues, **5**) Recovery of BOA-labeled glycans, **6**) MALD-TOF/MS and data analysis.

2.2.4. *N*-glycan Analysis by MALDI-TOF/MS

BOA-labeled *N*-glycans were directly dissolved with an equivalent volume of matrix solution (100 mM α-cyano-4-hydroxycin-namic acid diethylamine salt), after which 2.5 µL of each sample-matrix mixture was auto-spotted in quadruplicate on MTP 384 target plate (polished steel TF, Bruker Daltonics). Ultraflex III mass spectrometry that works based on matrix-assisted laser desorption/ionization-time of flight/mass spectrometry (MALDI-TOF/MS) was used during which mass spectra were acquired in an automated manner using

AutoXecute flexControl software (Bruker Daltonics, Germany) in reflector, positive ion mode, typically summing 1000 shots for each spot. The obtained mass spectra were further analyzed using FlexAnalysis v. 3 Software (Bruker Daltonics, Germany). The intensities from monoisotopic peaks of each quadruplicated spectra were normalized using known concentration of an internal standard and then averaged. This data was used for further statistical and quantitative comparison. Detected *N*-glycans were selected based on their quantitative reproducibility after evaluated using calibration curve of serially diluted human serum standards. By using experimental masses, *N*-glycan compositions were assigned to each peak from an online available GlycoMod Tool (http://br.expasy.org/tools/glycomod/).

2.2.5. Statistical Analysis

Data for *N*-glycan expression level were analyzed using SPSS software. Multiple comparisons among normal control and clinical stage groups were performed using a bonferroni one-way analysis of variance (ANOVA) while comparison between two (NC vs whole BC) groups was performed using Independent-Samples T-Test. Mean value differences were considered to be significant at 95% confidence interval ($p \leq 0.05$). The diagnostic potential of significantly differed individual glycans or glyco-subclasses was further analyzed by receiver operator characteristic (ROC) test. Area under the curve (AUC) value generated from ROC test was used to assess the diagnostic accuracy of potential glycan biomarkers. Accordingly, AUC values of 0.9-1, 0.8-0.9, 0.7-0.8, and <0.7 indicate a "highly accurate", "accurate", "moderately accurate", and uninformative test, respectively. We also used Graph Pad prism 5 software to show the data distribution in scatter dot plot.

2.3. Results

2.3.1. Patient Characteristics

Descriptive information on the selected study participants is presented in Table 2.1. All the data were collected during the time of diagnosis and individuals who were pregnant, were receiving any treatment, or were with a medical history of disease which can affect the glycan profile, were excluded from the study. Age and gender were matched among the clinical groups. However, with the majority of the patients belonging to the young age, still the control group was found to be slightly younger (Table 2.1A). Overall, the relatively younger age of the study population (uncommon in most prior cancer studies) can be taken as an advantage to rule out the impact of aging on glycan profile. Ductal carcinoma was found to be the most common breast cancer type, comprising about 78% of the cases (Table 2.1B).

Table 2.1. Demographic characteristics of all study participants and Clinico-pathological features of breast cancer patients

A. *Demographic characteristics of normal controls and breast cancer patients of stages I-IV*									
		Mean ± SD			**Number of subjects per age range**				
Status	n (%)	Age at diagnosis	BMI (kg/m2)	Age at Menarche	20-35	36-45	46-55	56-65	>65
NC	24 (20.17)	31.54 ± 7.17	22.17 ± 2.97	14.29 ± 1.46	16	7	1		
BC-I	19 (15.97)	42.74 ± 14.04	22.14 ± 2.52	15.32 ± 1.53	8	3	3	5	
BC-II	23 (19.33)	43.7 ± 12.76	22.01 ± 3.01	15 ± 1.38	6	10	3	3	1
BC-III	25 (21)	41.4 ± 12.02	21.78 ± 3.02	14.48 ± 1.69	9	6	8	1	1
BC-IV	28 (23.53)	43.25 ± 13.84	22.89 ± 2.92	14.39 ± 1.4	10	8	6	2	2
Total (N)	119				49	34	21	11	4

| | | Histological | | Site of | | Type of | |
Tumor Size	n (%)	Grade	n (%)	lesion	n (%)	BC	n (%)
< 2 cm	17 (17.9)	Poorly differentiated	25 (26.3)	Right	33 (34.7)	Ductal	74 (77.9)
2-5 cm	38 (40)	Moderately differentiated	49 (51.6)	Left	51 (53.7)	Lobular	16 (16.8)
> 5cm	30 (31.6)	Well differentiated	21 (22.1)	Bilateral	11 (9.1)	Mucinous	5 (5.3)
Metastasized	10 (10.5)						

Pathological investigations for tumor stage, tumor size, and metastasis were performed based on TNM classification system. n: frequency, NC: normal control, BC: breast cancer, SD: standard deviation, BMI: body mass index

2.3.2. Quantitative Reproducibility Test

Before selection of the detected N-glycans, each peak was evaluated for its quantitative reproducibility using serially diluted standard human serum samples that had been experimented in the same plate beside the patient samples. The peak area of each glycan detected in the dilution series (0.5×, 0.75×, 1×, 1.25×, 1.5×, 1.75×, 2×, and 2.25×) was normalized using known concentration of an internal standard, after which a standard calibration curve was plotted for each glycan as shown in Figure 2.2. Since each serum or IgG sample was spotted in quadruplicates during matrix-assisted laser desorption/ionization-time of flight/mass spectrometry (MALDI-TOF/MS) analysis, four normalized spectral data for each sample were averaged before statistical analysis. Quantitative reliability was then judged based on minimum outlier scores, good slope linearity, and significant level of the correlation

coefficient across the standard curve. Accordingly, glycan peaks meeting these criteria were selected and considered for further statistical analysis in the main study samples.
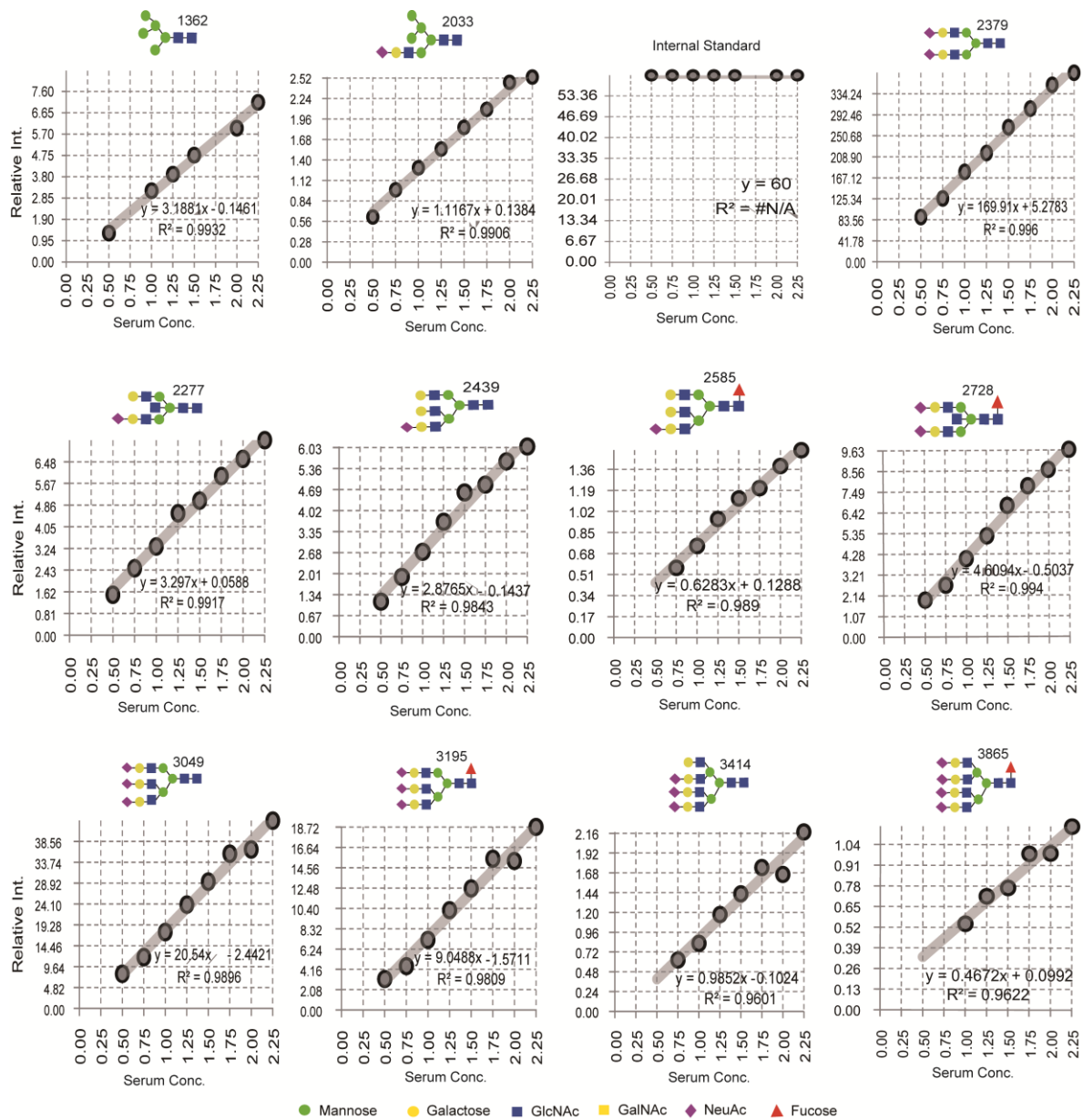
Figure 2.2. Human serum standard calibration curve showing quantitative reproducibility of some common N-glycans. Each glycans are coded by numbers showing their molecular weight (m/z value) placed near to their structure.

2.3.3. Total Serum *N*-glycan Profiles

Using an integrated protocol for glycan release and purification, coupled with MALDI-TOF/MS analysis, 46 N-glycans (Table 2.2) could be detected and quantitatively verified in the serum of entire study samples. Structurally, 37 (80.43%) of the detected *N*-glycans belong to complex type whereas high-mannose and hybrid types comprise 5 (10.87%) and 4 (8.9%), respectively. The large-scale mass spectra of total serum *N*-glycans (Figure 2.3) have shown reproducible spectral patterns with nonnegligible differences in the peak intensity of many glycans (shown in Figure 2.4) between patients and controls.

Table 2.2. Estimated composition of 46 N-linked glycans identified from human serum glycoproteins.

Pea#	m/z	Glycan composition	Type	In IgG fraction
1	1362.48109	(Hex)2 + (Man)3 (GlcNAc)2	High-Mannose	
2	1444.53419	(HexNAc)2 + (Man)3 (GlcNAc)2	Complex	+
3	1524.53392	(Hex)3 + (Man)3 (GlcNAc)2	High-Mannose	+
4	1565.56047	(Hex)2 (HexNAc)1 + (Man)3 (GlcNAc)2	Hybrid	+
5	1590.59210	(HexNAc)2 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	+
6	1606.58702	(Hex)1 (HexNAc)2 + (Man)3 (GlcNAc)2	Complex	+
7	1686.58675	(Hex)4 + (Man)3 (GlcNAc)2	High-Mannose	
8	1708.61871	(Hex)1 (HexNAc)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
9	1752.64493	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	+
10	1768.63985	(Hex)2 (HexNAc)2 + (Man)3 (GlcNAc)2	Complex	+
11	1793.67148	(HexNAc)3 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	+
12	1848.63958	(Hex)5 + (Man)3 (GlcNAc)2	High-Mannose	
13	1854.67662	(Hex)1 (HexNAc)1 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
14	1870.67154	(Hex)2 (HexNAc)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Hybrid	+
15	1911.69809	(Hex)1 (HexNAc)2 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
16	1914.69776	(Hex)2 (HexNAc)2 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	+
17	1955.72431	(Hex)1 (HexNAc)3 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	+
18	2010.69241	(Hex)6 + (Man)3 (GlcNAc)2	High-Mannose	
19	2032.72437	(Hex)3 (HexNAc)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Hybrid	

20	2057.75600	(Hex)1 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	+
21	2073.75092	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
22	2117.77714	(Hex)2 (HexNAc)3 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	+
23	2175.78261	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3 (GlcNAc)1	I.S	
24	2219.80883	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	+
25	2260.83538	(Hex)1 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
26	2263.83505	(Hex)2 (HexNAc)3 (Deoxyhexose)2 + (Man)3 (GlcNAc)2	Complex	
27	2276.83030	(Hex)2 (HexNAc)3 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
28	2336.85144	(Hex)3 (HexNAc)4 + (Man)3 (GlcNAc)2	Complex	
29	2378.86199	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
30	2422.88821	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	+
31	2438.88313	(Hex)3 (HexNAc)3 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
32	2483.89335	(Hex)3 (HexNAc)1 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Hybrid	
33	2520.93623	(Hex)1 (HexNAc)5 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
34	2524.91990	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
35	2581.94137	(Hex)2 (HexNAc)3 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
36	2584.94104	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
37	2727.99928	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
38	2743.99420	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
39	2890.05211	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
40	3007.09472	(Hex)4 (HexNAc)5 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
41	3049.10527	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
42	3109.12641	(Hex)4 (HexNAc)4 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
43	3195.16318	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
44	3414.23748	(Hex)4 (HexNAc)4 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
45	3560.29539	(Hex)4 (HexNAc)4 (Deoxyhexose)1 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
46	3719.34855	(Hex)4 (HexNAc)4 (NeuAc)4 + (Man)3 (GlcNAc)2	Complex	
47	3865.40646	(Hex)4 (HexNAc)4 (Deoxyhexose)1 (NeuAc)4 + (Man)3 (GlcNAc)2	Complex	

The m/z values represent experimental masses of N-glycans tagged with O-benzylhydroxylamine hydrochloride. Peak number 23 is an internal standard (I.S: disialyloctasaccharide). The "+" sign in the last column indicates those glycans which were also detected in IgG fraction of the samples (see details in chapter 3).
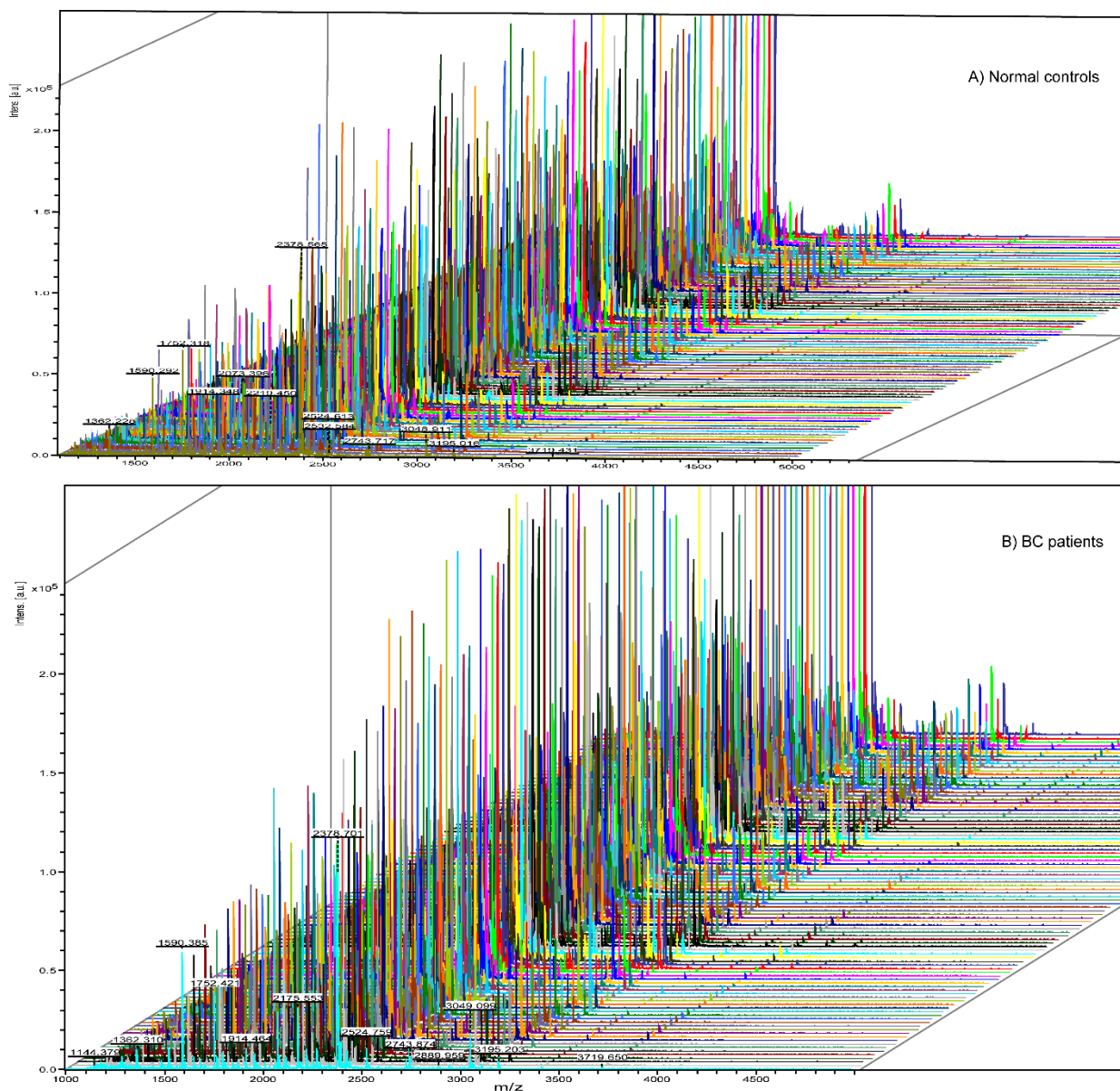
Figure 2.3. Stacked-view MALDI-TOF Mass Spectra from large-scale serum *N*-glycomics of normal controls (**A**) and BC patients (**B**). The raw mass spectra were further subjected to compositional analysis and structural elucidation.
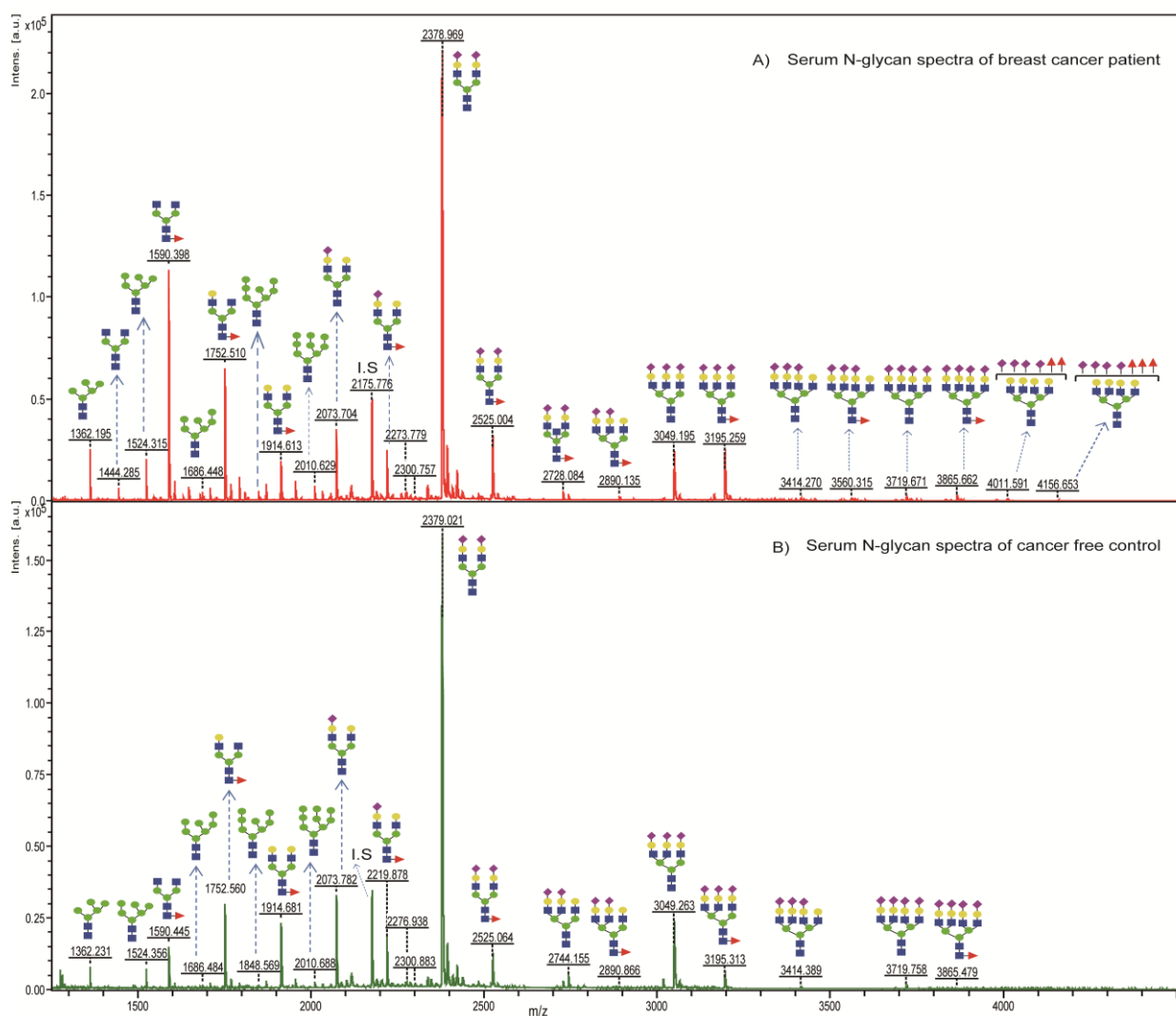
Figure 2.4. Representative MALDI-TOF/MS spectra of serum N-glycans derived from individuals with BC (A) and free of BC (B). Greater peak intensities appeared in the patient spectra imply the up-regulation of serum level of the respective glycoforms. The two glycans with m/z of 4011.591 and 4156.663 were not considered for quantitative comparison since their detection level was not quantitatively reproducible in all patients.

2.3.4. Serum Glycan Abundance Noticed Greatly in Breast Cancer Patients

The serum expression levels and structures showing monosaccharide compositions of the entire detected *N*-glycans are shown in Figure 2.5. To identify individual glycans that showed

differential expression patterns between patients and controls, we performed independent-samples T-test using the quantified data of each *N*-glycan peak. The majority of *N*-glycans exhibited up-regulated expression in which the levels of 35 N-glycans were significantly higher in the sera of whole BC patients compared to the controls (Table 2.3). In contrary, only two glycans showed reduction in the patient group in a slightly significant (m/z 1915, p = 0.048) and a non-significant (m/z = 2220) manner.
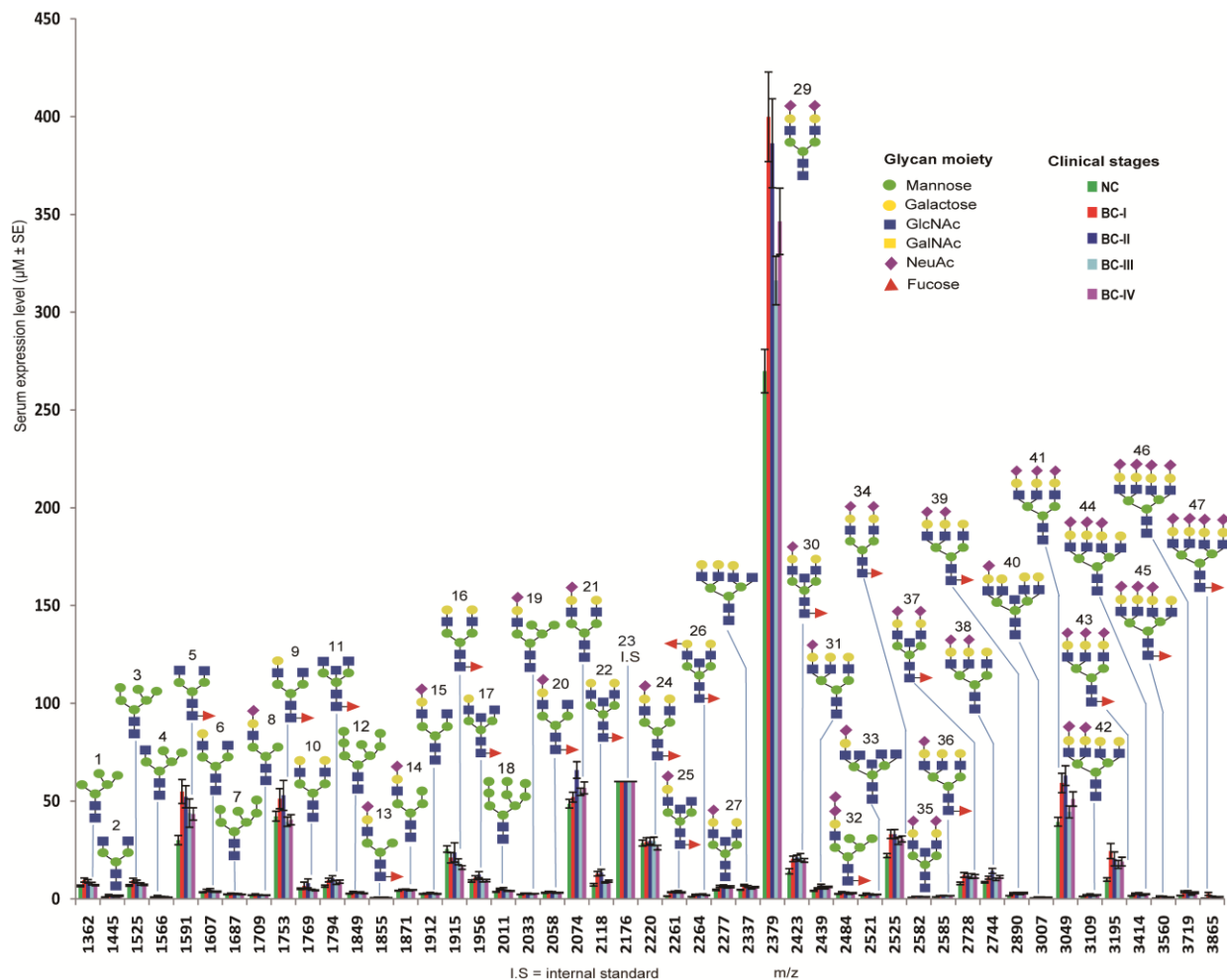
Figure 2.5. Expression levels and structures of total *N*-glycans isolated from healthy and breast cancer human serum glycoproteins. Peak numbers are indicated over the structures. Peak number 23 detected at m/z 2176 was an internal standard (I.S) co-experimented with each sample in order to allow absolute quantification of the glycan level.

Table 2.3. List of serum *N*-glycans whose expression level statistically differed between whole cancer patients (BC) and normal controls (NC).

| Peak # | m/w | Serum Expression Level (AVG μM ± SE) | | *p*-value (t-test) |
		NC	BC	
1	1362	6.65 ± 0.44	8.20 ± 0.34	0.042
2	1445	0.79 ± 0.11	1.55 ± 0.14	P<0.0001
5	1591	30.07 ± 2.39	47.94 ± 2.64	P<0.0001
11	1794	6.53 ± 0.55	9.26 ± 0.49	0.001
13	1855	0.55 ± 0.05	0.70 ± 0.03	0.033
14	1871	4.20 ± 0.15	4.57 ± 0.08	0.047
16	1915	25.54 ±1.62	19.68 ± 1.34	0.048
18	2011	3.57 ± 0.17	4.53 ± 0.17	0.009
19	2033	2.20 ± 0.11	2.63 ± 0.06	0.002
21	2074	48.64 ± 2.22	57.49 ± 1.58	0.011
22	2118	7.21 ± 0.66	10.92 ± 0.52	0.002
25	2261	1.43 ± 0.11	3.50 ± 0.18	P<0.0001
26	2264	1.14 ± 0.09	2.10 ± 0.10	P<0.0001
27	2277	4.60 ± 0.33	6.29 ± 0.24	0.001
28	2337	4.65 ± 0.20	6.10 ± 0.19	P<0.0001
29	2379	269.88 ± 11.06	358.85 ± 9.81	P<0.0001
30	2423	14.09 ± 1.34	20.63 ± 0.73	P<0.0001
31	2439	4.12 ± 0.20	6.01 ± 0.20	P<0.0001
32	2484	2.53 ± 0.12	2.99 ± 0.10	0.008
33	2521	1.86 ± 0.09	2.32 ± 0.07	P<0.0001
34	2525	22.11 ± 1.07	31.46 ± 0.98	P<0.0001
35	2582	0.71 ± 0.05	1.13 ± 0.05	0.001
36	2584	1.00 ± 0.07	1.57 ±0.05	P<0.0001
37	2728	7.89 ± 0.62	11.81 ± 0.51	P<0.0001
38	2744	8.49 ± 0.43	11.58 ± 0.44	P<0.0001
39	2890	1.60 ± 0.10	2.80 ± 0.16	P<0.0001
40	3007	0.46 ± 0.04	0.77 ± 0.03	P<0.0001
41	3049	39.42 ± 2.27	53.86 ± 2.10	P<0.0001
42	3109	1.25 ± 0.07	1.98 ± 0.10	P<0.0001
43	3195	9.88 ± 0.91	20.10 ± 1.36	P<0.0001
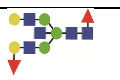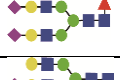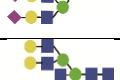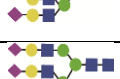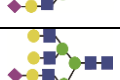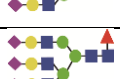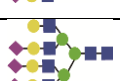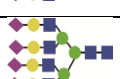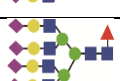
44	3414	1.50 ± 0.11	2.44 ± 0.12	P<0.0001
45	3560	0.33 ± 0.03	1.00 ± 0.07	P<0.0001
46	3719	1.66 ± 0.16	3.16 ± 0.20	P<0.0001
47	3865	0.33 ± 0.03	1.31 ± 0.14	P<0.0001

All glycan peaks except m/z 1915 showed increased abundance (p≤0.05) in the BC patients compared to NC. Peak numbers (peak #) are according to Figure 2.5, as shown over each glycan structure

2.3.5. Evaluation for Diagnostic Performance of Selected Serum *N*-glycans

The serum gylcans that had shown significant (p≤0.05) expression alteration during ANOVA or T-test analysis were further considered for receiver operating characteristic (ROC) test. Their ability to distinguish cancer patients from controls was then evaluated based on their area under the curve (AUC) value generated from ROC analysis. As a result, 17 serum N-glycans with highest diagnostic performance (detailed in Table 2.4) were suggested as biomarkers for different BC stages and/or entire cancer patients. All these glycan structures belong to complex type N-glycans that include 4 bisecting, 8 core fucosylated, and 12 sialylated to a different degree with multiply branched antennas. Each of these identified glyco-biomarkers could demonstrate significant abundance with "highly accurate" (AUC: 0.9-1) or "accurate" (AUC: 0.8-0.9) diagnostic performance at an early stage (stage I, II, or both) or all patients, whereas no powerful candidate biomarker was obtained exclusively for the late stages (BC-III or BC-IV) of the disease. As shown in Table 2.4 and Figure 2.7, only two glycans that have bisected and fucosylated structures (m/w 2261 and 2264) for BC-III, and 3 monosialylated glycans (m/w 2261, 2439, 3007) for BC-IV met the biomarker criteria and adequately distinguished the patients from the controls.

Table 2.4. Details of suggested serum glyco-biomarkers for breast cancer stages (I-IV) based on ROC analysis.

		Breast Cancer (BC) Patients vs Healthy Controls										
m/z	Glycan structure	BC-I (n=19)		BC-II (n=23)		BC-III (n=25)		BC-IV (n=24)		Whole BC (N=95)		Can be Biomarker for
		AUC	*p*	AUC	*p*	AUC	*p*	AUC	*p*	AUC	*p*	
1591		**0.833**	**0.014**	0.781	0.027	0.630	0.538	0.744	0.555	0.741	*P<0.0001*	I
2118		**0.858**	**0.006**	**0.814**	**0.001**	0.681	1.00	0.708	1.00	0.761	0.002	I, II
2261		**0.958**	**0.029**	**0.961**	**0.005**	**0.886**	**0.006**	**0.970**	**0.02**	**0.945**	*P<0.0001*	I- IV, whole BC
2264		0.846	0.13	**0.885**	**0.023**	**0.856**	**0.017**	0.843	0.309	**0.857**	*P<0.0001*	II, III, whole BC
2337		**0.873**	*P<0.0001*	0.779	0.007	0.650	1.00	0.768	0.034	0.761	*P<0.0001*	I
2379		**0.868**	*P<0.0001*	**0.844**	*P<0.0001*	0.725	0.60	0.772	0.016	0.796	*P<0.0001*	I, II
2439		0.771	0.359	**0.881**	*P<0.0001*	0.757	0.114	**0.862**	*P<0.0001*	0.821	*p<0.0001*	II, IV, whole BC
2525		**0.84**	0.001	0.795	*P<0.0001*	0.773	0.035	**0.814**	**0.007**	0.806	*p<0.0001*	I, whole BC
2744		0.765	0.589	**0.82**	*P<0.0001*	0.643	1.00	0.738	0.136	0.739	*P<0.0001*	II
3007		**0.875**	*P<0.0001*	**0.883**	*P<0.0001*	0.692	0.603	**0.84**	**0.004**	0.818	*P<0.0001*	I, II, IV, whole BC
3049		**0.827**	**0.009**	**0.815**	*P<0.0001*	0.587	1.00	0.692	0.286	0.721	*P<0.0001*	I, II
3109		0.752	0.285	**0.839**	**0.001**	0.618	0.954	0.796	0.033	0.751	*P<0.0001*	II
3195		**0.80**	**0.003**	0.723	0.05	0.755	0.347	0.778	0.102	0.763	*P<0.0001*	I
3414		**0.816**	**0.016**	**0.837**	*P<0.0001*	0.576	1.00	0.778	0.043	0.747	*P<0.0001*	I, II
3560		**1**	**0.003**	**1**	**0.018**	0.882	0.752	0.935	0.348	**0.944**	*P<0.0001*	whole BC
3719		**0.826**	**0.006**	**0.871**	**0.001**	0.621	1.00	0.745	0.035	0.762	*P<0.0001*	I, II
3865		**0.985**	**0.001**	1	0.846	0.963	1.00	0.973	0.927	**0.977**	*P<0.0001*	I, whole BC

AUC value ≥ 0.8 and level of significant, *p*≤0.05 at 95% CI were considered as double criteria

to evaluate the distinguishing potential of a glycan between patients and normal controls. (AUC

value vs diagnostic accuracy: 0.9-1 = "highly accurate", 0.8-0.9 = "accurate", 0.7-0.8 = "moderately accurate", <0.7 = "uninformative test"). The bold AUC and p-values indicate the glycans at the respective stage could fulfill both the criteria and thus were selected as candidate biomarkers.

Dot plots and ROC curves for some of the identified candidate biomarkers of stage I, stage II, or all BC patients compared with NC are diagramed in Figure 2.6. Strong correlation with the disease stages was observed in which a substantial increase in the glycans abundance was predominant at an early stage of the disease. Comparing with the control group, individual expression scores within the cancer groups have shown dispersed distribution.
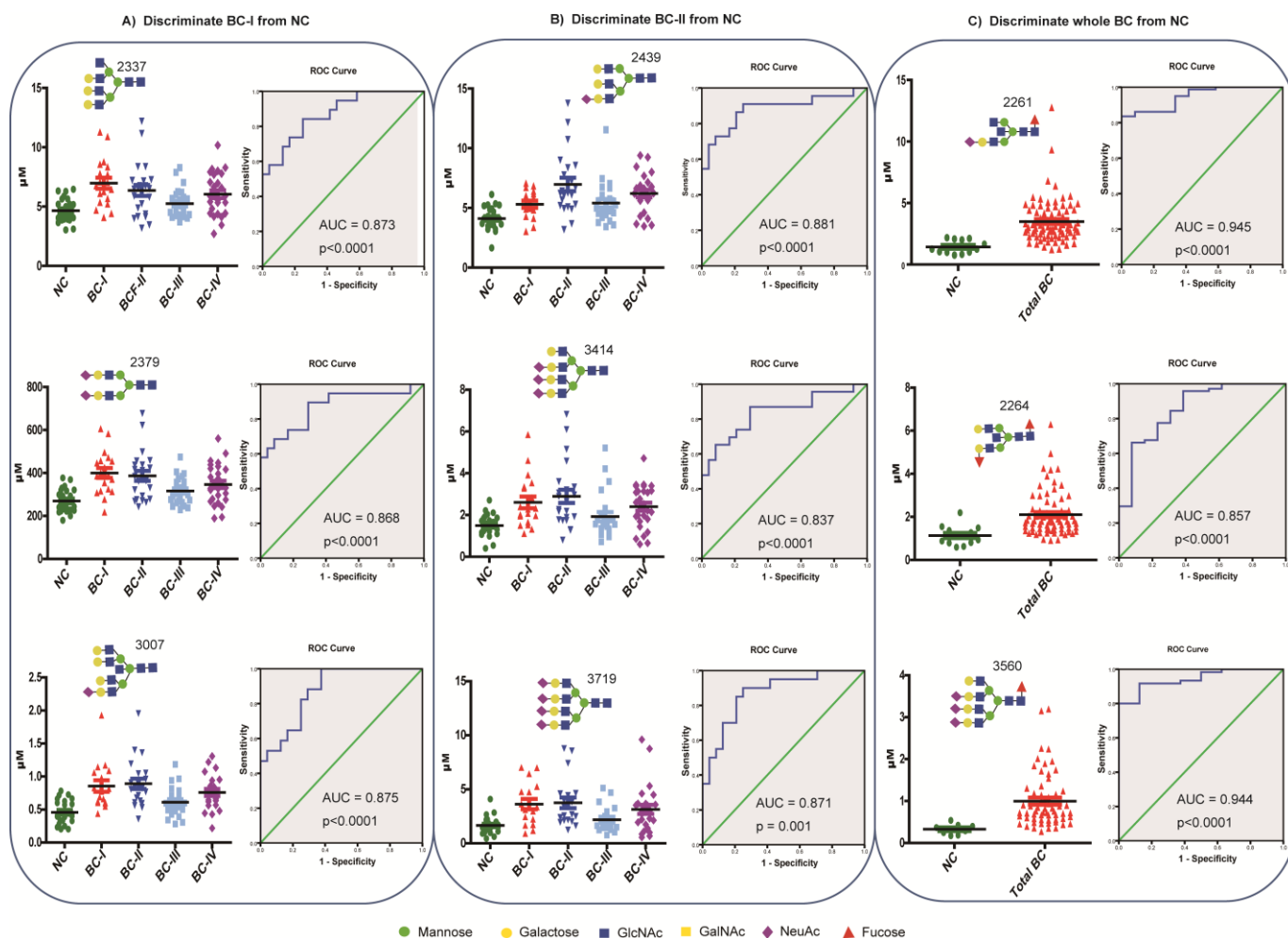
Figure 2.6. Dot plot expression and ROC curves of selected serum N-glycans up-regulated in BC patients comparing with NC. Area under the curve (AUC) value is displayed to indicate the discrimination power of the glycan between the patients and the controls.

As clearly stated in Figure 2.7, discriminating power of the candidate biomarkers that fulfilled both criteria ($p \leq 0.05$ and AUC ≥ 0.8) was either specific to stage I and stage II or shared by multiple cancer stages. Within the 17 selected candidate biomarkers, 13 glycoforms could predict and classify BC stage I patients from controls while 12 glycoforms did similarly for BC state II patients.
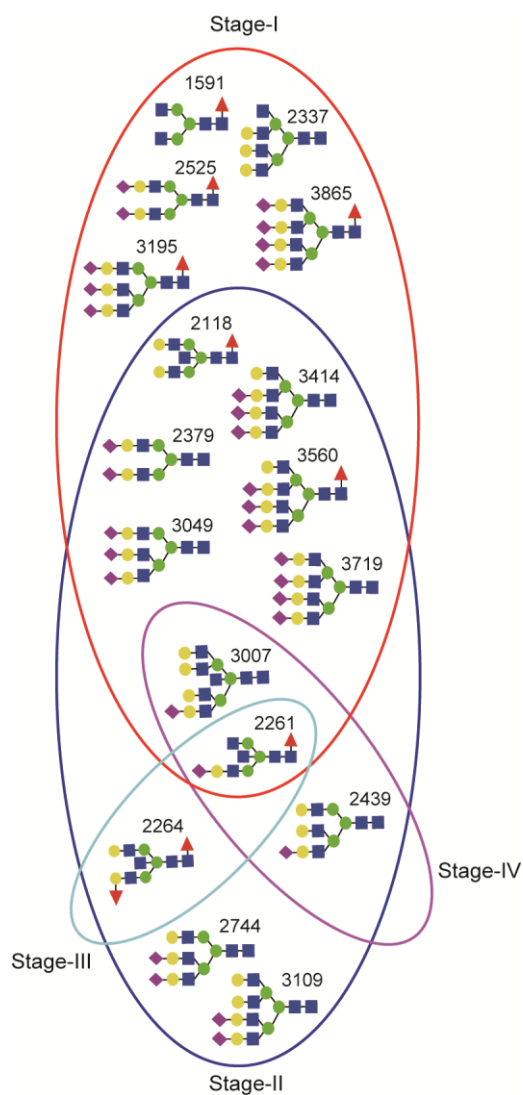
Figure 2.7. Venn diagram showing BC stage specific and overlapping N-glycans identified and selected as candidate biomarkers.

As an important observation, the most branched and sialylated structures have shown marked abundance in the serum of all cancer stages with a greater increase at an early stage (Figure 2.8). Apart from being multiply sialylated, the first (m/z 3049 vs 3195), the second (m/z 3414 vs 3560), and the third (m/z 3719 vs 3865) pairs are biosynthetically consistent only differed by core-fucose moiety.
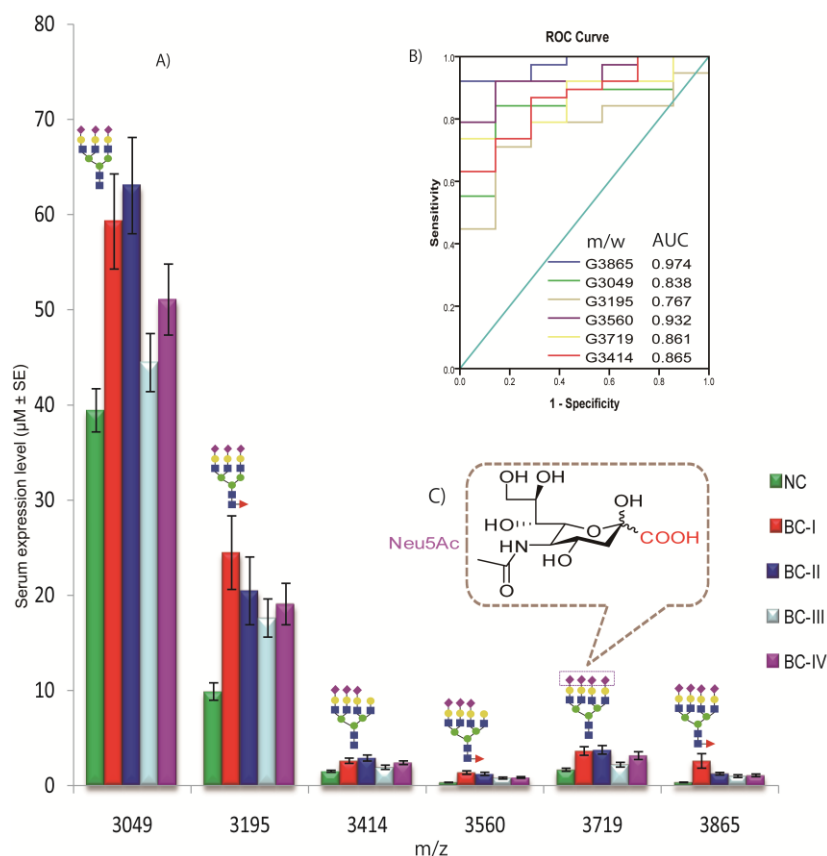
Figure 2.8. Serum expression abundance of hyperbranched and hypersialylated glyco-biomarkers. A) Glycan expression level of the NC and BC stages of I-IV, B) ROC curve and AUC value for their potential to distinguish the whole BC patients from NC, C) chemical structure of sialic acid whose carboxyl group (COOH) has functional significance for cancer cell migration and metastasis.

2.3.6. Serum Glyco-subclasses Predicting Early Stage Breast Cancer

Apart from individual serum *N*-glycan alteration, group of glycotypes sharing structural features were also compared among different BC stages and controls. With no glyco-subclass altered significantly at stage III, we found statistically significant abundance in core-fucosylated, bisecting, bi-, tri-, and tetra-antennary, bi-, tri-, and tetra-sialylated *N*-glycans in the serum of stages I and II patients. From ROC analysis, greater diagnostic accuracy to predict early-stage BC was noticed in bi-antennary, bi-sialylated, tri-antennary, and tri-sialylated

glycans (AUC = 0.895, .89, 0.863, 0.884, respectively) for stage I patients at p<0.001. Moreover, tri-antennary, and tri-sialylated glycans demonstrated an AUC value of 0.88 and 0.902 to distinguish stage II patients from NC (Figure 2.9). Larger abundance of these glycotypes associated with the early stages of the disease agrees with the aforementioned expression pattern of individual candidate biomarkers.
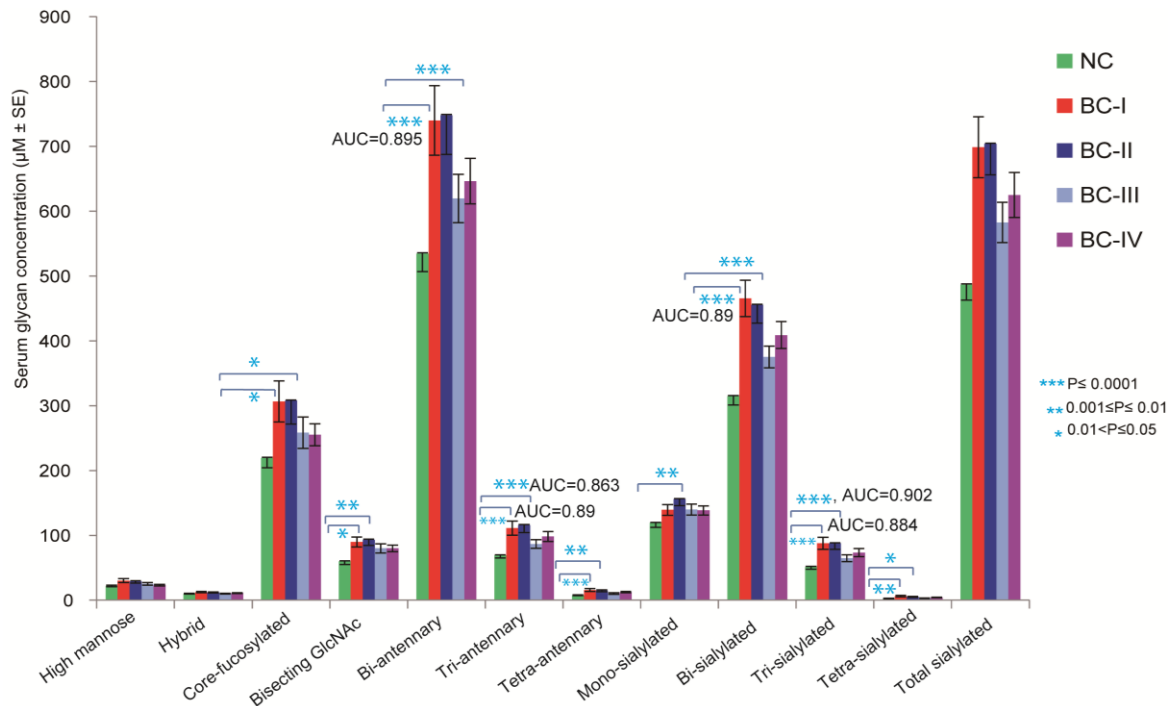
Figure 2.9. Serum abundance pattern of glycotypes sharing structural features, comparing BC patients of stage I-IV with controls. NC: normal control, BC: breast cancer, AUC: Area under the curve.

2.3.7. Some Serum *N*-glycans Associated with Age

Age associated serum glycosylation changes were observed in 10 N-glycans (Figure 2.10A) among of which glycans with m/z of 1591, 2261, and 3195 had been selected earlier as potential biomarkers for early detection of BC. In order to clear up whether aging or cancer stage had impact on elevation of these three glycans, we checked their expression pattern across all age

ranges and clinical stages. Subsequently, a significant abundance of the glycans was noticed only at age > 56, given that more than 87 % of our study subjects are under this age (Table 2.1A, Figure 2.10B) where level of these glycans differed non-significantly age wise. Moreover, individuals belonging to BC stage I and stage II had higher abundance for these glycans at lower age ranges than at older age, suggesting that cancer stage and age had different patterns with correspondent characteristic glycans. Altogether, the observation that few of the selected candidate biomarkers showed age association only in few old patients (age > 56) hardly affects their specificity to predict early stage breast cancer.
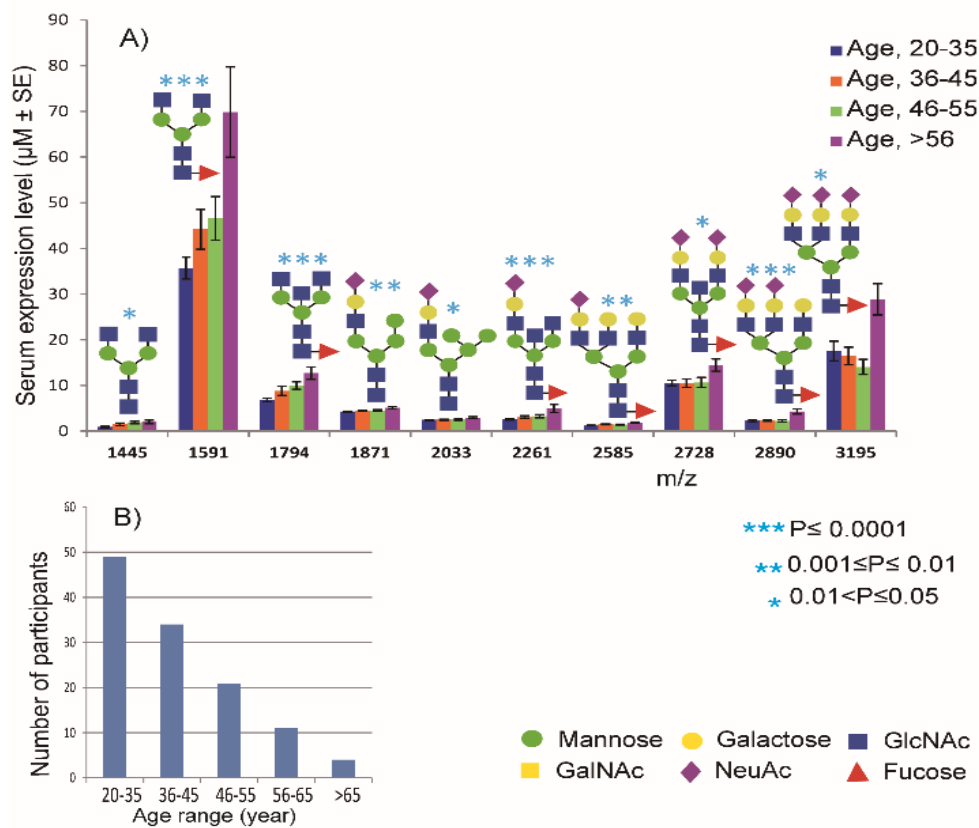
Figure 2.10. Association of serum N-glycan expression pattern with age. **A**) *N*-glycans significantly elevated at older age (> 56 years), **B**) Bar plot showing frequency distribution of the number of study subjects across different age ranges.

2.4.   Discussion

Apart from the cancer genome, deciphering posttranslational protein glycosylation alterations has been of critical importance as a promising target for discovering novel diagnostic and therapeutic agents [2]. Despite significant advances in breast cancer health care system, the current diagnostic methods give procedural discomfort and suffer from lack of reliability and specificity to predict the disease at its early stage. Hence, novel serological biomarkers and drug targets are still needed to further improve breast cancer diagnosis and treatment [29]. The present study quantitatively profiled whole serum and purified IgG *N*-glycosylation (discussed in chapter 3) alterations in Ethiopian breast cancer patients and matched controls. To the best of our knowledge, this is the first comprehensive report simultaneously addressing whole serum and purified IgG *N*-glycomics of breast cancer patients and more importantly employing study samples from native black African population whose glycosylation profile has not been studied elsewhere. We used a sensitive and an efficient glycoblotting method for glycan enrichment, after which we could identify and quantify a wide range of *N*-gycans by MALDI-TOF/MS analysis. After careful data processing and analysis steps, seventeen biosynthetically relevant complex type serum *N*-glycans (Table 2.4) could specifically distinguish breast cancer patients from normal controls. Structurally, majority of them belong to core-fucosylated, multiply branched and sialylated *N*-glycan types. With an overall quantitative up-regulation in all the breast cancer stages compared to NC, these serum glycans showed the highest abundance in the earlier stage (stage I and II) patients. ANOVA and ROC test analysis independently confirmed that these seventeen *N*-glycans had high discriminating power to differentiate cancer stages from the healthy controls. Eight of these putative glyco-biomarkers could mutually distinguish (AUC = 0.8-1) both stage I and II

patients from NC within which one glycan (m/z 2261) demonstrated high diagnostic performance in all stages of I-IV (Figure 2.7).

Quantitative analysis of glyco-subclasses sharing the same structural residue (fucose, bisecting GlcNAc, antenna, or sialic acid) further intensified the predominant branching and sialylation features associated with early stage patients compared to controls. Among such glycotypes, it was observed that total bi-antennary and bi-sialylation for stage I, total tri-antennary and tri-sialylation for stage II strongly predicted and classified them from NC (Figure 2.9). These results illustrate that aberrant serum glycan alterations in the breast cancer pathogenesis begin at an early stage of the disease and might be considered for promising diagnostic biomarker identification. The current results are different from a previous study [13] that suggested eight sialylated serum *N*-glycans (none of them had bisecting GlcNAc) as biomarkers associated with breast cancer. The authors reported that the predominant change (down-regulation of lower m/z and up-regulation of higher m/z glycans) of the suggested markers was observed in the late stage (stage IV) patients compared to controls. The candidate biomarkers identified in the present study comprise asialo, bisecting, and hperbranched/hypersialylated glycans including only three of their reported glycan markers (m/z 2744, 3560, and 3719). Notably, our candidate biomarkers showed only an up-regulation pattern that was primarily noticed at the earlier stages (stage I or II) compared with the controls. As they used C18 column for glycan purification and permethylation (classical strategy)-based MALDI-TOF/MS analysis of relative intensities, it should be clear that the methodology and patient background of the two studies are quite different that makes comparisons complicated. Thus, we hypothesize that the *N*-glycosylation changes associated with breast cancer progression in young black subjects seem to be unique and needs further attention.

Several studies have reported that hypersialylation is a crucial feature in the progression of many cancers, where by the negative charges in the terminal sialic acids of sugars interfere with epithelial cadherin-mediated cell–cell adhesion, enhancing the migratory and metastatic capacity of tumor cells [2, 30]. Furthermore, higher expressions of the glycosyltransferases responsible for core fucosylation (FUT8), branching (GnT-V), and sialylation (both α-2, 3 and α-2, 6 sialytransferases) have been associated with a greater potential for motility, invasion and metastasis in breast cancer [31-34]. The up-regulation of a bisected tetraantennary glycan (m/z 3007) in the cancer groups was however unexpected observation as the *N*-acetylglucosaminyltransferase III (GnT-III) and its bisecting GlcNAc products have been reported to inhibit further branching in the biosynthesis pathway and suppress cancer metastasis [35]. Kinetic studies on GnT-V, another subsequent key enzyme for the synthesis of branched N-glycans, indicated that the enzyme often use the bisecting GlcNAc as a substrate with a very low Vmax value and produce branched bisecting structures [36]. Taking this into account, finding such a bisected and branched structure whose abundance was strongly associated with BC stages I, II, and IV may provide implications for the drastic up-regulation of the preceding steps in the biosynthetic machinery that could unusually satisfy the low affinity of the branching enzyme. Additionally, this glycan seems to be unique to Ethiopian population as it was detected in neither of our previous glycomics results from various diseases and matched controls experimented in similar settings.

To ensure the reliability of our analytical procedure, quadruplicates of each sample were subjected to MALDI-TOF/MS analysis and the data for glycan levels were obtained from absolute quantification unlike many prior reports that had measured relative abundances. As such, the versatility of our glycoblotting method has been evidenced by its suitability for glycan purification from diverse samples (serum, cell lines, tissues, and CSF) [10-12, 20-24] using only 10 µL of sample aliquots and overnight reaction.

Previous findings from various studies highlighted that early phase aggressive type of BC is commonly noticed in the young and black population [37-40], accounting for the racial disparity in breast cancer mortality. In this context, as a first report employing native blacks Africans for protein glycosylation study, the findings of over sialylation, branching, and core-fucosylation at an earlier stage of the disease may partly reflect the racial disparity in breast cancer aggressiveness and mortality.

Altogether, our findings reveal characteristic tumor-associated glycoforms which demonstrate significant distinguishing potential not only between the whole BC patients and the healthy controls but also in a cancer stage dependent manner. The current study, however, had limitation to reveal the carrier proteins of the identified candidate biomarkers. Also, the suggested biomarkers from relatively small sample size demands further study on a larger sample size of well-matched case-control subjects before considering them in the clinical area.

2.5.   References

1. Jin X., Mu P. Targeting breast cancer metastasis. Breast Cancer (Auckl) 2015; 9:23-34. doi: 10.4137/BCBCR.S25460.

2. Pinho S.S., Reis C.A. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015; 15:540-55. doi: 10.1038/nrc3982.

3. Stowell S.R., Ju T., Cummings R.D. Protein glycosylation in cancer, Annu Rev Pathol. 2015; 10:473-510. doi: 10.1146/annurev-pathol-012414-040438.

4. Clerc F., Reiding K.R., Jansen B.C., Kammeijer G.S., Bondt A., Wuhrer M. Human plasma protein N-glycosylation, Glycoconj J. 2016; 33:309-343. doi: 10.1007/s10719-015-9626-2.

5. Anderson N.L., Anderson N.G. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteosmics 2002; 1:845-67. https://doi.org/10.1074/mcp.R200007-MCP200.

6. Christopher A.C., Stefani N.T., Lori J.S., Daniel W.C. Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics 2016; 13:1. doi: 10.1186/s12014-015-9102-9.

7. An HJ., Miyamoto S., Lancaster K.S., Kirmiz C., Li B., Lam K.S., Leiserowitz G.S., Lebrilla C.B. Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. J Proteome Res. 2006; 5:1626-1635. doi: 10.1021/pr060010k.

8. Sethi M.K., Hancock W.S., Fanayan S. Identifying N-Glycan biomarkers in colorectal cancer by mass spectrometry. Acc. Chem. Res. 2016; 49:2099-2106. doi: 10.1021/acs.accounts.6b00193.

9. Qin R., Zhao J., Qin W., Zhang Z., Zhao R., Han J., Yang Y., Li L., Wang X., Ren S., Sun Y., Gu J. Discovery of non-invasive glycan biomarkers for detection and surveillance of gastric cancer. J Cancer 2017; 8:1908-1916. doi: 10.7150/jca.17900.

10. Nouso K., Amano M., Ito Y.M., Miyahara K., Morimoto Y., Kato H., Tsutsumi K., Tomoda T., Yamamoto N., Nakamura S. Clinical utility of high-throughput glycome analysis in patients with pancreatic cancer. J Gastroenterol. 2013; 48:1171–1179. doi: 10.1007/s00535-012-0732-7.

11. Ishibashi Y., Tobisawa Y., Hatakeyama S., Ohashi T., Tanaka M., Narita S., Koie T., Habuchi T., Nishimura S., Ohyama C., Yoneyama T. Serum tri- and tetra-antennary N-glycan is a potential predictive biomarker for castration-resistant prostate cancer. Prostate 2014; 74:1521-1529. doi: 10.1002/pros.22869.

12. Kamiyama T., Yokoo H., Furukawa J., Kurogochi M., Togashi T., Miura N., Nakanishi K., Kamachi H., Kakisaka T., Tsuruga Y. Identification of novel serum biomarkers of hepatocellular carcinoma using glycomic analysis. Hepatology 2013; 57:2314-2325. doi: 10.1002/hep.26262.

13. Kyselova Z., Mechref Y., Kang P., Goetz J.A., Dobrolecki L.E., Sledge G.W., Schnaper L., Hickey R.J., Malkas L.H., Novotny M.V. Breast cancer diagnosis and prognosis through escapequantitative measurements of serum glycan profiles. Clin Chem. 2008; 54:1166-1175. doi: 10.1373/clinchem.2007.087148.

14. Haakensen V.D., Steinfeld I., Saldova R., Shehni A.A., Kifer I., Naume B., Rudd P.M., Børresen-Dale A-L., Yakhini Z. Serum N-glycan analysis in breast cancer patients – Relation to tumour biology and clinical outcome. Mol Oncol. 2016; 10:59-72. doi: 10.1016/j.molonc.2015.08.002.

15. De Leoz M.L., Young L.J., An H.J., Kronewitter S.R., Kim J., Miyamoto S., Borowsky A.D., Chew H.K., Lebrilla C.B. High-mannose glycans are elevated during breast cancer progression. Mol Cell Proteomics 2011; 10:M110.002717. doi: 10.1074/mcp.M110.002717.

16. Goetz J.A., Mechref Y., Kang P., Novotny M.V. Glycomic profiling of invasive and non-invasive breast cancer cells. Glycoconj J. 2009; 26:117–131. https://doi.org/10.1007/s10719-008-9170-4.

17. Kirmiz C., Li B., An H.J., Clowers B.H., Chew H.K., Lam K.S., Ferrige A., Alecio R., Borowsky A.D., Sulaimon S. A serum glycomics approach to breast cancer biomarkers. Mol Cell Proteomics 2007; 6:43-55, DOI: 10.1074/mcp.M600171-MCP200.

18. Lee L.Y., Thaysen-Andersen M., Baker M.S., Packer N.H., Hancock WS., Fanayan S. Comprehensive N-glycome profiling of cultured human epithelial breast cells identifies unique secretome N-glycosylation signatures enabling tumorigenic subtype classification. J. Proteome Res. 2014; 13:4783-4795. doi: 10.1021/pr500331m.

19. Nishimura S. Toward automated glycan analysis. Adv. Carbohydr. Chem. Biochem. 2011; 65:219–271. doi: 10.1016/B978-0-12-385520-6.00005-4.

20. Hatakeyama S., Amano M., Tobisawa Y., Yoneyama T., Tsuchiya N., Habuchi T., Nishimura S., Ohyama C. Serum N-glycan alteration associated with renal cell carcinoma detected by high-throughput glycan analysis. J Urol. 2014; 191:805–813. doi: 10.1016/j.juro.2013.10.052.

21. Gizaw, S.T.; Koda, T.; Amano, M.; Kamimura, K.; Ohashi, T.; Hinou, H.; Nishimura, S-I. A comprehensive glycome profiling of Huntington's disease transgenic mice. Biochim Biophys Acta. 2015; 1850:1704-1718, doi: 10.1016/j.bbagen.2015.04.006.

22. Gizaw S.T., Ohashi T., Tanaka M., Hinou H., Nishimura S. Glycoblotting method allows for rapid and efficient glycome profiling of human Alzheimer's disease brain, serum and cerebrospinal fluid towards potential biomarker discovery. Biochim Biophys Acta 2016; 1860(8):1716-1727. https://doi.org/10.1016/j.bbagen.2016.03.009 PMID: 26968461.

23. Inafuku S., Noda K., Amano M., Ohashi T., Yoshizawa C., Saito W., Murata M., Kanda A., Nishimura S, Ishida S. Alteration of N-glycan profiles in diabetic retinopathy. Invest. Ophthalmol. Vis. Sci. 2015; 56:5316–5322. doi: 10.1167/iovs.15-16747.

24. Terashima M., Amano M., Onodera T., Nishimura S., Iwasaki N. Quantitative glycomics monitoring of induced pluripotent- and embryonic stem cells during neuronal differentiation. Stem Cell Res. 2014; 13:454-464. doi: 10.1016/j.scr.2014.10.006.

25. Urita A., Matsuhashi T., Onodera T., Nakagawa H., Hato M., Amano M., Seito N., Minami A., Nishimura S., Iwasaki N. Alterations of high-mannose type N-glycosylation in human and mouse ostesarthritis cartilage. Arthritis Rheum. 2011; 63:3428–3438. doi: 10.1002/art.30584.

26. Kita Y., Miura Y., Furukawa J., Nakano M., Shinohara Y., Ohno M., Takimoto A., Nishimura S. Quantitative glycomics of human whole serum glycoprotein based on the standardized protocol for liberating N-glycans. Mol Cell Proteomics 2007; 6:1437-1445. doi: 10.1074/mcp.T600063-MCP200.

27. Furukawa J., Shinohara Y., Kuramoto H., Miura Y., Shimaoka H., Kurogochi M., Nakano M., Nishimura S. Comprehensive approach to structural and functional glycomics based on chemoselective glycoblotting and sequential tag conversion. Anal Chem. 2008; 80:1094-1101. doi: 10.1021/ac702124d.

28. Nishimura S., Niikura K., Kurogochi M., Matsushita T., Fumoto M, Hinou H., et al. High-throughput protein glycomics: combined use of chemoselective glycoblotting and MALDI-TOF/TOF mass spectroscopy. Angew Chem Int Ed Engl. 2004; 44(1):91-96. https://doi.org/10.1002/anie.200461685

29. Misek D.E., Kim E.H. Protein biomarkers for the early detection of breast cancer. Int J Proteomics 2011; 2011:343582. http://dx.doi.org/10.1155/2011/343582.

30. Pearce O.M., Läubli H. Sialic acids in cancer biology and immunity. Glycobiology 2016; 26:111-128. doi: 10.1093/glycob/cwv097.

31. Tu C.F., Wu M.Y., Lin Y.C., Kannagi R., Yang R.B. FUT8 promotes breast cancer cell invasiveness by remodeling TGF-β receptor core fucosylation. Breast Cancer Res. 2017; 19:111. doi: 10.1186/s13058-017-0904-8.

32. Handerson T., Camp R., Harigopa, M., Rimm D., Pawelek J. Beta1,6-branched oligosaccharides are increased in lymph node metastases and predict poor outcome in breast carcinoma. Clin. Cancer Res. 2005; 11:2969-2973. doi: https://doi.org/10.1158/1078-0432.CCR-04-2211.

33. Recchi M.A., Hebbar M., Hornez L., Harduin-Lepers A., Peyrat J.P., Delannoy P. Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer. Cancer Res. 1998; 58:4066-4070, http://cancerres.aacrjournals.org/content/58/18/4066.

34. Cui H., Lin Y., Yue L., Zhao X., Liu J. Differential expression of the alpha 2,3-sialic acid residues in breast cancer is associated with metastatic potential. Oncol. Rep. 2011; 25:1365-1371. doi: 10.3892/or.2011.1192.

35. Zhao Y., Nakagawa T., Itoh S., Inamori K., Isaji T., Kariya Y., Kondo A., Miyoshi E., Miyazaki K., Kawasaki N., et al. N-Acetylglucosaminyltransferase III antagonizes the effect of N-acetylglucosaminyltransferase V on α3β1 integrin-mediated cell migration. J Biol Chem. 2006; 281:32122-32130. doi: 10.1074/jbc.M607274200.

36. Sasai K., Ikeda Y., Fujii T., Tsuda T., Taniguchi N. UDP-GlcNAc concentration is an important factor in the biosynthesis of beta1,6-branched oligosaccharides: regulation based on the kinetic properties of N-acetylglucosaminyltransferase V. Glycobiology 2002; 12:119-127. https://doi.org/10.1093/glycob/12.2.119.

37. Richardson L.C., Henley S.J., Miller J.W., Massetti G., Thomas C.C. Patterns and trends in black-white differences in breast cancer incidence and mortality – United States, 1999-2013, MMWR Morb Mortal Wkly Rep. 2016; 65:1093–1098. doi: 10.15585/mmwr.mm6540a1.

38. Swanson G.M., Haslam S.Z., Azzouz F. Breast cancer among young African-American women. Cancer 2002; 97:273-279, https://doi.org/10.1002/cncr.11025.

39. Assi H.A., Khoury K.E., Dbouk H., Khalil L.E., Mouhieddine T.H., El Saghir N.S. Epidemiology and prognosis of breast cancer in young women, J Thorac Dis. 2013; 5: S2-S8. DOI: 10.3978/j.issn.2072-1439.2013.05.240.

40. Morris G.J., Mitchell E.P. Higher incidence of aggressive breast cancers in African-American women: A review, J Natl Med Assoc. 2008; 100:698-702. https://doi.org/10.1016/S0027-9684(15)31344-4.

Chapter 3

IgG-Focused *N*-Glycomics for Breast

Cancer Biomarker Discovery

3.1. Introduction

Immune response alterations involving chronic activation of the humoral immunity have been implicated in the pathogenesis of breast cancer [1]. Immunoglobulins (Ig) are group of glycoproteins that have antibody activity (including IgA, IgD, IgE, IgG, IgM), accounting for about 20% of plasma proteins. It is known that among all, IgG is the most abundant (> 80%) Ig and represents the central role in the humoral immune response. Each of IgG heavy chains contains a single bi-antennary *N*-glycan covalently attached to asparagine 297 of their fragment crystallizable (Fc) region in which minor changes in this glycan composition or abundance can significantly alter the conformation of the Fc region, altering the interaction with receptor proteins and thus modulating its effector function [2, 3]. Lack of core-fucose in the Fc glycan has been demonstrated to enhance the ability of IgG to induce antibody-dependent cell-mediated cytotoxicity (ADCC), whereas increased galactosylation, and sialylation have been reported to be associated with anti-inflammatory effect of IgG [4, 5], Figure 3.1. On the other hand, a recently reported finding show that Fc-galactosylation specifically increases the antibodies' affinity for C1q binding and their efficacy to induce complement activation and complement-dependent cytotoxicity (CDC) without changing its affinity with Fcγ receptors (and thus ADCC) [6]. Addition of terminal sialic acid to IgG *N*-glycan converts its function from pro-to anti-inflammatory by decreased binding to Fcγ receptors [7].
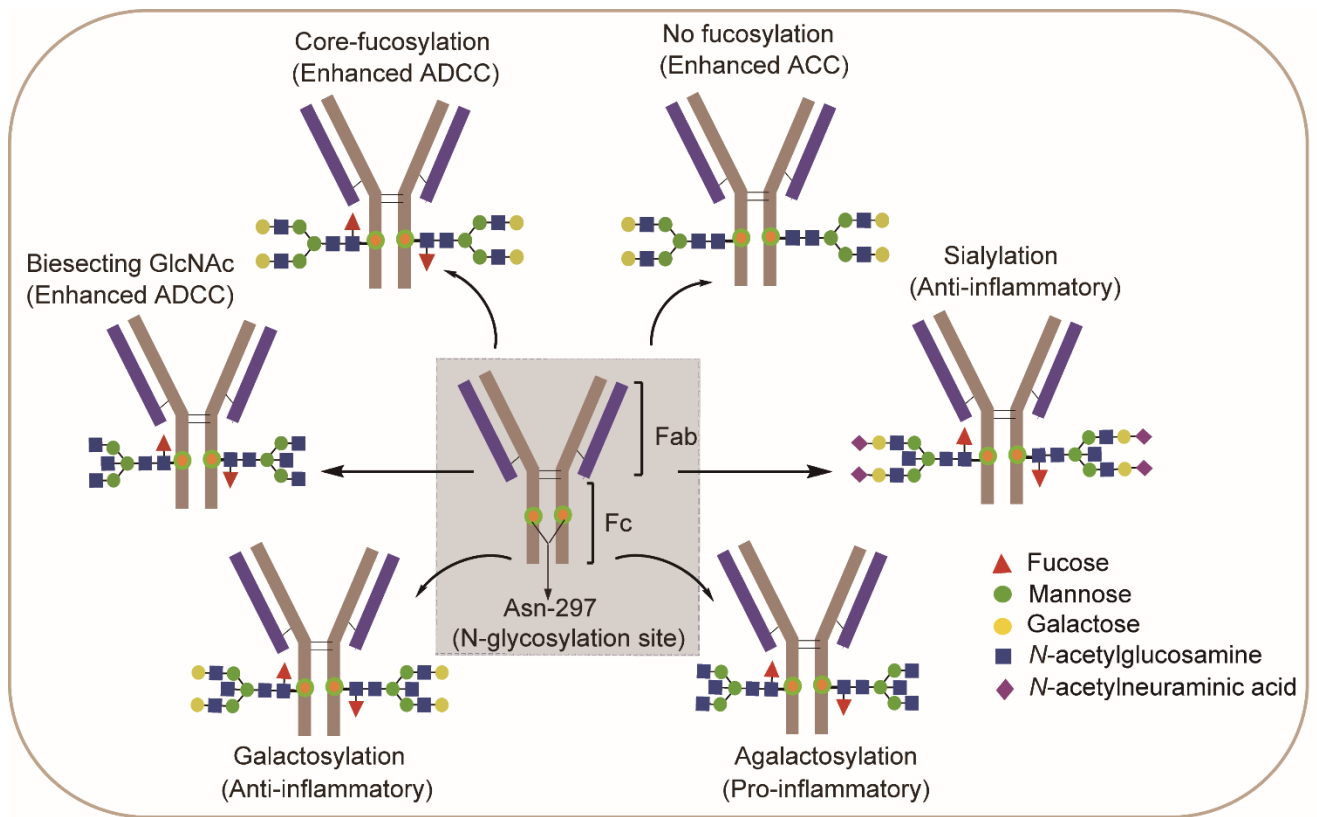
Figure 3.1. IgG glycosylation alterations: multiple functional implications

The large abundance of IgG and its role as a representative protein of the immune system makes it an ideal protein to investigate the role of protein-specific glycosylation in cancer. Altered glycosylation patterns of IgG Fc N-glycan have been observed in colorectal, gastric, urothelial, ovarian, and breast cancers [8-12]. Having observed informative patient-associated glycosylation alterations in the total serum *N*-glycome profile, we have extended our effort to IgG-focused *N*-glycomics of similar study subjects for deeper insights from the immune system perspective.

3.2. Immunoglobulin G Purification from Whole Serum

IgG was purified from serum by affinity chromatography using 96-well protein G spin plate (Thermo Scientific) that is applied by fixing on another wash/collection plate as per the manufacturer's protocol (Figure 3.2). First, 50 µL of serum was diluted 4 times with binding buffer (PBS, pH 7.2) and applied to the wells. The plate assembly was placed on a plate shaker and incubated for 30 minutes with moderate agitation, followed by centrifugation at $1,000 \times g$ for 1 minute with discarding the flow-through. The resin on the purification plate was washed 4 times by adding 300 µL of PBS to each well and then centrifuging at $1,000 \times g$ for 1 minute, discarding the flow-through each time. Before elution, 20 µL of neutralization buffer (1 M Tris-HCl, pH 9) was added to each well of new collection plate to maintain IgG stability. Subsequently, IgG was eluted by adding 200 µL of 0.1 M formic acid, pH 2.7 and centrifuging at $1,000 \times g$ for 1 minute. This step was repeated twice for complete elution. After confirming the purity of the IgG fraction by SDS-PAGE, it was dried using speed vac and then reconstituted in 50 µL of pure water to be applied for glycan release and purification by glycoblotting and then MALDI-TOF/MS analysis (similar protocol as described in chapter 2, section 2.2.3).
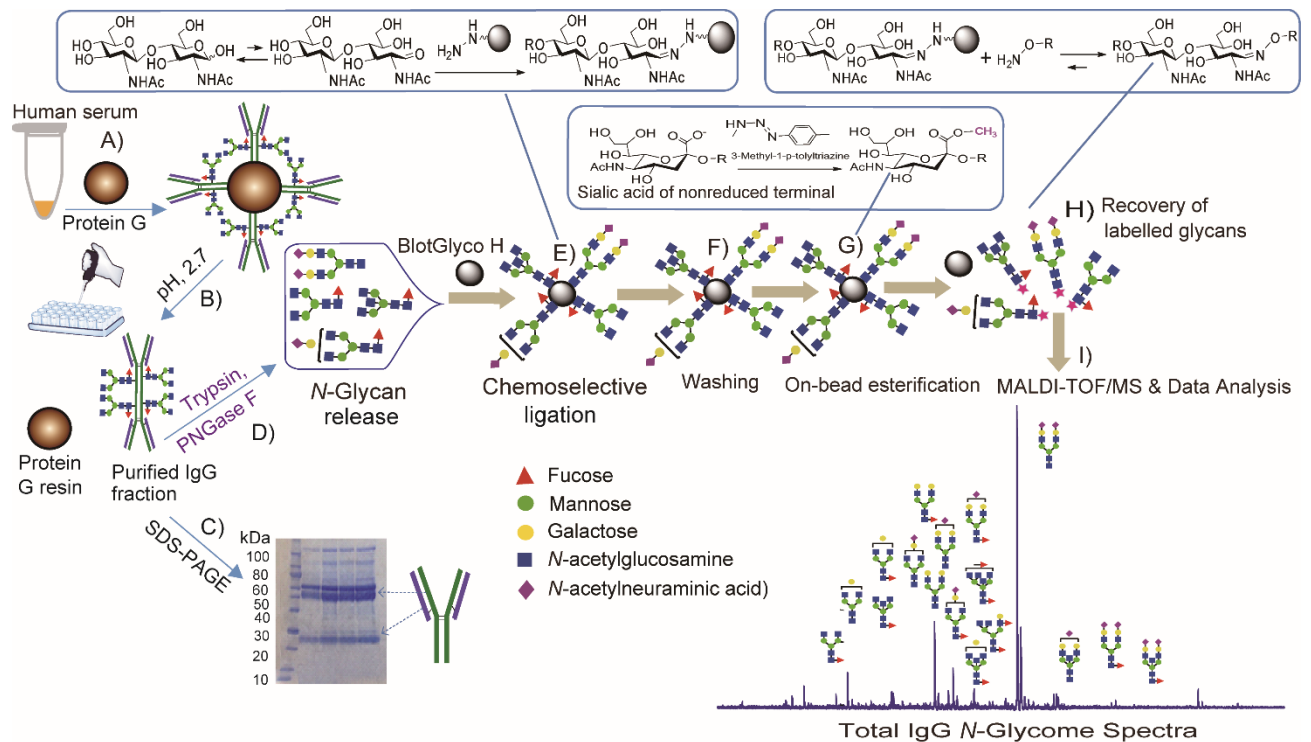
Figure 3.2. Schematic workflow of glycoblotting-based MALDI-TOF/MS analysis of *N*-glycomes derived from purified IgG fraction. In summary, the major steps include: **A**) Purification of IgG fraction from whole serum using protein G resin, **B**) IgG separation from bead using elution buffer, **C**) Confirmation of IgG purity by SDS-PAGE, **D**) Enzymatic release of glycans from IgG, **E**) Chemoselective capturing of reducing sugars onto a hydrazide-functionalized BlotGlycoH beads, **F**) Washing to remove all impurities, **G**) On-bead methyl esterification of sialic acid residues, **H**) Recovery of BOA-labeled glycans, **I**) MALD-TOF/MS and data analysis.

3.3.   Immunoglobulin G-linked *N*-glycans and their Diagnostic Potential

To know if the aberrant glycosylation pattern of total serum *N*-glycome was associated with the immune system, we focused on in-depth IgG fraction *N*-glycome profiling of the entire study subjects by applying the glycoblotting-assisted *N*-glycan capturing coupled with MALDI-TOF/MS based quantitative analysis. About one-third of the total glycans that had been detected in the whole serum were appeared in the IgG fraction of the samples as well (indicated with "+" sign in Table 2.2). The collective mass spectra for IgG *N*-glycome of normal controls and breast cancer patients is shown in Figure 3.3, from which recognizable variations in terms of peak intensity and detection patterns between BC patients and controls can be seen in the individual mass spectrum (Figure 3.4).
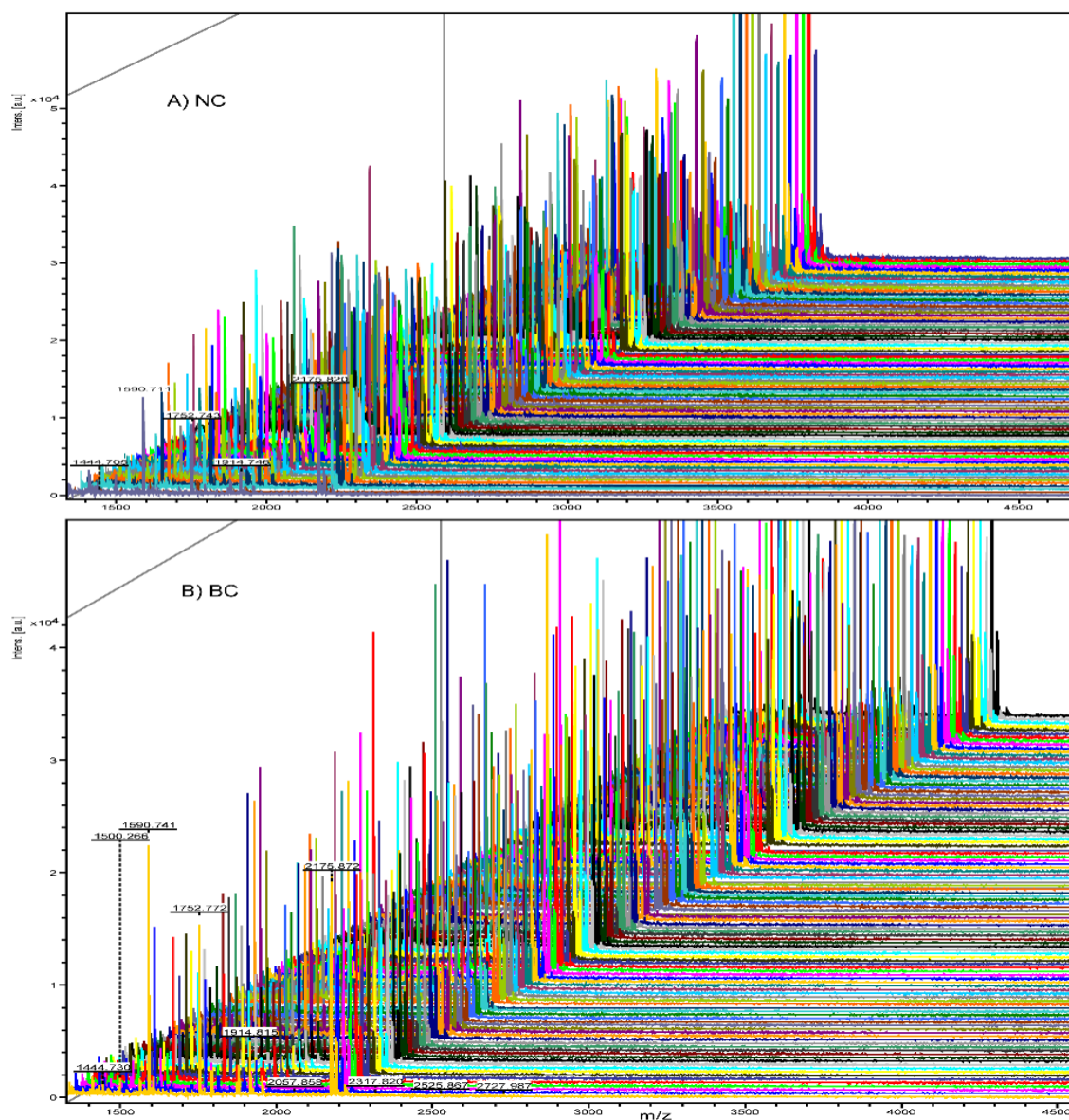
Figure 3.3. Large-scale MALDI-TOF Mass Spectra of IgG N-glycans of normal controls (**A**) and BC patients (**B**).
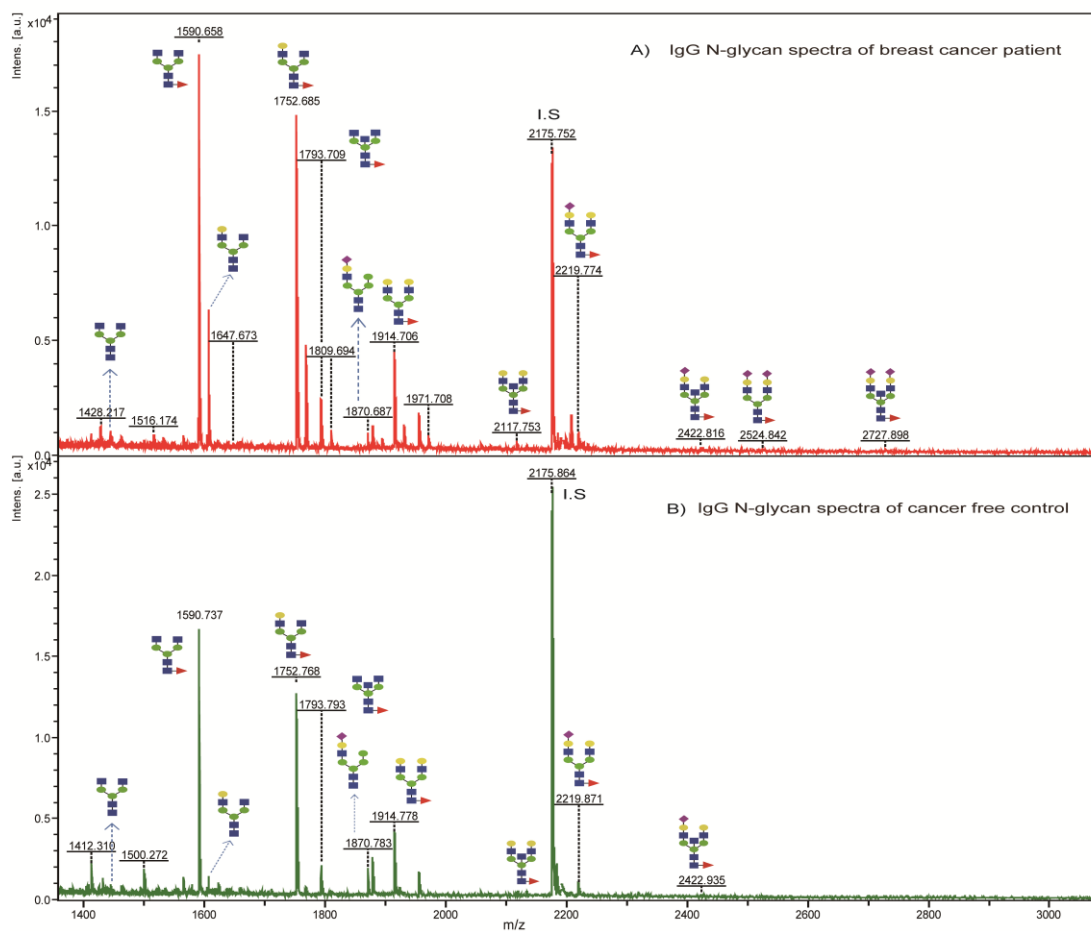
Figure 3.4. MALDI-TOF/MS spectra of IgG *N*-glycans acquired from individuals with BC (**A**) and free of BC (**B**).

Quantitative comparison result illustrated statistically significant elevation of 5 glycans (m/z 1445, 1591, 1753, 1794, and 2423) in the whole patient group (Figure 3.5A), of which 3 glycans (m/z 1591, 1753, 1794) had substantial increase in the stage II patients comparing with the controls (Figure 3.5B). Overall, 12 of the 15 detected IgG glycans belong to complex biantennary type within which core-fucosylation was noticed as a structural feature in 9 of them, including the 3 abundantly expressed in patients. Moreover, in the same way as in serum, only two IgG glycans (m/z 1915 and 2220) showed an overall slight reduction in the patient group.
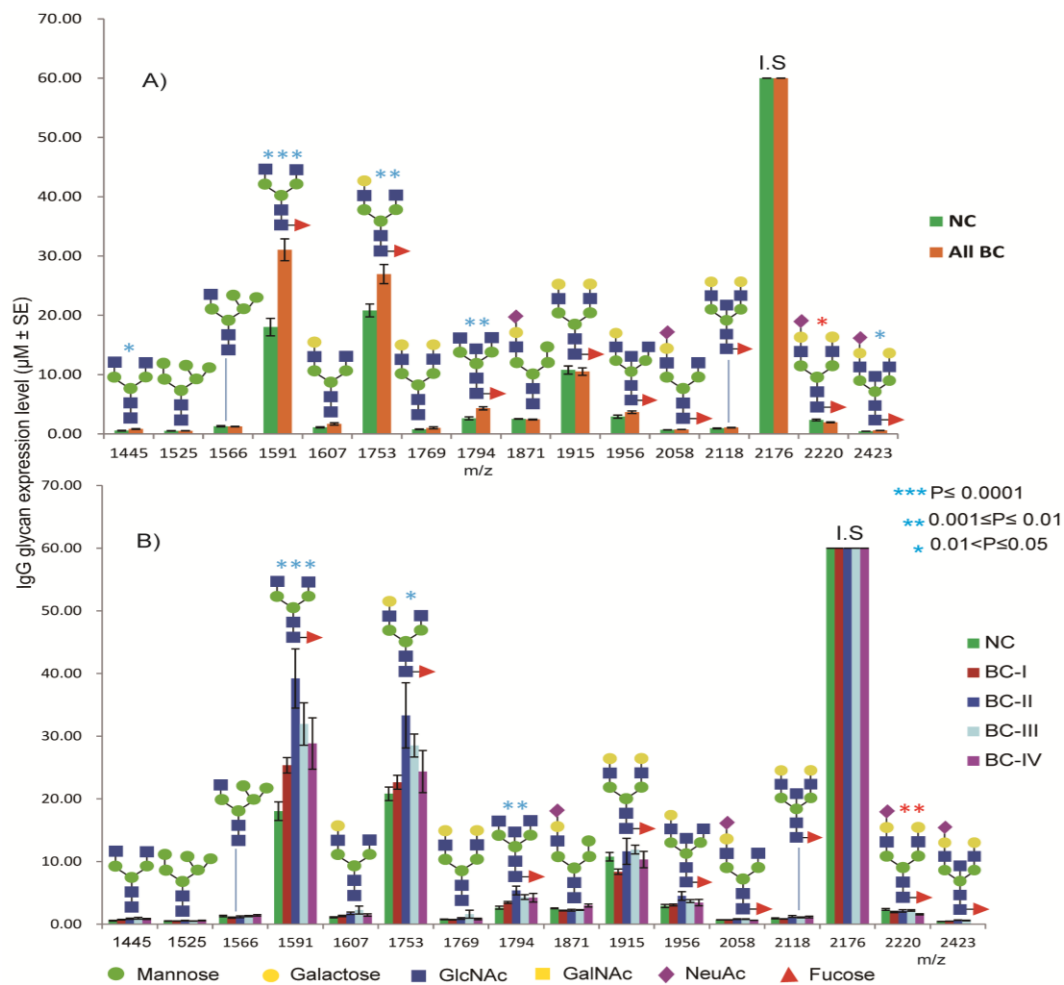
Figure 3.5. Expression levels and structures of N-glycans analyzed from purified IgG fraction. **A)** All breast cancer patients vs normal controls, **B)** Breast cancers stages I-IV vs normal controls. m/z 2176 is an internal standard (I.S), blue and red asterisk = significant increase and decrease, respectively in the glycan level of BC group comparing with NC.

ROC test analysis indicated that, two IgG glycoforms (m/z 1591 and 1794) could distinguish all BC patients and more strongly stages-II patients from the controls (AUC = 0.944 and 0.921, respectively; Figure 3.6A, B). It should be noted that these glycans have shown quite similar pattern of abundance across the IgG and serum of all cancer stages and controls (Figure 3.6B, C), reflecting that their carrier glycoprotein in the patients' serum is mainly IgG.
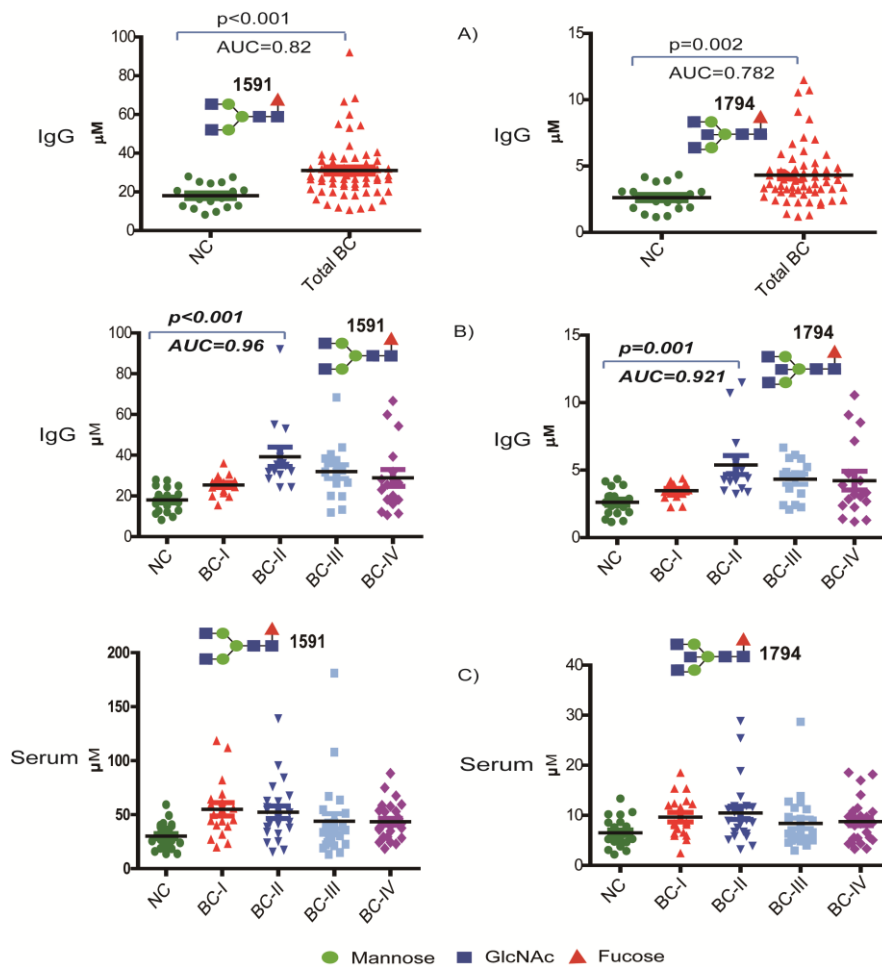
Figure 3.6. Dot plot illustrating two *N*-glycans highly expressed in IgG of BC patients. **A)** Their distinguishing potential between whole BC patient group and NC, **B)** Their high discrimination potential particularly between breast cancer stage II (BC-II) patients and NC. **C)** Their similar serum expression pattern (with that of IgG shown in **B**) across the four cancer stages (I-IV).

To determine IgG as a quantitatively reliable carrier of certain circulatory *N*-glycans, ratio of glycan levels quantified from IgG fraction to that of whole serum were calculated and summarized in Table 3.1. Thus, among the glycans whose larger proportion originated from

IgG in BC patients include the glycans with m/z 1445 (53.25%), 1591 (64.76%), 1607 (41.72%), 1753 (59.41%), 1794 (46.64%), 1871 (53.24%), and 1915 (53.44%).

Table 3.1. Percentage of IgG *N*-glycan level relative to their concentration in serum.

m/z	Glycan	Glycan concentration (µM)				% Contributed by IgG	
		Serum (NC)	Serum (BC)	IgG (NC)	IgG (BC)	% (NC)	% (BC)
1445		0.79	1.55	0.56	0.83	**70.87**	**53.25**
1525		6.88	8.17	0.52	0.54	7.57	6.56
1566		0.88	1.03	1.29	1.25	**146.04**	**121.10**
1591		30.07	47.94	18.03	31.05	**59.94**	**64.76**
1607		3.41	4.05	1.10	1.69	**32.12**	**41.72**
1753		42.26	45.38	20.82	26.96	**49.26**	**59.41**
1769		5.19	5.76	0.77	1.02	14.89	17.76
1794		6.53	9.26	2.62	4.32	**40.19**	**46.64**
1871		4.20	4.57	2.53	2.43	**60.21**	**53.24**
1915		25.54	19.68	10.79	10.52	**42.23**	**53.44**
1956		9.13	10.36	2.92	3.66	**31.94**	**35.29**
2058		3.02	3.29	0.69	0.75	22.71	22.85
2118		7.21	10.92	0.94	1.06	13.07	9.69
2220		28.54	28.52	2.34	1.96	8.21	6.86
2423		14.09	20.63	0.44	0.57	3.13	2.77

NC = normal control group, BC = whole breast cancer group. Indicated in bold are those *N*-glycans whose quantity in IgG showed considerable contribution to their expression level in serum.

3.4. Discussion

It is unclear and hardly reported that which kinds of carrier proteins are involved in breast cancer associated serum *N*-glycan alterations. In a more detailed analysis, we purified the IgG fraction from the whole serum of all study participants and then integrated (Figure 3.2) to our high-throughput glycoblotting method for subsequent glycan release and purification. MALDI-TOF/MS based quantitative analysis confirmed that many of the biantennary glycans including some of the suggested serum biomarkers (m/z 1591 and 2118) for early stage BC were originated from serum IgG (Table 3.1). Two core-fucosylated and agalactosylated IgG glycans (m/z 1591 and 1794) were significantly up-regulated in the breast cancer patients, more specifically distinguishing patients with stage II breast cancer from normal controls. Their expression patterns in IgG and serum of controls and stages I-IV patients were found to be consistent (Figure 3.6A-C); evidencing that IgG is a reliable carrier for a larger proportion of their concentration in serum. Being one of the major *N*-glycosylated serum proteins and an essential part of the humoral immune system, IgG structure and function is importantly modulated by the abundance and types of glycoforms attached to its fragment crystallizable (Fc) region [3]. Absences of core fucose in the IgG Fc domain has been shown to enhance its antibody-dependent cell-mediated cytotoxicity (ADCC) function by increasing its affinity towards cellular Fc receptors, a finding increasingly being used to improve the efficacy of afucosylated therapeutic antibodies [4, 13, 14]. Furthermore, increased IgG Fc galactose has recently been reported to enhance its complement-dependent cytotoxicity (CDC) activity via C1q binding [6, 14].

Based on our IgG glycomics result and the reported observations, our findings of increased fucosylated and agalactosylated glycan structures observed in the cancer group might be a

synergistic mechanism allowing the tumor cells to escape from the immune system. In line with the present results, increased IgG fucosylation (of non-sialylated glycans) and agalactosylation features have been previously reported in colorectal and gastric cancers [8, 9]. In contrast, a recent study by Kawaguchi-Sakita N, et al., [12] has reported increased galactosylated IgG *N*-glycans associated with Japanese breast cancer patients. Accordingly, seven galactosylated biantennary *N*-glycans when merged into a group showed significant abundance and predicted the probability of BC with a moderate accuracy whereas, unlike in the present study, no single IgG glycan could strongly distinguish patients of BC from NC. Such dissimilarities can be caused not only by differences in analytical procedures but also differences in the patient background such as race, life style, age (as the majority of their study subjects were already in the menopausal state; over 50s y/o), all of which have been reported to affect the gylcosylation pathway [15].

Apart from nonnegligible distinguishing features associated with BC, the IgG *N*-glycomics results clearly suggest that other unknown carrier proteins are greatly involved in the serum *N*-glycosylation alteration during breast cancer progression. In particular, serum proteins that carry the hyperbranched and hypersialylated candidate glyco-biomarkers of BC are yet to be identified. Alpha-1-acid glycoprotein (AGP), an acute phase protein synthesized in the liver and secreted into the circulation, is believed to be one potential carrier for the bi-, tri-, or tetra-antennary complex type glycoforms. However, little is known about the association between breast cancer progression and AGP glycosylation pattern [16].

In conclusion, the author comprehensively evaluated the whole serum and IgG *N*-glycosylation signatures of breast cancer patients and identified many novel *N*-glycan

candidate biomarkers showing strong potential for distinguishing early stage BC from NC. Apart from the whole serum, IgG focused *N*-glycome profiling provided evidences on BC-associated glycosylation alterations from the immune system perspective. Further study on direct identification of the intact glycoproteins using a glycoprteomics approach is expected to provide deeper understanding of specific biomarkers towards their clinical utility.

3.5. References

1. DeNardo D.G., Coussens L.M. Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. Breast Cancer Res. 2007; 9(4):212. doi: [10.1186/bcr1746] PMCID: PMC2206719.

2. Raju T.S. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. Curr Opin Immunol. 2008; 20(4):471-478. https://doi.org/10.1016/j.coi.2008.06.007.

3. Arnold J.N., Wormald M.R., Sim R.B., Rudd P.M., Dwek R.A. The impact of glycosylation on the biological function and structure of human immunoglobulins. Annu. Rev. Immunol. 2007; 25:21-50. doi: 10.1146/annurev.immunol.25.022106.141702.

4. Lauc G., Pezer M., Rudan I., Campbell H. Mechanisms of disease: The human N-glycome. Biochim Biophys Acta 2016; 1860(8):1574-1582. https://doi.org/10.1016/j.bbagen.2015.10.016.

5. Plomp R., Ruhaak L.R., Uh H.W., Reiding K.R., Selman M., Houwing-Duistermaat J., et al. Subclass specific IgG glycosylation is associated with markers of inflammation and metabolic health. Sci Rep. 2017; 7(1): 12325. doi: [10.1038/s41598-017-12495-0].

6. Peschke B., Keller C.W., Weber P., Quast I., Lünemann J.D. Fc-Galactosylation of human immunoglobulin gamma isotypes improves C1q binding and enhances complement-dependent cytotoxicity. Front Immunol. 2017; 8:646. doi: 10.3389/fimmu.2017.00646.

7. Kaneko Y., Nimmerjahn F., Ravetch J.V. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. Science 2006; 313(5787):670-673. doi: 10.1126/science.1129594.

8. Vučković F., Theodoratou E., Thaçi K., Timofeeva M., Vojta A., Štambuk J., Pučić-Baković M., Rudd P.M., Đerek L., Servis D., et al. IgG glycome in colorectal cancer. Clin Cancer Res. 2016; 22:3078-3086. doi: 10.1158/1078-0432.CCR-15-1867.

9. Kodar K., Stadlmann J., Klaamas K., Sergeyev B., Kurtenkov O. Immunoglobulin G Fc N-glycan profiling in patients with gastric cancer by LC-ESI-MS: relation to tumor progression and survival. Glycoconj. J. 2012; 29:57-66. doi: 10.1007/s10719-011-9364-z.

10. Tanaka T., Yoneyama T., Noro D., Imanishi K., Kojima Y., Hatakeyama S., Tobisawa Y., Mori K., Yamamoto H., Imai A.; et al. Aberrant N-glycosylation profile of serum immunoglobulins is a diagnostic biomarker of urothelial carcinomas. Int. J. Mol. Sci. 2017; 18:2632. https://doi.org/10.3390/ijms18122632.

11. Qian Y., Wang Y., Zhang X., Zhou L., Zhang Z., Xu J., Ruan Y., Ren S., Xu C., Gu J. Quantitative analysis of serum IgG galactosylation assists differential diagnosis of ovarian cancer. J Proteome Res. 2013; 12(9):4046-4055. doi: 10.1021/pr4003992.

12. Kawaguchi-Sakita N., Kaneshiro-Nakagawa K., Kawashima M., Sugimoto M., Tokiwa M., Suzuki E., Kajihara S., et al. Serum immunoglobulin G Fc region N-glycosylation profiling by matrix-assisted laser desorption/ionization mass spectrometry can distinguish breast cancer patients from cancer-free controls. Biochem Biophys Res Commun. 2016; 469:1140-1145. doi: 10.1016/j.bbrc.2015.12.114.

13. Li T., DiLillo D.J., Bournazos S., Giddens J.P., Ravetch J.V., Wang L-X. Modulating IgG effector function by Fc glycan engineering. Proc Natl Acad Sci USA. 2017; 114:3485-3490, doi: 10.1073/pnas.1702173114.

14. Li W., Zhu Z., Chen W., Feng Y., Dimitrov D.S. Crystallizable fragment gycoengineering for therapeutic antibodies development. Front Immunol. 2017; 8:1554. doi: 10.3389/fimmu.2017.01554.

15. Pucic M., Knezevic A., Vidic J., Adamczyk B, Novokmet M., Polasek O., Gornic O., Supraha-Goreta S., Wormald M.R., Redzic I. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human

populations. Mol Cell Proteomics 2011; 10(10):M111.010090. doi: 10.1074/mcp.M111.010090.

16. Smith K.D., Behan J., Matthews-Smith G., Magliocco A.M. Alpha-1-acid glycoprotein (AGP) as a potential biomarker for breast cancer. InTech. 2012; 9:201-222. http://dx.doi.org/10.5772/48177.

Chapter 4

Influence of Ethnic Variation on Human Serum *N*-glycome and the Identified Cancer-Relevant Glycan Biomarkers

4.1. Introduction

Many of the human serum proteins including alpha-1-acid glycoprotein, alpha-1-antitrypsin, alpha-2-macroglobulin, antithrombin-III, apolipoproteins, ceruloplasmin, fibrinogen, immunoglobulins, haptoglobin, hemopexin, and serotransferrin, are heavily glycosylated, making them targets for glyco-biomarker discovery and therapeutic opportunities [1]. In contrast to nucleic acids and proteins, biosynthesis of glycans is not template-driven but, rather, is a result of a complex network of metabolic and enzymatic reactions. Because of this and subsequent methodological difficulties, the field of glycomics has been lagging behind genomics and proteomics [2, 3]. Recent analytical advancements have transformed the glycomics field [4-6] and thus enabled discovery of glycan-based serum biomarkers for various cancers. However, bringing these glycan or glycoprotein markers to clinical practice has been hindered as their potential to distinguish cases from controls or disease stages seems to be inadequate and varies from country to country population wise, complicating their validity and clinical utility [2, 7-9]. For example, the FDA approved cancer biomarkers based on serum level of *O*-glycosylated mucin glycoproteins of carbohydrate antigens (CA125 for ovarian, CA27.29 and CA15-3 for breast, CA19-9 for pancreatic cancers) and *N*-glycosylated glycoproteins (α-fetoprotein for hepatocellular carcinoma, prostate-specific antigen for prostate cancer) lack the specificity and sensitivity to be used for early detection of cancer [9-10].

Previous population-based studies have reported the association of plasma *N*-glycan structure alterations (mostly increased plasma *N*-glycan complexity) with metabolic syndrome and higher risk of type 2 diabetes [11-12]. In several glycomics studies profiling immunoglobulin G (IgG) focused *N*-glycome among various ethnic populations, it was

emphasized that changes in IgG-linked glycan composition and abundance were correlated with hypertension [13-15], cardiovascular disease [16], blood lipids and dyslipidaemia [17]. Apart from several pathological studies that have profiled glycosylation pattern, comparative glycosylation studies among healthy subjects with various characteristics are often overlooked. Nevertheless, such studies on healthy people would be of clinical benefit not only by providing anticipatory insights but also by elucidating confounding factors that could lead to controversies on the identified biomarker [18]. In this context, a comprehensive study on human plasma *N*-glycan profile provided an evidence for the variability of some glycan levels with aging, life style and environmental factors [19]. Similarly, Ding N. and his group demonstrated that a healthy human serum *N*-glycan profile had shown considerable variations in age and sex dependent manner [20]. Furthermore, IgG focused *N*-glycomic study in a Han Chinese population revealed that changes in IgG *N*-glycan features significantly correlate with age [21]. To our knowledge, little is known about the association between ethnicity and healthy human serum *N*-glycome profile. Thus, this study was aimed to address the inter-ethnic differences in serum *N*-glycome among US origin control, South Indian, Japanese, and Ethiopian ethnic populations using a rapid glycoblotting-assisted MALDI-TOF/MS-based quantitative analysis. The present study clearly demonstrated that various high mannose, core-fucosylated, multiply branched and sialylated glycoforms illustrated an ethnic-specific expression pattern and marked alterations in their serum abundance.

4.2.   Materials and Methods

4.2.1.   Human Serum Samples

The study was performed in accordance with the ethical guidelines and protocols of the Declaration of Helsinki upon approval by the ethical review boards of Hokkaido University, Faculty of Advanced Life Sciences, Japan and Addis Ababa University, School of Medicine, Ethiopia. Informed consent was obtained from all volunteer participants. As summarized in Table 4.1, total of 54 healthy subjects having a relatively narrow age range and various ethnic composition were involved in the study. Serum from male Japanese (n = 10, age = 20s-30s y/o), and male South Indians (n = 10, age = 32.5 ± 5 y/o) were collected as part of Asia-Africa Science Platforms project. Similarly, serum samples from 24 female Ethiopians (31.54 ± 7 y/o) were collected in Black Lion Specialized Teaching Hospital of Addis Ababa University and carefully transported to Japan after freeze dried. US origin healthy control samples were from serum pool of different male donors whose age information was not provided as these samples were purchased from Sigma-Aldrich company (product # H4522). Serum samples from male Japanese hepatocellular carcinoma (HCC) patients (n = 11, age = 50s-60s y/o) were also included in the study. As inclusion criteria for the healthy ethnic groups, fully heathy (mentally and physically) adults (> 18 y/o), non-smokers, and non-obese individuals were included in the present study. Whereas, individuals who were pregnant, were receiving any treatment, or were with a medical history of disease such as cancer, diabetes, neurodegenerative, liver, or cardiovascular diseases which can affect the glycan profile, were excluded from the study. All serum samples were kept at -80°C until used in the subsequent experiment.

Table 4.1. Demographic characteristics of study participants.

Category	Number (n)	Gender	Age (years) Mean ± SD	Range
US control	10	M	---	---
Japanese	10	M		20s-30s
Indian	10	M	32.5 ± 5.039	27-43
Ethiopian	24	F	31.54 ± 7.175	23-50
HCC	11	M	---	50s-60s
Total	65			

While recruiting the Japanese subjects, age of each subject was not recorded in its exact value, rather as 20s (within 20-29.9 y/o), 30s (within 30-39.9 y/o), 50s (within 50-59.9 y/o), 60s (within 60-69.9 y/o). The US origin control serum is a pool collected from several male donors whose age information was not provided as it was purchased sample from Sigma-Aldrich company.

4.2.2. Reagents and Equipment

Ammonium bicarbonate 99% (ABC), 1-propanesulfonic acid, 2-hydroxyl-3-myristamido (PHM), 3-methyl-1-p-tolyltriazene (MTT), O-benzylhydroxylamine hydrochloride (BOA), and disialyloctasaccharide were purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan). Peptide N-glycosidase F (PNGase F) was purchased from New England Biolabs Inc. (Ipswich, Massachusetts, USA), whereas BlotGlyco H beads were purchased from Sumitomo Bakelite, Co. Ltd. (Tokyo, Japan). Dithiothreitol (DTT), iodoacetamide (IAA), trypsin, α-cyano-4-hydroxycin-namic acid diethylamine salt, 2-mercaptoethanol, and 5-N-Glycolylneuraminic acid (Neu5Gc) were from Sigma-Aldrich, Inc. (St. Louis, MO, USA). 5-N-acetylneuraminic acid (Neu5Ac) was purchased from Japan Food & Liquor Alliance Inc., Food & Bio Research Center, Inc. (Kyoto, Japan). 1,2-Diamino-4,5-

methylenedioxybenzene (DMB) was from Dojindo Laboratories (Kumamoto, Japan). Other reagents and solvents were obtained from Wako Pure Chemicals Co., (Tokyo, Japan), unless otherwise stated. SweetBlot[TM] (automated glycan processing and incubating machine) was from System Instruments Co., Ltd., (Hachioji, Japan). MultiScreen Slvinert[R] filter plates were purchased from Millipore Co., Inc. (Tokyo, Japan). All mass quantifications were done using MALDI-TOF/MS (Ultraflex III, Brukers Daltonics, Germany). HPLC analysis was performed on Hitachi D-7000, Hitachi High Technologies Co., Ltd. (Tokyo, Japan).

4.2.3. Release of Total *N*-glycans from Human Serum Glycoproteins

A method based on high-throughput glyco-technology previously developed and optimized in our laboratory [4-6, 22-25] was used for glycan release, purification, labeling and spotting. Initially, 10 µL of each serum sample was transferred to a 96 well polymerase chain reaction (PCR) plate and dissolved with 30 µL of freshly prepared 0.33M ammonium bicarbonate (ABC) containing 0.1% of PHM in 10mM ABC and the solution was incubated at 37°C for 10 min., serially diluted human serum standards were prepared by mixing about 50 µL serum aliquot of each healthy participant to give pooled sera, from which the concentration of standard serial dilutions (0.5×, 0.75×, 1.0×, 1.25×, 1.5×, 1.75×, 2.0×, and 2.25×) were adjusted using Milli-Q water and then included in the experiment as the resulting calibration curve helps to evaluate the linearity and reproducibility of the detect peaks. As an internal standard, 12 µL of 60 µM disialyloctasaccharide was also added and mixed in each well to aid eventual quantification of detection N-glycans. Solubilized proteins were reduced by 10 µL of 120 mM 1, 4-dithiothreitol (DTT) at 60°C for 30 min followed by alkylation with 20 µL of 123 mM iodoacetamide by incubation in dark for 1 hr. The mixture was then treated with 10 µL of 40 U/µL trypsin in 1 mM HCl at 37°C for 2 hrs. After heat-inactivation of the enzyme at 90 °C

for 10 minutes and then cooling to room temperature, N-glycans were released from trypsin-digested samples by incubation with 2 U of Peptide N-glycosidase F (PNGase F) at 37°C for 6 hrs.

4.2.4. Selective *N*-glycan Enrichment by Glycoblotting Method

Once *N*-glycans were enzymatically released, oligosaccharides carrying reducing terminal were chemically ligated with hydrazide-functionalized BlotGlyco H bead, allowing their selective capturing from the complex mixtures of serum originated biomolecules as diagrammed in Figure 4.1. In this glycoblotting-based quantitative N-glycomics strategy, 250 µL of BlotGlyco H bead was placed into each well of a MultiScreen Solvinert filter plate (Millipore) with vacuuming. 20 µL of PNGase F digested mixture containing released N-glycans was then mixed with the bead in each well, followed by the addition of 180 µL of 2% acetic acid (AcOH) in acetonitrile (ACN). To capture the N-glycans specifically onto beads via reversible hydrazone bonds, the plate was incubated at 80°C for 45 minutes. Next, the plate was washed twice with each 200 µL of 2 M guanidine-HCl in ABC, water, and 1% triethyl amine in methanol (TEA in MeOH). To cap unreacted hydrazide functional groups on beads, 10% acetic anhydride in MeOH was added with incubation for 30 min at room temperature and then removing the solution by vacuum. The beads were then washed twice with each 200 µL of 10 mM HCl, methanol, and dioxane. To prevent sialic acid dissociation under mild acidic condition or when directly ionized by MALDI-TOF/MS, on-bead esterification of its carboxyl groups was carried out by incubation with fresh 100 mM 3-methyl-1-p-tolyltriazene (MTT) in dioxane at 60°C for 90 minutes. This approach allows simultaneous analysis of neutral and acidic (sialylated) glycans in positive-ion detection mode. Subsequently, each well was washed twice using 200 µL of dioxane, water, methanol, and water. The glycans blotted on the beads

were labeled by trans-iminization reaction with 20 µL of 50 mM O-benzyloxyamine hydrochloride (BOA) and 180 µL of 2% AcOH in ACN with incubation at 80°C for 45 minutes. BOA-tagged N-glycans were finally eluted with 100 µL of Milli-Q water.
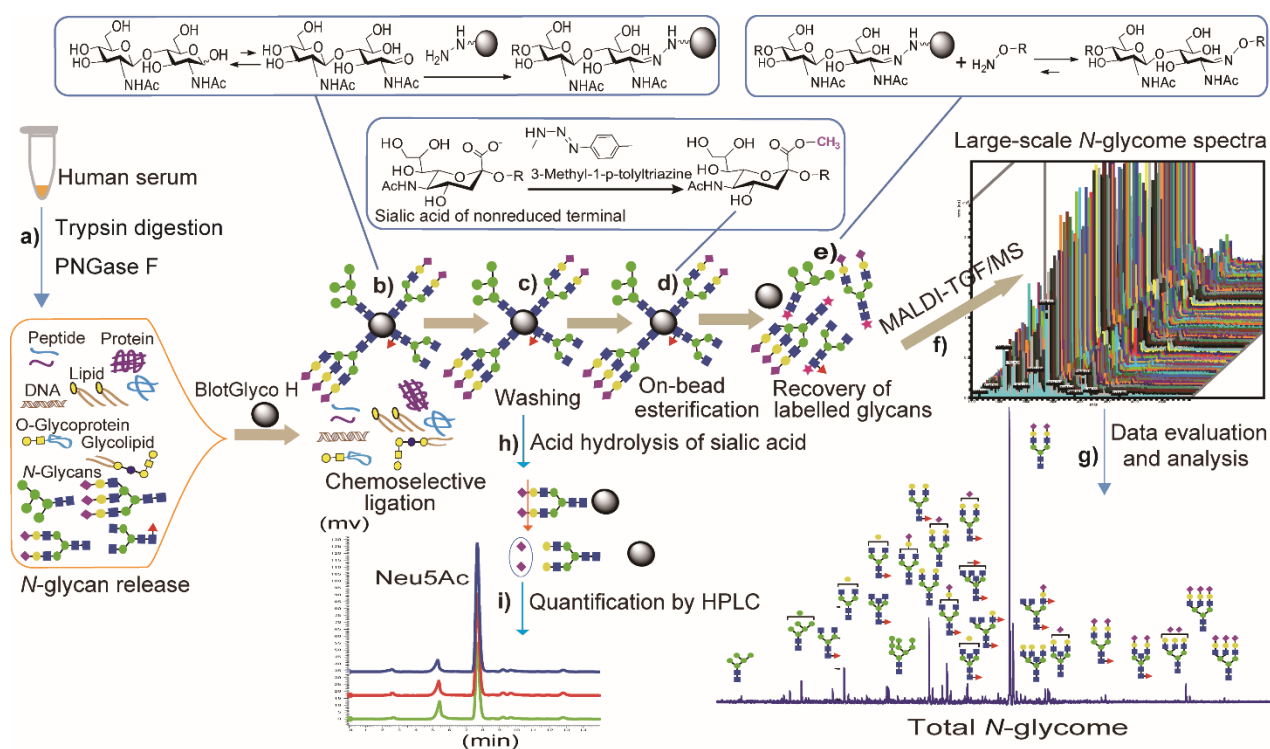
Figure 4.1. Schematic workflow of glycoblotting-assisted MALDI-TOF/MS and HPLC analysis. (**a**) Enzymatic treatment for serum protein digest and deglycosylation using trypsin and PNGase F, respectively, (**b**) Chemoselective capturing of reducing sugars onto a hydrazide-functionalized BlotGlyco H beads from complex mixture of biomolecules, (**c**) Washing to remove all impurities, (**d**) On-bead methyl esterification of sialic acid residues, (**e**) Recovery of BOA-labeled glycans, (**f**) MALDI-TOF/MS analysis, (**g**) Data processing and evaluation, (**h**) Cleavage of free sialic acid residues from captured N-glycans, (**i**) Free sialic acid labeling and quantification by fluorescence-HPLC.

4.2.5. Serum *N*-glycan Quantification by MALDI-TOF/MS

BOA-labeled N-glycans were directly dissolved with an equivalent volume of matrix solution (100 mM α-cyano-4-hydroxycin-namic acid diethylamine salt), after which 2.5 µL of each sample-matrix mixture was auto-spotted in quadruplicate on MTP 384 target plate (polished steel TF, Bruker Daltonics). Ultraflex III mass spectrometry that works based on matrix-assisted laser desorption/ionization-time of flight/mass spectrometry (MALDI-TOF/MS) was used during which mass spectra were acquired in an automated manner using AutoXecute flexControl software (Bruker Daltonics, Germany) in reflector, positive ion mode, typically summing 1000 shots for each spot. The obtained mass spectra were further analyzed using FlexAnalysis v. 3 Software (Bruker Daltonics, Germany). The intensities from monoisotopic peaks of each quadruplicated spectra were normalized using known concentration of an internal standard and then averaged. This data was used for further statistical and quantitative comparison. Detected N-glycans were selected based on their quantitative reproducibility after evaluated using calibration curve of serially diluted human serum standards. *N*-glycan structural compositions were assigned by GlycoMod (ExPASy proteomics server, Swiss Institute of Bioinformatics: http://br.expasy.org/tools/glycomod/) using experimental masses, and by CFG database (http://www.functionalglycomics.org/glycomics/publicdata/home.jsp).

4.2.6. Quantification of Free Sialic Acids Cleaved from Serum *N*-glycoproteins

To demonstrate the versatility of our comprehensive glycoblotting method for HPLC based sialic acid quantification, previously reported methods [26, 27] were slightly modified and integrated to our automated glycan enrichment approach. Briefly, after enzymatic release of serum N-glycans under the digestion conditions described above, 20 µL from each sample

containing released N-glycans was subjected to glycoblotting with BlotGlyco H beads in a similar procedure as described above except sialic acid esterification with MTT and subsequent procedures were excluded in this case. Next to acetyl capping and subsequent washing, 100 μL of 25 mM HCl was added to each well, with sealing and incubating of the plate at 80°C for 1 hr to release terminal sialic acids via selective cleavage of the α-glycosidic bond from the adjacent galactose residues (Figure 4.1h). The hydrolysate containing trimmed sialic acid was filtered and collected into PCR tubes. The filtrate was then reacted with 100 μL of 7 mM 1,2-diamino-4,5-methylenedioxybenzene (DMB) reagent (prepared by dissolving DMB−2HCl powder in equal volumes of 1 M 2-mercaptoethanol and 18 mM $Na_2S_2O_4$) with heating at 60°C for 2.5 hrs in dark, allowing labeling of sialic acids. This DMB reagent effectively derivatizes all type of sialic acids without any side reaction as it specifically reacts with their α-keto acid moiety (see Figure 4.2). After stopping the reaction by cooling in ice water, 50 μL of the mixture solution was transferred into vial tube from which 10 μL was auto-injected into fluorescence HPLC for analysis on a reversed-phase column. The column was eluted at a flow rate of 1 mL/min using MeOH/ACN/H_2O solution mixture (3:1:10, v/v). The fluorescence was detected at 448 nm using excitation at 373 nm in a D-7000 HPLC system equipped with an L-7485 fluorescence detector (Hitachi High-Technologies Co., Tokyo, Japan). Standard mixtures of *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc, a hydroxylated form of the common sialic acid, Neu5Ac) were run parallel with the samples to normalize the retention time of each peak on the column, allowing quantitative analysis.
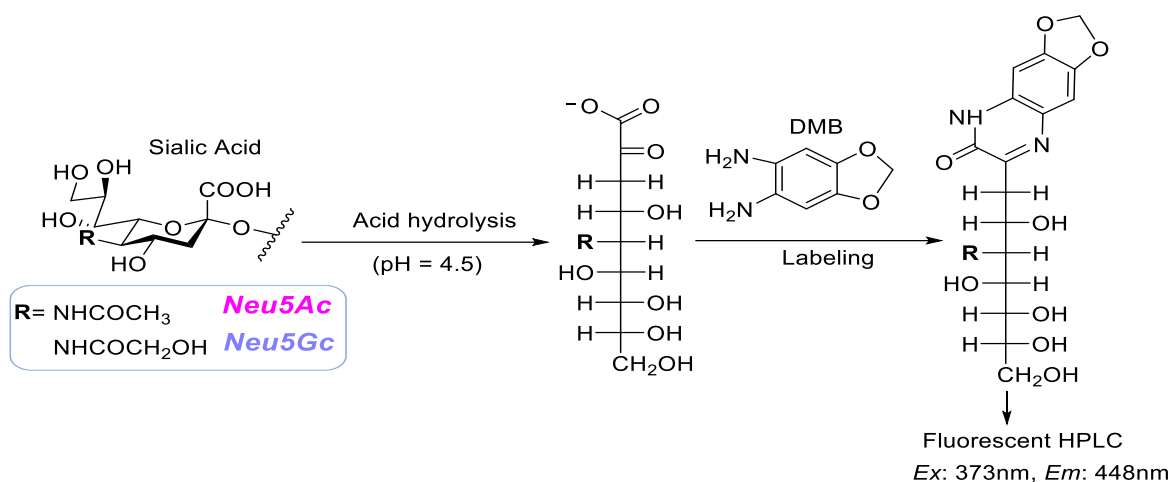
Figure 4.2. Scheme for hydrolytic cleavage of terminal sialic acid and its labeling for fluorescence detection. DMB: 1,2-diamino-4,5-methylenedioxybenzene, Neu5Ac: *N*-acetylneuraminic acid, Neu5Gc: *N*-glycolylneuraminic acid.

4.2.7. Statistical Analysis

N-glycan peaks detected in MALDI-TOF/MS spectra were annotated using FlexAnalysis 3.0 software (Bruker Daltonics, Germany). Normalized data for expression levels of *N*-glycans were analyzed using SPSS software. Multiple comparisons among the ethnic groups were done using one-way analysis of variance (ANOVA). P-values were adjusted for multiple testing using Bonferroni method and mean value differences were considered significant at 95% confidence interval ($p \leq 0.05$). We also used Graph Pad prism 5 to show the serum *N*-glycan and free sialic acid level of individual samples in a scatter dot plot.

4.3. Results

4.3.1. Quantitative Reproducibility Test Using Standard Human Serum

To select the reliably detected *N*-glycans, each peak was evaluated for its quantitative reproducibility using serially diluted standard human serum samples (0.5×, 0.75×, 1.0×, 1.25×, 1.5×, 1.75×, 2.0×, and 2.25×) that were simultaneously experimented in the same plate beside the study samples. The peak intensity of each glycan was first normalized using known concentration of an internal standard. Standard calibration curve for each *N*-glycan across the dilution series was plotted using the normalized data as shown in Figure 4.3. *N*-glycan peaks that met the criteria of at least $p < 0.05$ detection reliability, accessibility in at least six of the total eight human serum standards, and minimum outlier scores were selected and considered for quantitative comparison in the result of main study samples.
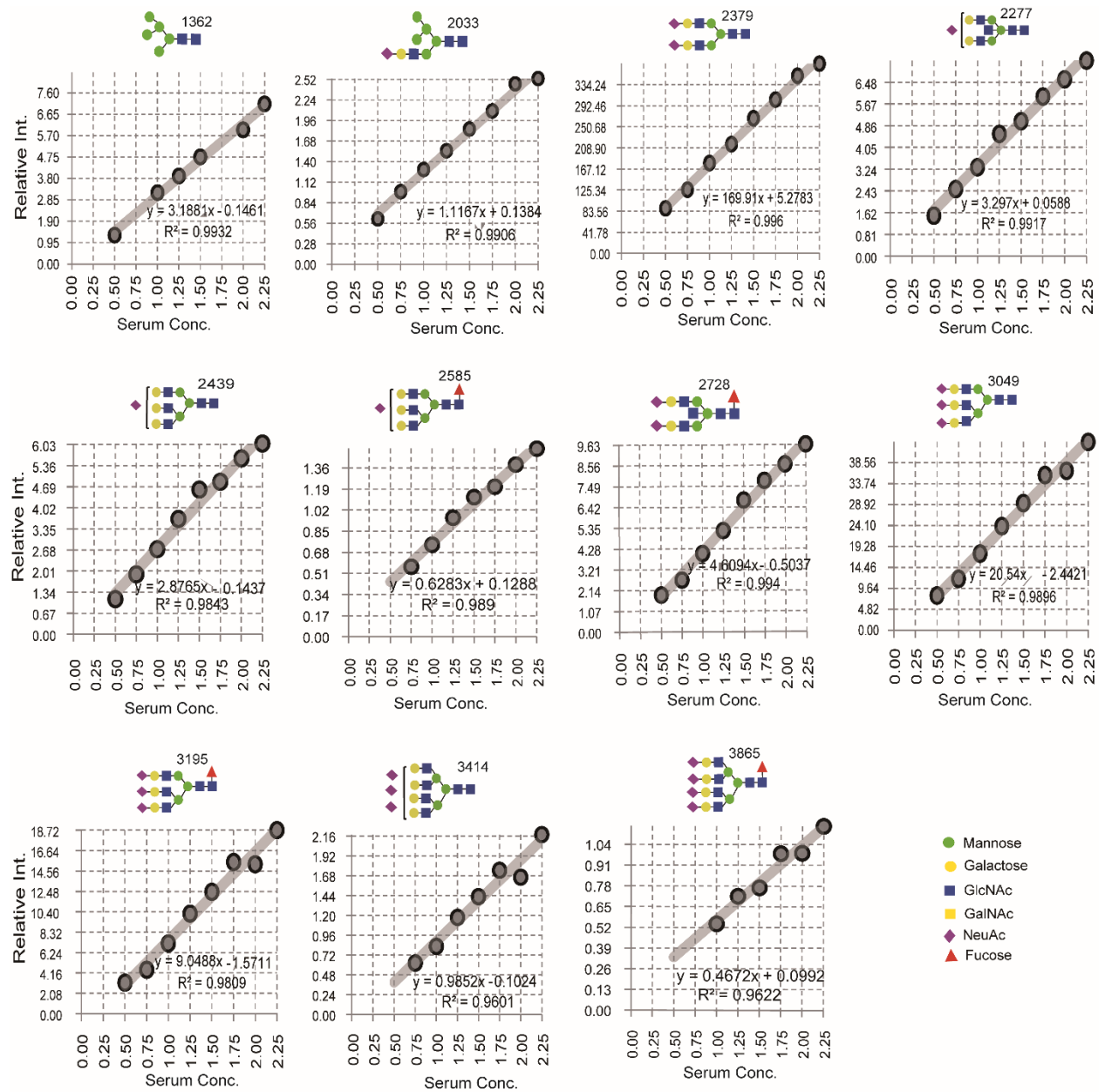
Figure 4.3. Human serum standard calibration curve of some common *N*-glycans. Glycans that have shown quantitatively reproducible detection across the various standard serum concentrations were considered and selected in the case of the main study samples. For identification purpose, individual glycans are coded or labeled by their molecular weight (*m/z* value) shown near to their structure.

4.3.2. Inter-Ethnic Variation in the Total Serum *N*-glycan Profile

This comprehensive *N*-glycomics study mainly emphasized investigating the association between serum glycan expression pattern and ethnic variation among 54 healthy individuals comprising Japanese, South Indians, Ethiopians, and US origin controls. Mass spectral data acquired by MALDI-TOF/MS analysis demonstrated noticeable variation in the peak intensities of several glycans with the highest signals appeared in the Ethiopian ethnic group (Figure 4.4). Taking the entire study samples, detection of 51 *N*-glycans (Table 4.1) was quantitatively verified of which 33 *N*-glycans were common to all ethnic groups. Detection level of the remaining glycans was < 5 µM and exclusive either to the Ethiopian ethnic group (13 glycans, Table 4.1 highlighted in gold color shade) or to all the other ethnic groups (5 glycans, Table 4.1 highlighted in light green color shade). Here, it should be clear that glycans whose detection profile was not quantitatively reliable or limited to some samples were not considered to avoid biased conclusions. Overall, 40 (78.43%) of the detected *N*-glycans belong to complex type whereas high-mannose and hybrid types comprise 5 (9.8%) and 6 (11.76%), respectively.
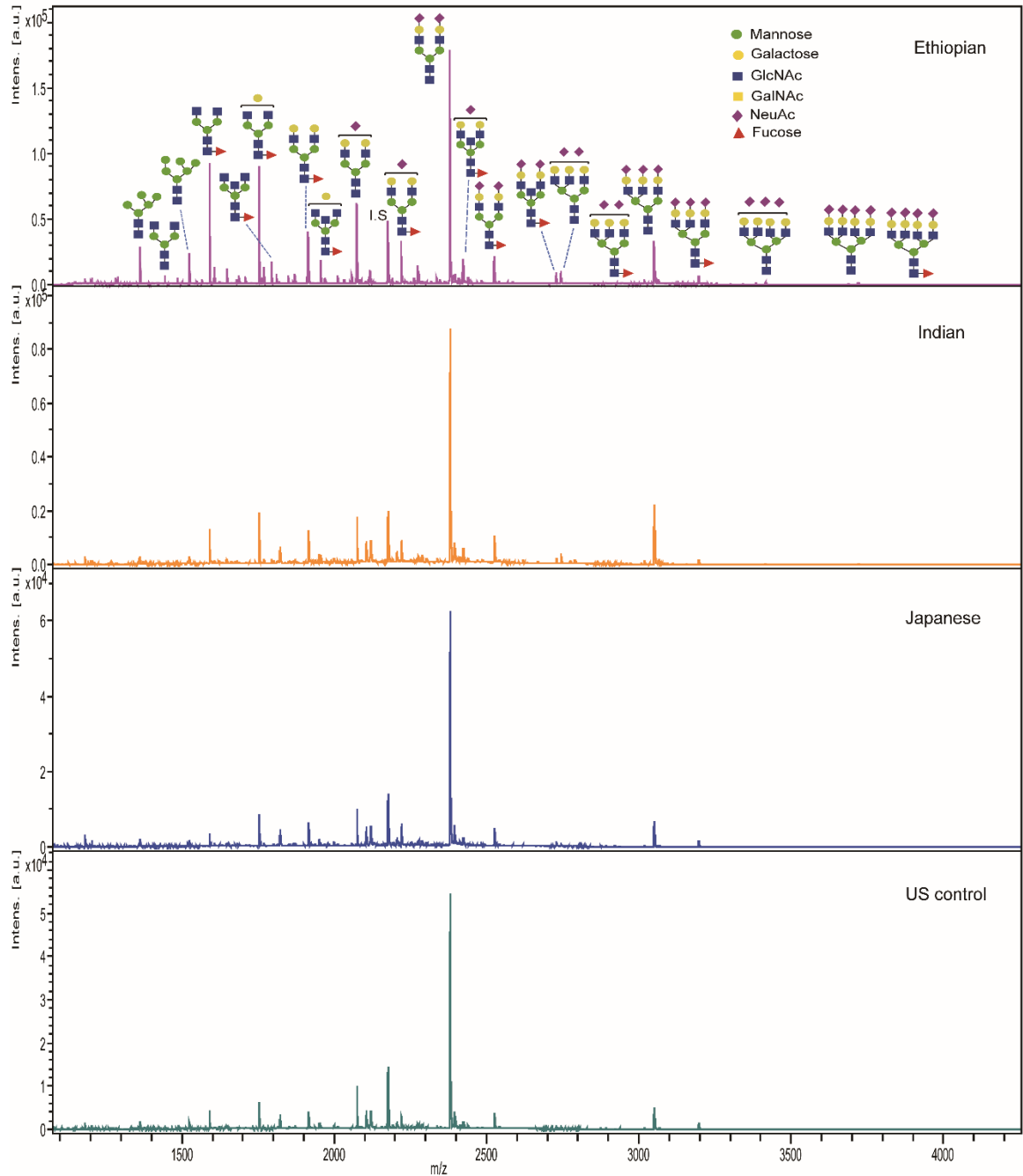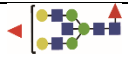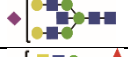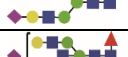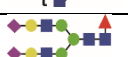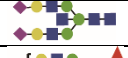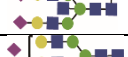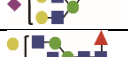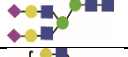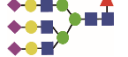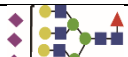
Fig 4.4. Representative MALDI-TOF/MS spectra of human serum *N*-glycans acquired from subjects of various ethnicity. Structures for some *N*-glycans that demonstrated differential peak intensities are indicated over the respective glycan peak.

Table 4.2. Estimated composition of 51 *N*-linked glycans identified from human serum glycoproteins.

Peak #	m/z	Glycan composition	Type	Structure
1	1362.48109	(Hex)2 + (Man)3 (GlcNAc)2	High-Mannose	
2	1444.53419	(HexNAc)2 + (Man)3 (GlcNAc)2	Complex	
3	1524.53392	(Hex)3 + (Man)3 (GlcNAc)2	High-Mannose	
4	1565.56047	(Hex)2 (HexNAc)1 + (Man)3 (GlcNAc)2	Hybrid	
5	1590.59210	(HexNAc)2 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
6	1606.58702	(Hex)1 (HexNAc)2 + (Man)3 (GlcNAc)2	Complex	
7	1686.58675	(Hex)4 + (Man)3 (GlcNAc)2	High-Mannose	
8	1708.61871	(Hex)1 (HexNAc)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
9	1752.64493	(Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
10	1768.63985	(Hex)2 (HexNAc)2 + (Man)3 (GlcNAc)2	Complex	
11	1793.67148	(HexNAc)3 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
12	1848.63958	(Hex)5 + (Man)3 (GlcNAc)2	High-Mannose	
13	1854.67662	(Hex)1 (HexNAc)1 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
14	1870.67154	(Hex)2 (HexNAc)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Hybrid	
15	1873.874	(Hex)3 (HexNAc)1 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Hybrid	
16	1911.69809	(Hex)1 (HexNAc)2 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
17	1914.69776	(Hex)2 (HexNAc)2 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
18	1955.72431	(Hex)1 (HexNAc)3 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
19	1996.967	(HexNAc)4 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
20	2010.69241	(Hex)6 + (Man)3 (GlcNAc)2	High-Mannose	
21	2032.72437	(Hex)3 (HexNAc)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Hybrid	
22	2057.75600	(Hex)1 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
23	2073.75092	(Hex)2 (HexNAc)2 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
24	2117.77714	(Hex)2 (HexNAc)3 (Deoxyhexose)1 + (Man)3 (GlcNAc)2	Complex	
25	2166.284	(Hex)3 (HexNAc)1 (Deoxyhexose)3 + (Man)3(GlcNAc)2	Hybrid	
26	2175.78261	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3 (GlcNAc)1	I.S	
27	2219.80883	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
28	2260.83538	(Hex)1 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	

29	2263.83505	(Hex)2 (HexNAc)3 (Deoxyhexose)2 + (Man)3 (GlcNAc)2	Complex	
30	2276.83030	(Hex)2 (HexNAc)3 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
31	2301.444	(HexNAc)4 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
32	2336.85144	(Hex)3 (HexNAc)4 + (Man)3 (GlcNAc)2	Complex	
33	2378.86199	(Hex)2 (HexNAc)2 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
34	2422.88821	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
35	2438.88313	(Hex)3 (HexNAc)3 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
36	2483.89335	(Hex)3 (HexNAc)1 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Hybrid	
37	2520.93623	(Hex)1 (HexNAc)5 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
38	2524.91990	(Hex)2 (HexNAc)2 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
39	2581.94137	(Hex)2 (HexNAc)3 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
40	2584.94104	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
41	2727.99928	(Hex)2 (HexNAc)3 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
42	2743.99420	(Hex)3 (HexNAc)3 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
43	2787.89	(Hex)3 (HexNAc)4 (Deoxyhexose)1 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
44	2890.05211	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
45	3007.09472	(Hex)4 (HexNAc)5 (NeuAc)1 + (Man)3 (GlcNAc)2	Complex	
46	3049.10527	(Hex)3 (HexNAc)3 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
47	3109.12641	(Hex)4 (HexNAc)4 (NeuAc)2 + (Man)3 (GlcNAc)2	Complex	
48	3195.16318	(Hex)3 (HexNAc)3 (Deoxyhexose)1 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
49	3414.23748	(Hex)4 (HexNAc)4 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
50	3560.29539	(Hex)4 (HexNAc)4 (Deoxyhexose)1 (NeuAc)3 + (Man)3 (GlcNAc)2	Complex	
51	3719.34855	(Hex)4 (HexNAc)4 (NeuAc)4 + (Man)3 (GlcNAc)2	Complex	
52	3865.40646	(Hex)4 (HexNAc)4 (Deoxyhexose)1 (NeuAc)4 + (Man)3 (GlcNAc)2	Complex	

The *m/z* values indicate experimental masses of *N*-glycans tagged with BOA for enhancing their ionization potentials. Glycans whose peaks # is highlighted by color shade had generally weaker detection profile in which those highlighted in gold color shade were specific to the

Ethiopian group while those highlighted in light green shade were common to the remaining groups. Peak number 26 detected at *m/z* 2175.78 represents an internal standard (I.S).

We quantitatively compared the expression levels of commonly detected *N*-glycans among the ethnic groups by ANOVA test. Ethnicity-associated variation over a wide range of serum *N*-glycans with a consistent highest abundance (except one glycan drastically declined; *m/z* of 2118) in the Ethiopian subjects ($p < 0.001$) was observed [Figures 4.5A-C]. Structurally, these significantly altered glycoforms are predominantly composed of high mannose, core-fucosylated, hyperbranched or hypersialylated types. Moreover, the trend for glycan abundance in Ethiopians tends to be isolated from the other healthy ethnic groups, rather seems closer to that of Japanese hepatocellular carcinoma (HCC) patients whose serum was concurrently re-examined for *N*-glycan profiling. Compared to the US controls, the Japanese and the Indian healthy human serum glycome showed a unidirectional increase in expression level whereas a statistically non-significant expression difference was noticed between the two Asian ethnic groups. As an exception, the result pointed out that a bisect-type *N*-glycan having *m/z* of 2118 showed a significant ($p < 0.001$) concentration alteration among all ethnic groups with the highest in the Indian and the lowest in the Ethiopian groups [see *m/z* 2118 in Fig 5B].

A) High mannose

***means P≤ 0.001
**means 0.001<P≤ 0.01
*means 0.01<P≤0.05

B) Core-fucosylated biantennary
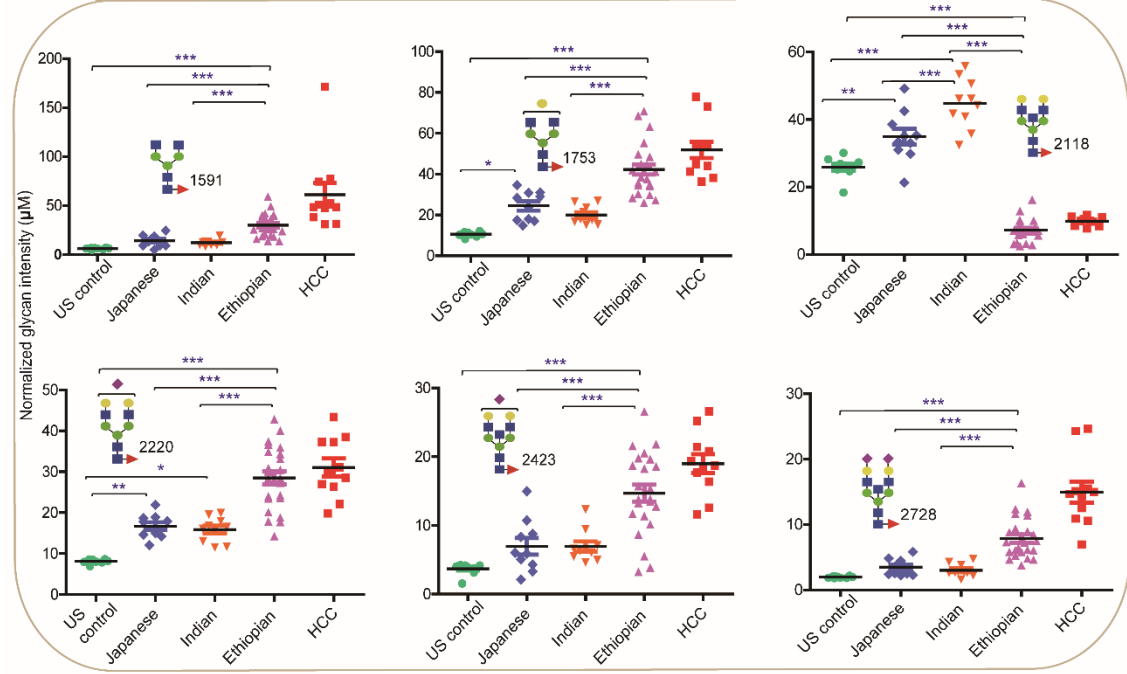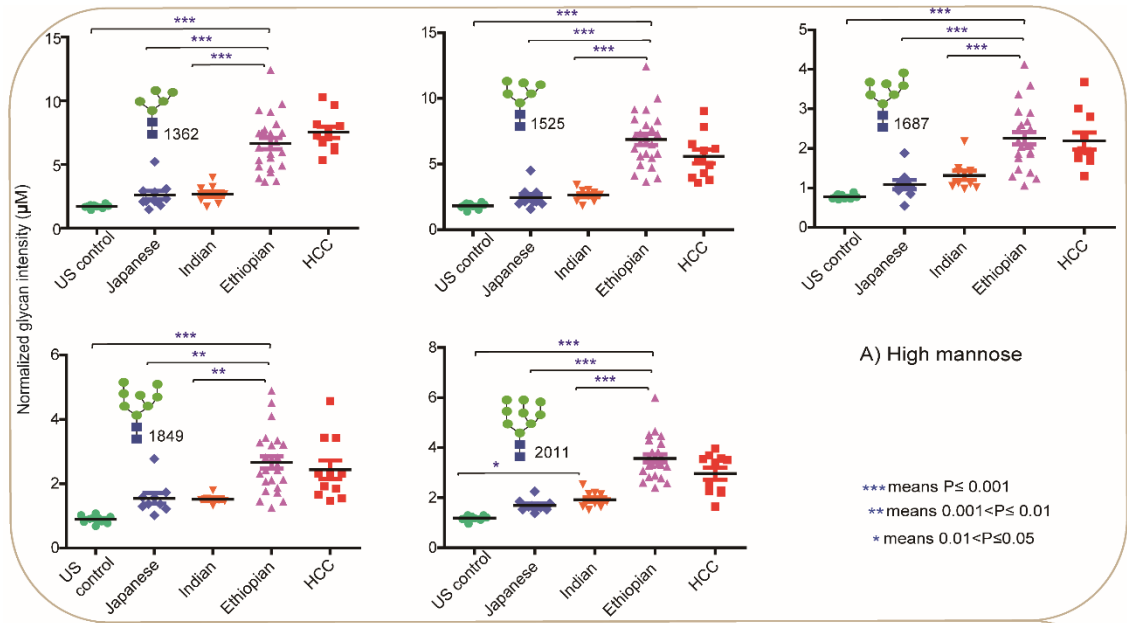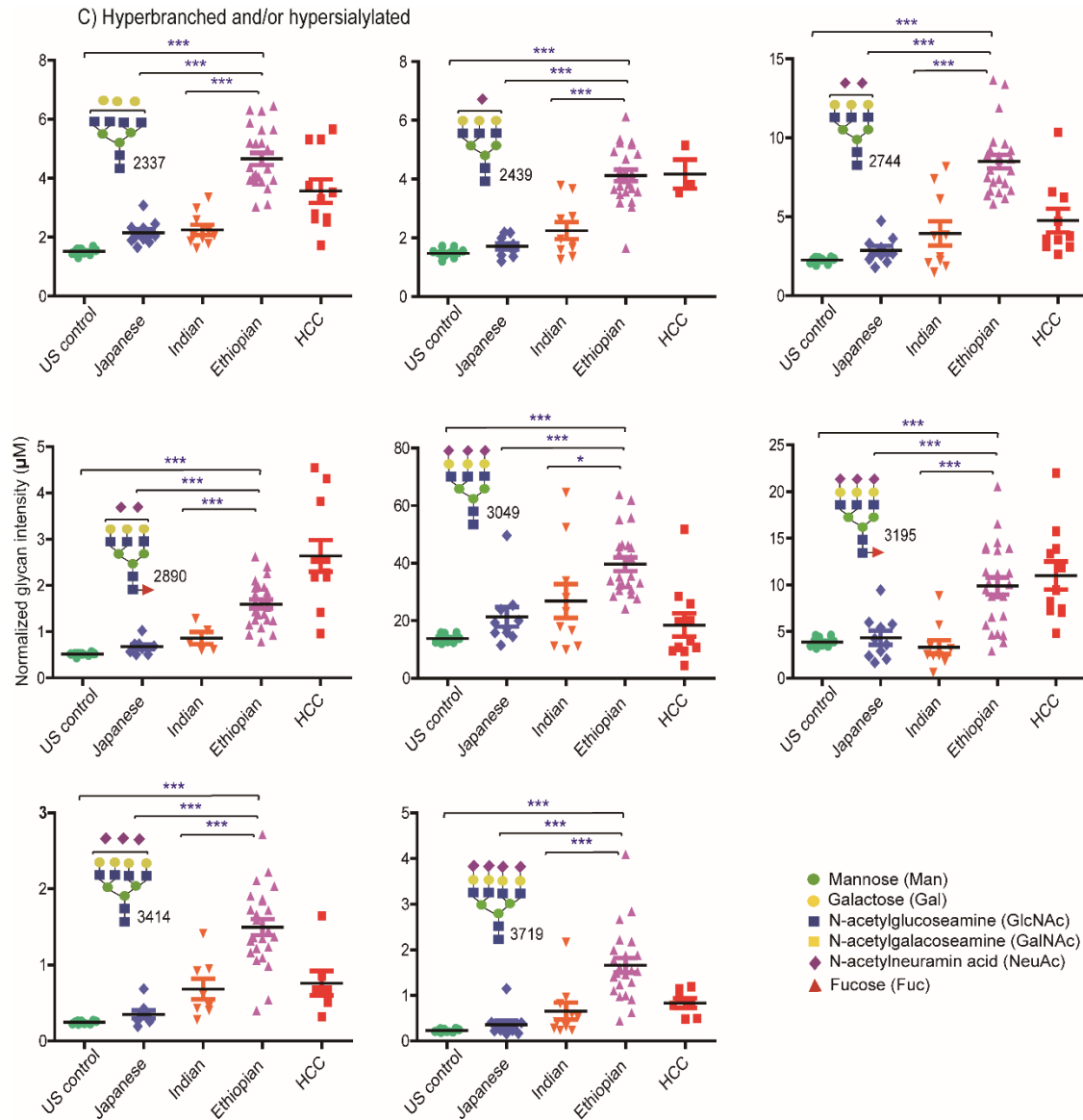
Figure 4.5. Dot plot illustrating comparative expression abundance of selected serum *N*-glycans. Proposed glycan structure and statistically significant differences among the groups is indicated for each glycoform**. A**) High mannose type, **B**) Core-fucosylated and biantennary structures, **C**) Hyperbranched and/or hypersialylated structures. HCC: Hepatocellular carcinoma patients.

Since Ethiopian subjects were females, for more clarity on whether differences in ethnicity, gender, or age seem to have more marked effects, we further included one Indian female (age = 43) and one Japanese female (age = 40s), whose serum *N*-glycome pattern in comparison with one Ethiopian female (age = 40) is diagramed in Figure 4.6. For further emphasis on the gender effect, spectral *N*-glycome profile of Indian and Japanese subjects of both sexes is provided in Figure 4.7. within which peak intensity of many *N*-glycans varied mainly in an ethnicity dependent manner, but irrespective of the gender difference. Altogether, these results show clear differences in the glycan abundance that were more strongly associated with ethnic differences than gender or age variation. The effect of age on serum glycan level within each ethnic group of this study was statistically non-significant.
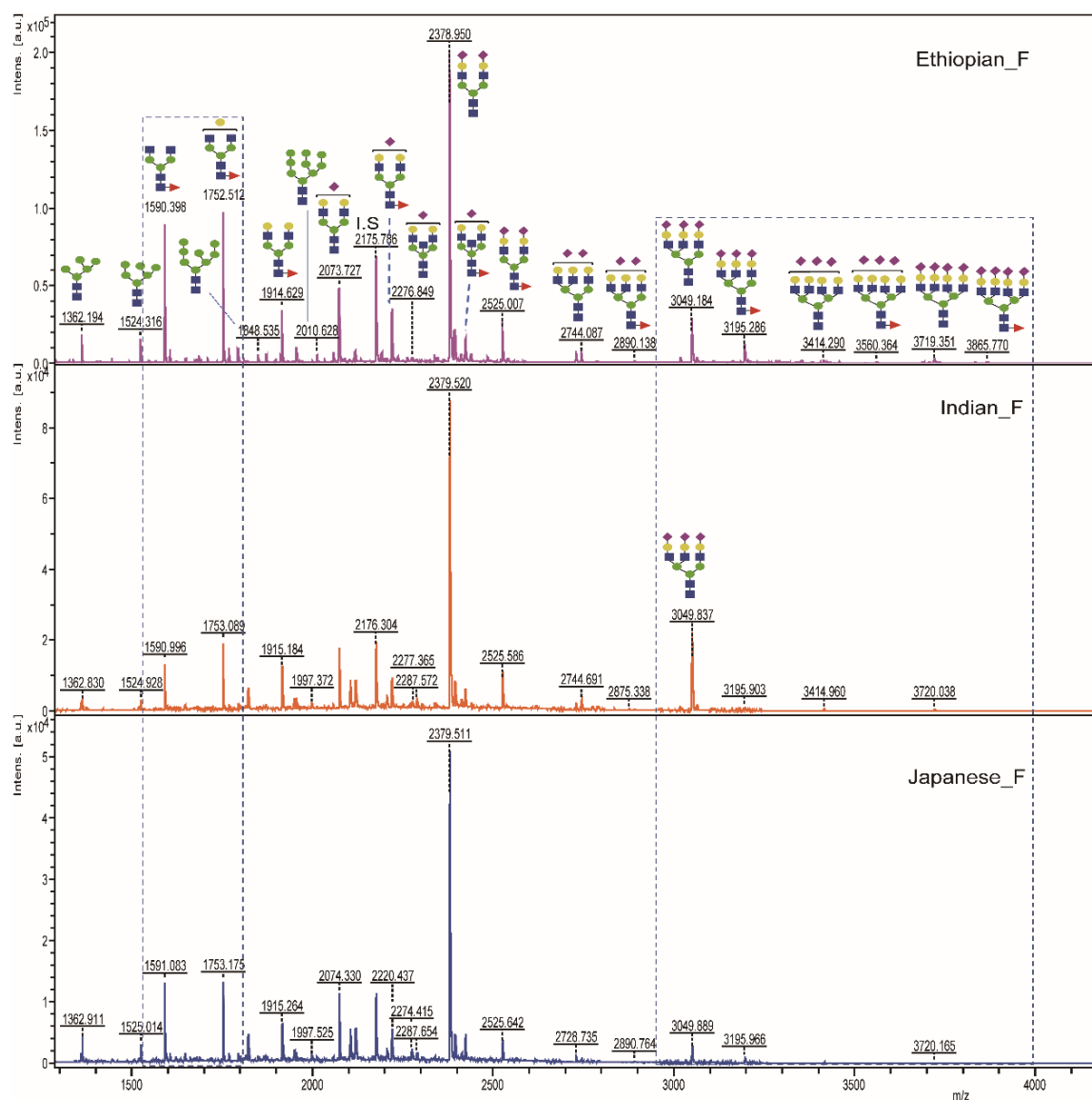
Figure 4.6. Comparative serum *N*-glycome spectra of age-matched females varying in ethnicity. In attempt to provide more clarity on whether difference in ethnicity or gender had marked effect on the *N*-glycan profile, we have included one Indian female (age = 43) and one Japanese female (age = 40s), whose serum *N*-glycome spectra is comparatively presented with that of one Ethiopian female (age= 40). Some of the *N*-glycan peaks showing marked variations among the three female subjects are highlighted by the dotted line shape. Considering this result from gender and age matched samples, most serum glycoforms showed abundant peak intensity in the Ethiopian subject, while one triantennary trisialylated glycan (m/z ~3049) demonstrated highest intensity in the Indian sample. These results from few female subjects intensify the

variations observed in serum *N*-glycan profile result when Indian, Japanese, and US male subjects were considered as well, evidencing the profound influence of ethnicity on the *N*-glycosylation signature of the study groups.
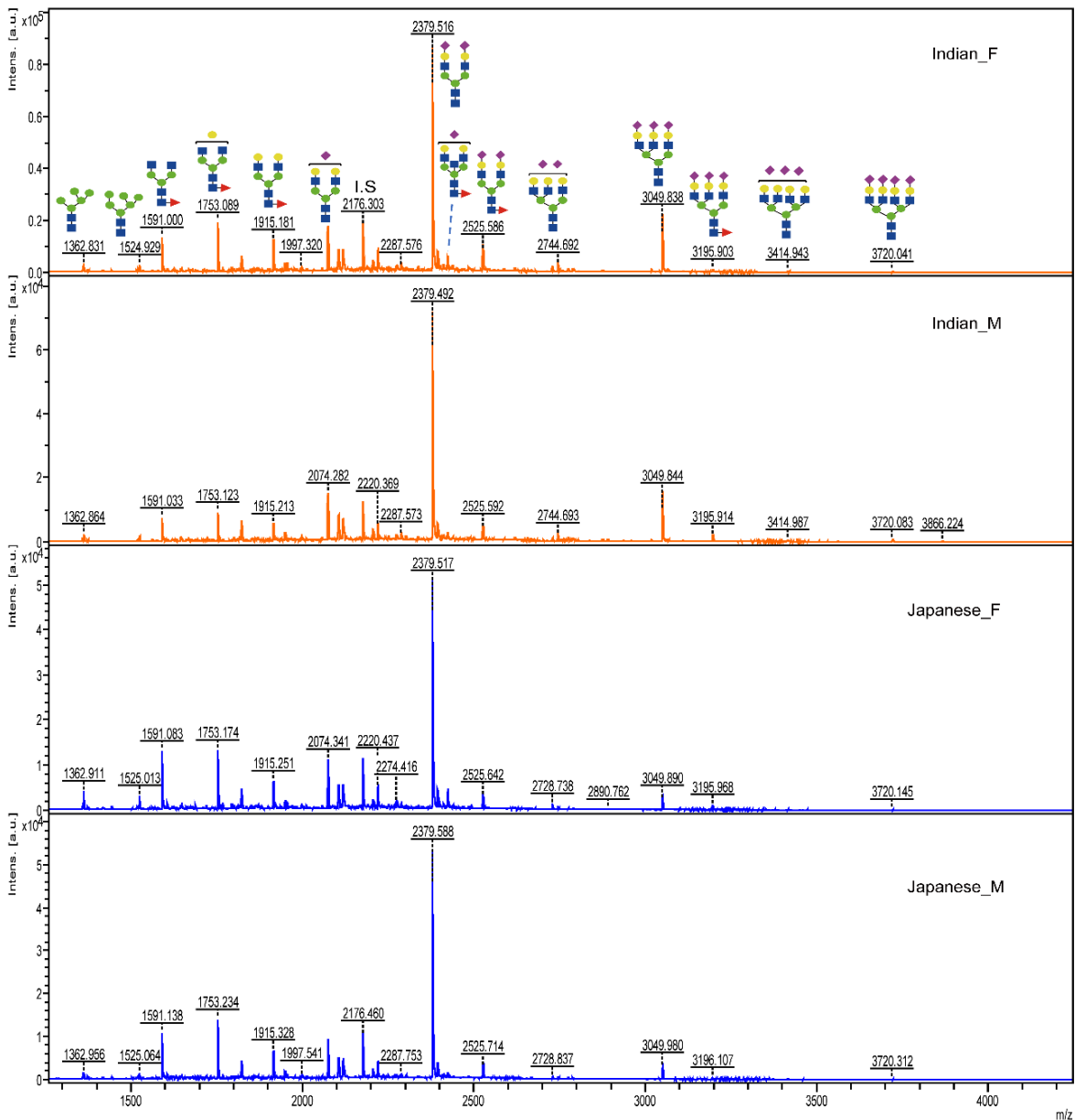
Figure 4.7. Serum *N*-glycome spectra showing Indian and Japanese of both sexes. Age of subjects: Indian female = 43, Indian male = 39, Japanese female = 40s, Japanese male = 30s. Variation in *N*-glycan peak intensity is clearly shown in an ethnicity dependent manner in

which glycans such as the core-fucosylated (m/z 1591 and 1753) illustrated abundant expression in the Japanese subjects, whereas hyperbranched and hypersialylated glycans (m/z 3049 and 3195) demonstrated up-regulated expression in the Indian subjects, irrespective of the gender difference.

To address the inter-individual variation of each *N*-glycan level from one ethnic data series to another, we determined coefficient of variation (CoV) for each glycan as the ratio of the standard deviation to the mean of glycan expression. The degree of variation was found to be lowest across the entire glycans in the US control sample. On the other hand, there were no consistent inter-subject differences among Japanese, Indian and Ethiopian ethnicities up to *m/z* of 2500, after which the hyperbranched and hypersialylated glycans showed widely dispersed expression pattern in the Indian and the Japanese groups (Figure 4.8).
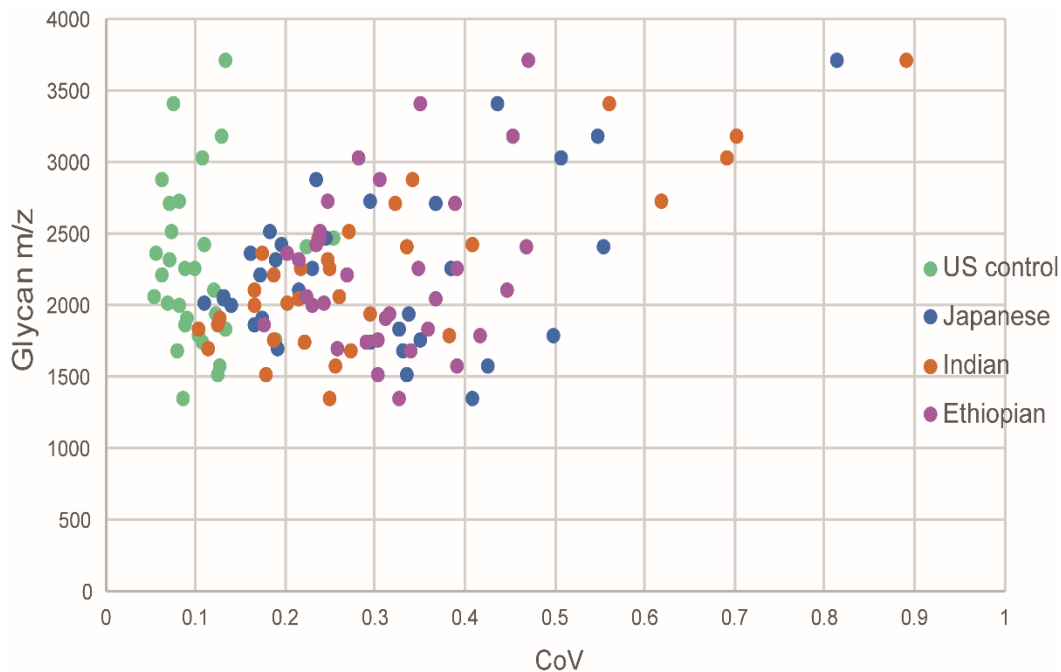
Figure 4.8. CoV illustrating the inter-subject extent of variation within each ethnic group. Y-axis shows *m/z* values of detected *N*-glycans, X-axis shows the coefficient of variation (CoV) value representing how far the glycan concentration was scattered within subjects of each ethnic group.

4.3.3. Glycotyping Analysis

After *N*-glycans were stratified into glyco-subclasses that share certain structural features of core-fucosylation, bisecting, sialylation, and branching (more details provided in Table 4.2), their overall expression pattern among the study groups is shown in Figure 4.9. The result clearly revealed that profound abundance of these glycotypes (except the bisecting) was found to be associated with the Ethiopian population. Ethnic-based greater disparities in the serum level of the glycoforms was mainly pronounced towards the higher *m/z* tri-/tetra-sialylated and tri-/tetra-antennary structures. Next to Ethiopians, consistent declining trend in the expression level was noticed in Indians, Japanese, and finally US controls, respectively. Minimum inter-ethnic differences in the level of bisecting glycoforms was observed across all ethnicities.

Table 4.3. List of *N*-glycans considered for glyco-subclass analysis.

Core-fuco	Bisecting	Mono/Di-Sia	Tri/Tetra-Sia	Tri/Tetra-Anten
1591	1794	1709	3049	2337
1753	1956	1871	3195	2439
1794	2118	2033	3414	2890
1915	2277	2058	3719	3049
1956	2728	2074		3195
2058		2220		3414
2118		2277		3719
2220		2379		
2264		2423		
2423		2439		
2525		2525		
2728		2728		
2890		2744		
3195		2890		

Only glycans that were detected in all the ethnic groups have been considered for the glyco-subclass analysis. There is a chance that a glycan can be counted in more than one group when it contains more than one structural features as per the grouping mechanism. m/z values are given as label with each glycan structure.
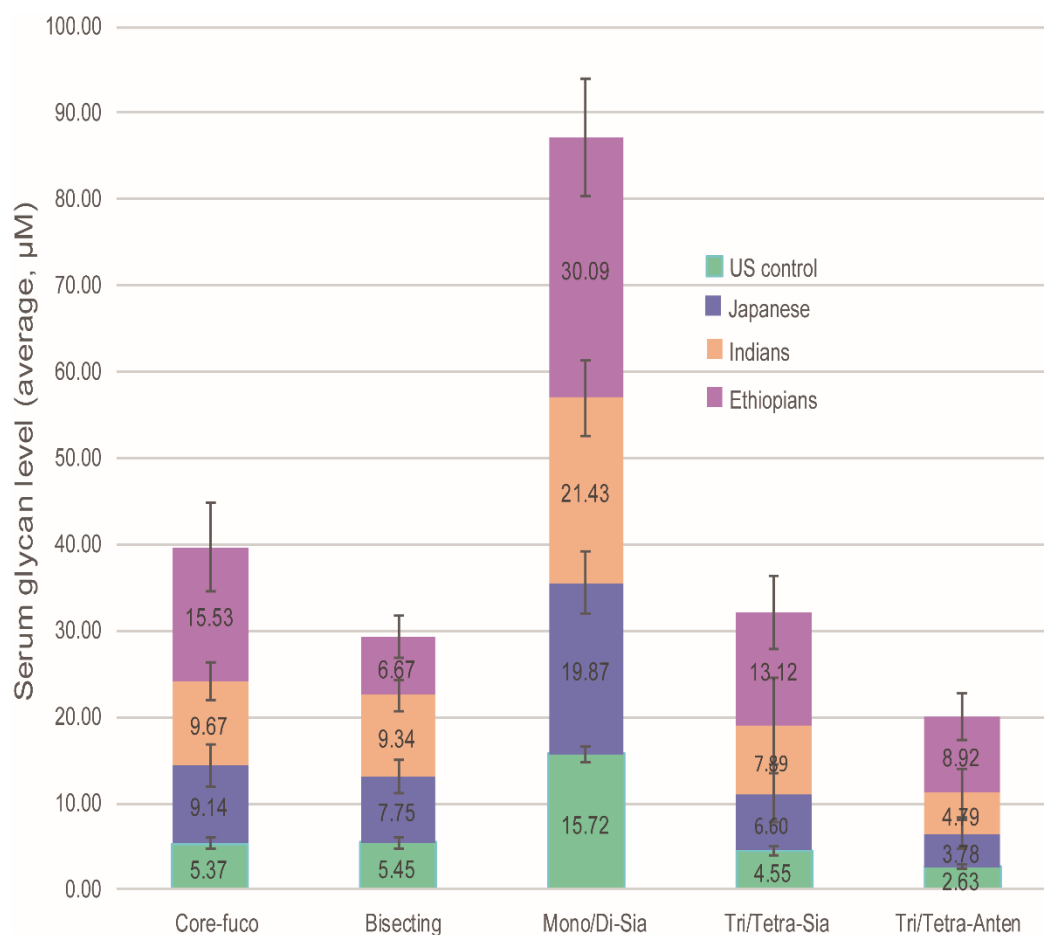
Figure 4.9. Ethnic-based serum level of some common glyco-subclasses. *N*-glycans were grouped in terms of core-fucosylation, bisecting, mono- or di-sialylation, tri-or tetra-sialylation, tri- or tetra-antennary features. Error bars indicate standard errors of the means.

4.3.4. Sialic Acid Quantification in Human Serum

We also performed a glycoblotting-assisted HPLC-based quantification of free sialic acid species cleaved from sialylated *N*-glycans captured by glycoblotting of human serum glycoproteins. 1,2-diamino-4,5-methylenedioxybenzene (DMB) facilitated selective labeling of sialic acid moieties was performed prior to HPLC-fluorescence detection. Standard solutions of *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) in 50-750 µM concentration ranges were analyzed in parallel with the serum samples. Samples from four

ethnic groups (US controls, Japanese, Indians and Ethiopians) were used in this experiment. The result demonstrated a consistent signal intensity pattern for the standard Neu5Gc and Neu5Ac in a concentration dependent manner as shown by HPLC profiles in Figure 4.10A), suggesting the reliability of the quantification method. Focusing on the study samples, highest peak intensity for Neu5Ac was primarily found in the Ethiopian group (Figure 4.10B) whose absolute concentration (average $\mu M \pm SE = 234.23 \pm 24.15$), as determined based on the standard Neu5Ac calibration curve (Figure 4.11A), was significantly higher ($p < 0.001$, $p = 0.002$, $p = 0.01$) comparing with that of US controls (105.51 ± 9.39), Japanese (139.84 ± 20.86), and Indians (155.69 ± 14.27), respectively (Figure 4.11B). Apart from inter-ethnic variation, Neu5Ac level was observed to be very low in only a few of the Ethiopian subjects, implicating the need to consider inter-individual disparities in the glycosylation process. However, Neu5GC was identified in none of the study subjects which is not surprising as this type of sialic acid normally does not exist in healthy human serum glycoproteins [28].
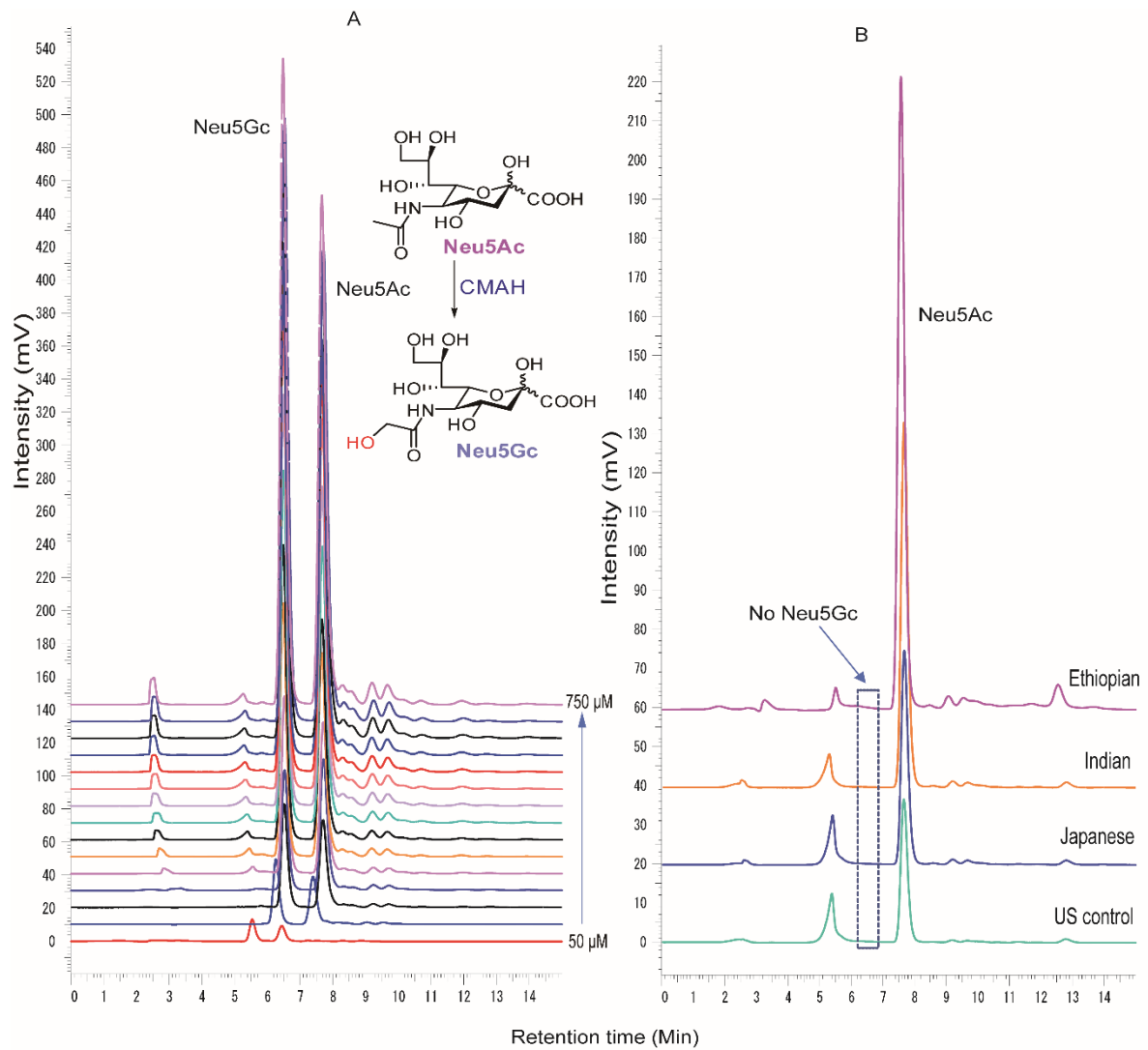
Figure 4.10. HPLC chromatogram showing quantitation of free sialic acids. **A**) standard Neu5Gc and standard Neu5Ac, **B**) Neu5Ac derived from captured serum *N*-glycans. Neu5Ac: *N*-acetylneuraminic acid, Neu5Gc: *N*-glycolylneuraminic acid. CMAH: Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase (enzyme that catalyzes conversion of Neu5Ac to Neu5Gc).
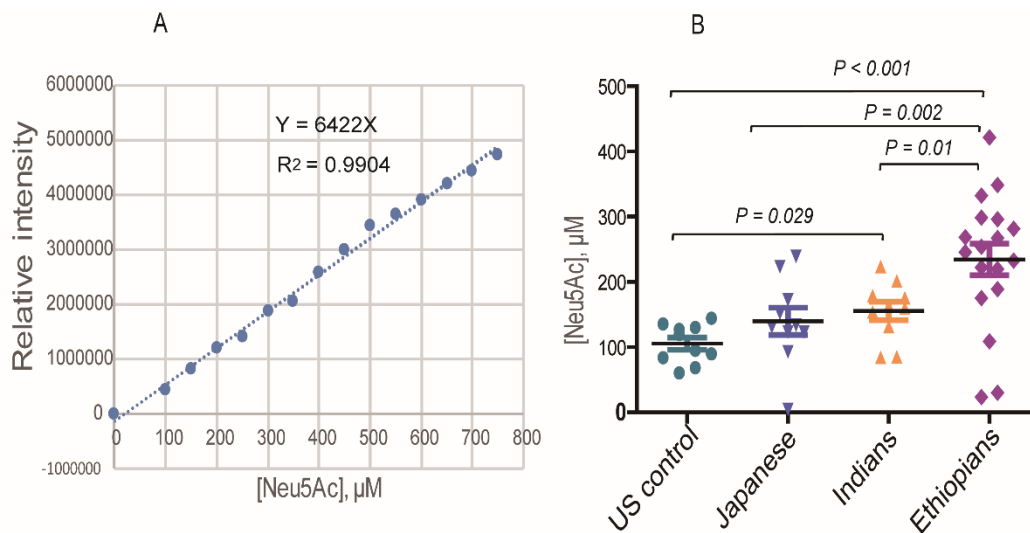
Figure 4.11. HPLC-based Neu5Ac quantification. **A)** standard Neu5Ac calibration curve, **B)** Dot plot showing Neu5Ac expression levels derived and quantified from human serum *N*-glycans of four ethnic groups.

Furthermore, HPLC chromatogram produced from serum Neu5Ac level of three age-matched females varying in ethnicity (Ethiopian 40 y/o, Indian, 43 y/o, Japanese 40s y/o) is provided in Figure 4.12, which shows marked differences in their Neu5Ac peak intensity, being most abundant in the Ethiopian subject. This observation was consistent with the result of total serum *N*-glycome spectra of the same individuals [Figure 4.6], all of which indicating that healthy human serum *N*-glycosylation pattern seems to be affected more strongly by ethnic difference than gender or age variation of the participants.
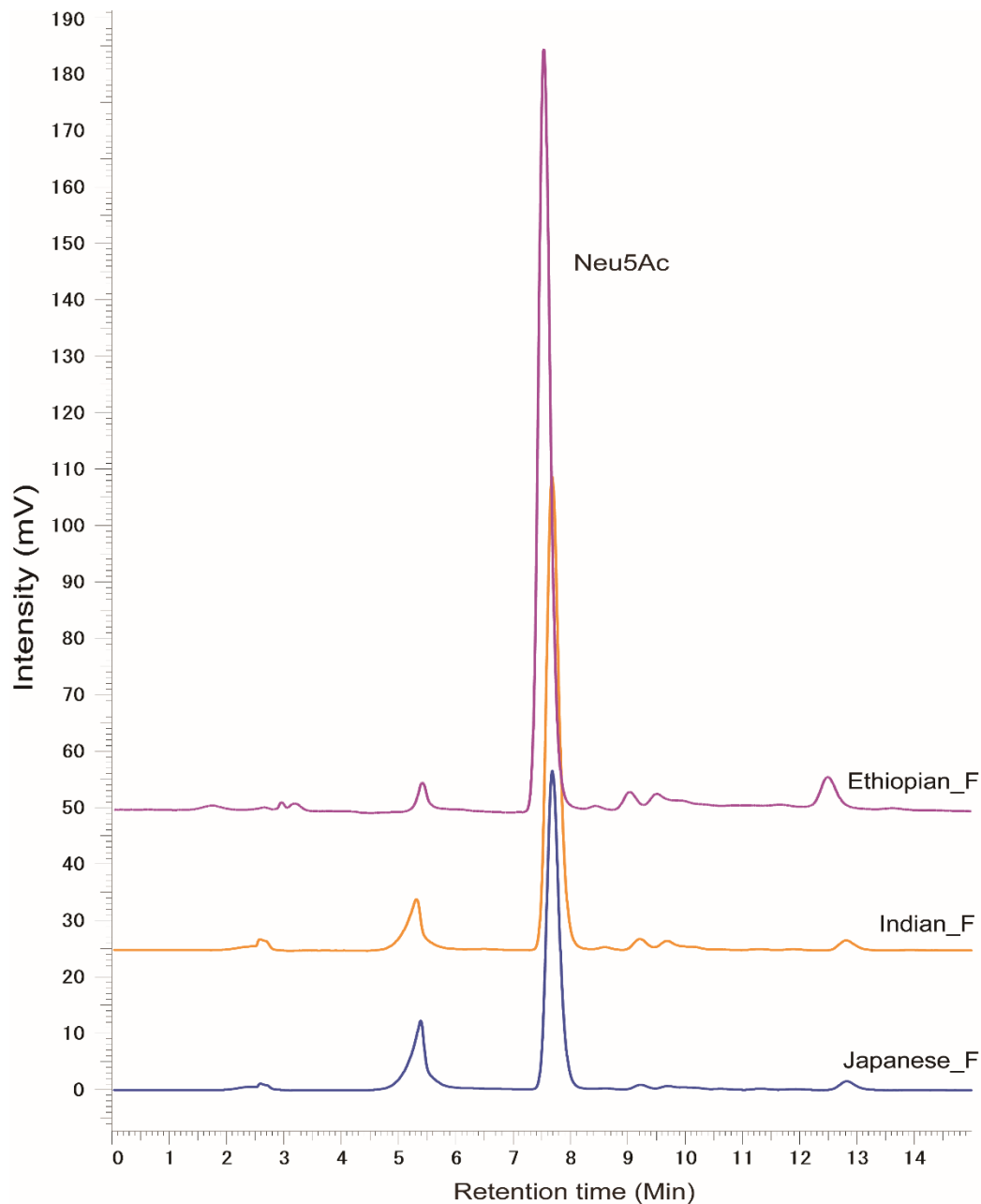
Figure 4.12. HPLC chromatogram for free Neu5Ac captured from serum of three age-matched females varying in ethnicity. The female subjects were each of Ethiopian (age= 40), Indian (age = 43), and Japanese (age = 40s). This Neu5Ac sialic acid result from age- and gender-matched subjects provides an evidence for the strong influence of ethnic deference on the sialylation pattern of human serum glycoproteins and strengthens our total sialylated *N*-glycan result that demonstrated ethnic-associated variation, being most abundant in the Ethiopian subject.

4.4. Discussion

Aiming at investigating the inter-ethnic physiological variations in serum *N*-glycome profile, we performed a glycoblotting-assisted MALDI-TOF/MS-based quantitative analysis focusing on 54 healthy subjects of various ethnicity. Total of 51 *N*-glycoforms could be identified and reproducibly quantified with an evident differential expression pattern among the ethnic groups. Among all, Ethiopian ethnic group exhibited the most isolated detection trend with greatly increased expression levels of particularly high mannose, Core-fucosylated, multi-antennary, and multi-sialylated glycans. Notably, some less intense glycans had exclusive occurrence in which 13 glycan structures were specific to Ethiopians whereas other 5 glycoforms were identified to be associated with the other ethnicities. Healthy Indian and Japanese serum glycoproteins seem to have mostly similar *N*-glycan composition and abundance with slight differences in the expression level of few *N*-glycoforms. It is interesting to note that one glycoform carrying a core-fucose and a bisecting GlcNAc structures (m/z 2118) showed a clear-cut expression difference among the four study groups. It is the only glycoform whose serum level was markedly decreased in Ethiopians comparing with the other ethnic groups. Declining of this glycan in Ethiopian subjects is most probably due to its involvement as a substrate in the subsequent biosynthetic steps, the product of which (*m/z* 2423 and *m/z* 2728 in Figure 4.5B) could be up-regulated comparing with the other ethnic populations. We have found high inter-individual variations in the expression level of multiply branched and sialylated glycan structures in the Indian and Japanese subjects (Figure 4.8). Combining both observations of lower inter-individual variation and higher glycan concentration (compared with Indians and Japanese) found in Ethiopians, it can be suggested that total serum glycan concentration was elevated across the entire Ethiopian subjects. Marked differences in glycotyping analysis, in which up to 2-3 times greater abundance particularly in the multi-

branched or multi-sialylated glycan features, were associated with Ethiopian ethnic group (Figure 4.9). This observation was consistent with the above-mentioned differential expression pattern of the individual *N*-glycans.

Increasing evidences show that glycan abundance is regulated by multiple molecular mechanisms that rely on metabolic interplay among genes, sugars, proteins, and lipids [2, 29]. In glycoproteins, the amino acid sequences of the core protein mostly remain stable while the glycan moiety undergoes faster alterations in response to physiological, pathological, and environmental stimuli [30-32]. Despite previous studies that have reported the correlation between changes in plasma/IgG glycosylation patterns and aging process [19-21], there was no age wise significant variation of healthy serum *N*-glycome profile in the current study, probably due to the relatively younger age range of the healthy participants. With quite different background of our study subjects, the observed changes in the present study may be explained by genetic, environmental, (nutritional habits, physical activity, exposure to pathogens, stress level), and sociocultural variations, while the degree to which each factor influences serum glycan level may depend on the specific ethnic population. These variations have been emphasized by previous reports addressing the variability, heritability, genetic and environmental determinants of human plasma and IgG *N*-glycome using chromatographic techniques [19, 33, 34]. Given the complexity of glycosylation at cellular and molecular levels, these diverse factors can ultimately affect the activity of glycosyltransferases and glycosidases enzymes that orchestrate the glycan biosynthetic and degradative pathways [35]. In this context, despite the non-template-based synthesis of glycans, their structures and expression abundance have been reported to be regulated by genetic and epigenetic factors [36, 37], partly accounting for the inter-individual glycan signature variations observed in health and disease conditions. It was also pointed out that variations in the composition of plasma protein *N*-

glycans, particularly increased branching, galactosylation and sialylation features, have been associated with metabolic syndrome related risk factors and higher risk of developing type 2 diabetes [11, 12]. While literatures on the association of healthy human serum glycosylation signatures with ethnicity are scarcely available, numerous studies have witnessed ethnic-specific differences in a number of biochemical markers [38, 39]. Although appreciating population diversity during comprehensive experimental studies is mostly uncommon, we have considered black Ethiopians/Africans in the present study which can be taken as an advantage as pattern of glycosylation in black population, from physiological or pathological perspective, has not been addressed elsewhere.

Surprisingly, some of the specific *N*-glycoforms observed to be exclusive or elevated in the Ethiopian ethnic group including the hyper-branched structures with sialic acid residues (mz 1362, 1591, 3195, 3560, 3865) have previously been identified as sensitive serum biomarkers as a significant recurrence factor of HCC using large-scale Japanese samples [22]. In the present study, these glycans could demonstrate nearly similar serum expression pattern between the Ethiopian group and the HCC group (11 Japanese patients) up on simultaneous experimentation (Figure 4.5). Similarly, among the hyperbranched glycans strongly increased in Ethiopians ($p < 0.001$), it was demonstrated that the glycans with m/z 2337, 2439, and 2890 could become promising prognostic biomarkers in renal cell carcinoma [40] while m/z 3049 and 3414 have been associated significantly with metastatic castration-resistant status in prostate cancer [41]. Hence, alteration in serum *N*-glycan profile seems not exclusive for pathological conditions as the present results also clearly demonstrated among healthy subjects in an ethnicity dependent manner. Altogether, these observations emphasize the substantial impact of ethnic differences in human serum *N*-glycome variation, the ignorance of which may

provide unclear and imprecise conclusion of the diagnosis by using glycan-related disease biomarkers.

Free sialic acid quantitation result among the four ethnic groups revealed nonnegligible ethnic differences in the serum Neu5Ac level in which the highest abundance has been shown in the Ethiopians, compared to the remaining groups [Figure 4.10B and Figure 4.11B]. These informative ethnic-associated variations in the free sialic acid residue further strengthen our MALDI-TOF/MS-based quantification results that demonstrated a consistent declining trend in the expression levels of sialylated glyco-subclasses among Indians, Japanese, and US controls, respectively. The non-detection of Neu5Gc in the current result agrees with the fact that humans do not naturally produce it because of the species-specific embryonic inactivating mutation of the gene encoding for CMP-Neu5Ac hydroxylase enzyme that converts Neu5Ac to Neu5Gc [28]. In addition, it is clear that there is little influence of exogenously incorporated Neu5Gc to the *N*-glycan biosynthesis of major serum glycoproteins. Neu5Ac, being a chief contributor of the anion layer of cellular surfaces in human, greatly modulates cell to cell repulsion, ligand-receptor interaction, immunogenicity, half-life of circulatory proteins, glomerular filtration, neural plasticity and cognitive development [42-44].

Given that more than half of human proteins are glycosylated [45], considerable interest still exists in identifying the specific carrier proteins to which those glycans attach. Bi-antennary serum *N*-glycans are reported to be carried mainly by IgG, a major serum glycoprotein and an essential part of the immune system, whose structural stability, binding and effector functions are greatly influenced by the type of *N*-glycan attached [46]. Recent inter- and intra- population studies on IgG *N*-glycome profile have increasingly evidenced the association between alteration in IgG *N*-glycome profile and hypertension [13-15].

Particularly, ethnic-based differences in IgG *N*-glycome have been observed with significantly reduced galactosylation and sialylation features in European hypertensive subjects, (but non-significantly in Chinese cases), comparing to their healthy counterparts [13]. In another IgG subclass-specific *N*-glycomic study, Liu JN et al consistently found a marked decrease in galactosylation of IgG1, IgG2, and IgG4, as well as sialylation of IgG1 and IgG2 among northwestern Chinese hypertensive individuals of four different ethnic categories [15]. Further association of IgG glycosylation alterations (loss of galactose and sialic acid, along with addition of bisecting GlcNAc) with blood lipid profile was proposed to cause dyslipidaemia [17], whereas more core fucose and bisecting GlcNAc structures were found to be strongly associated with atherosclerotic plaque [16]. Apart from the role of IgG glycovariants in switching on and off the pro- and anti-inflammatory functions of IgG (and hence its contribution in disease pathogenesis), these reports are suggestive for the possibility that individual variation in IgG *N*-glycan profile may influence the extent of susceptibility to the conditions of diseases. The hyperbranched and hypersialylated glycans most of which have shown ethnic-based differences in the present study are possibly originated from Alpha-1-acid glycoprotein (AGP). It is one of the heavily *N*-glycosylated serum proteins carrying mainly high molecular weight glycans [47] among of which our group has recently succeeded in developing a focused glycoproteomics strategy to directly quantify serum level of AGP carrying a tri-antennary glycoform in multiple cancer types [48].

The Versatility of our comprehensive glycoblotting method is evident because many glycoforms have reproducibly been profiled from diverse biological samples including serum, cell lines, tissues, and cerebrospinal fluid [4-6, 22-25, 40, 41, 49, 50]. Importantly, unlike several prior reports that had measured relative abundances, our systematic strategy could concurrently quantify the absolute concentrations of whole serum glycome and their free sialic acid terminals only from 10 µL of serum aliquot.

In conclusion, our inter-ethnic group glycomics result strongly revealed noticeable variations among the ethnic populations with high mannose, core-fucosylated, multi-antennary and multi- sialylated glycans, as well as the predominant sialic acid (Neu5Ac) demonstrated highest abundance in Ethiopians. The result further indicated some of the glycans that have shown profound expression alteration may not be useful candidates to be biomarkers of various diseases due to their large inter-ethnic and inter-individual variation. Despite the general scope of the present study, we were able to obtain interesting and informative results on the associations between ethnic difference and distinct changes in protein glycosylation which may become helpful for further in-depth investigations in the area. Due to limited samples to comprehensively address the gender and age effects, the current results are preliminary, and thus cannot be generalized to the target populations. In a large-scale study employing these and other ethnic compositions, we further need to investigate the correlation between human glycoforms and various confounding factors including gender and age. Establishing database for healthy human glycome variations among multi-ethnic populations is important as it further improves and accelerates the clinical utility of glycomics and glycoproteomics fields.

4.5. References

1. Clerc F., Reiding K.R., Jansen B.C., Kammeijer G.S., Bondt A., Wuhrer M. Human plasma protein N-glycosylation. Glycoconj J. 2016; 33(3):309-343. doi: 10.1007/s10719-015-9626-2

2. Hart G.W., Copeland R.J. Glycomics hits the big time. Cell 2010; 143(5):672-676. https://doi.org/10.1016/j.cell.20 10.11.008

3. Cummings R.D., Pierce J.M. The challenge and promise of glycomics. Chem Biol. 2014; 21(1):1-15. https://doi.org/10.1016/j.chembiol.2013.12.010

4. Nishimura S., Niikura K., Kurogochi M., Matsushita T., Fumoto M., Hinou H., et al. High-throughput protein glycomics: combined use of chemoselective glycoblotting and MALDI-TOF/TOF mass spectroscopy. Angew Chem Int Ed Engl. 2004; 44(1):91-96. https://doi.org/10.1002/anie.200461685

5. Nishimura S. Toward automated glycan analysis. Adv Carbohydr Chem Biochem. 2011; 65:219-71. https://doi.org/10.1016/B978-0-12-385520-6.00005-4.

6. Gizaw S.T., Koda T., Amano M., Kamimura K., Ohashi T., Hinou H., Nishimura S. A comprehensive glycome profiling of Huntington's disease transgenic mice. Biochim Biophys Acta 2015; 1850(9):1704-1718. https://doi.org/10.1016/j.bbagen.2015.04.006

7. Phuong L.M., Nicolas W., Mark H.G. Challenges related to developing serum-based biomarkers for early ovarian cancer detection. Cancer Prev Res (Phila). 2011; 4(3): 303-306. doi: 10.1158/1940-6207.CAPR-11-0053

8. Eleftherios P.D. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? BMC Med. 2012; 10:87. https://doi.org/10.1186/1741-7015-10-87

9. Duffy M.J. Serum tumor markers in breast cancer: are they of clinical value? Clin Chem. 2006; 52(3):345-351. doi:10.1373/clinchem.2005.059832 PMID: 16410341.

10. Kirwan A., Utratna M., O'Dwyer M.E., Joshi L., Kilcoyne M. Glycosylation-based serum biomarkers for cancer diagnostics and prognostics. Biomed Res Int. 2015; 2015:490531. http://dx.doi.org/10.1155/2015/490531

11. Lu J.P., Knežević A., Wang Y.X., Rudan I., Campbell H., Zou Z.K., et al. Screening novel biomarkers for metabolic syndrome by profiling human plasma N-glycans in Chinese Han and Croatian populations. J Proteome Res. 2011; 10(11):4959-4969. doi: 10.1021/pr2004067

12. Keser T., Gornik I., Vučković F., Selak N., Pavić T., Lukić E., Gudelj I., et al. Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. Diabetologia 2017; 60(12):2352-2360. doi: 10.1007/s00125-017-4426-9

13. Wang Y., Klarić L., Yu X., Thaqi K., Dong J., Novokmet M., et al. The association between glycosylation of immunoglobulin G and hypertension: A multiple ethnic cross-sectional study. Medicine (Baltimore) 2016; 95(17):e3379. doi: 10.1097/MD.0000000000003379

14. Gao Q., Dolikun M., Štambuk J., Wang H., Zhao F., Yiliham N., et al. Immunoglobulin G N-Glycans as potential postgenomic biomarkers for hypertension in the Kazakh population. OMICS 2017; 21(7):380-389. doi: 10.1089/omi.2017.0044

15. Liu J.N., Dolikun M., Štambuk J., Trbojević-Akmačić I., Zhang J., Wang H., et al. The association between subclass-specific IgG Fc N-glycosylation profiles and hypertension in the Uygur, Kazak, Kirgiz, and Tajik populations. J Hum Hypertens. 2018; 32(8-9):555-563. doi: 10.1038/s41371-018-0071-0

16. Menni C., Gudelj I., Macdonald-Dunlop E., Mangino M., Zierer J., Bešić E., et al. Glycosylation profile of immunoglobulin G Is cross-sectionally associated with

cardiovascular disease risk score and subclinical atherosclerosis in two independent cohorts. Circ Res. 2018; 122(11):1555-1564. doi: 10.1161/CIRCRESAHA.117.312174

17. Liu D., Chu X., Wang H., Dong J., Ge S.Q., Zhao Z.Y., et al. The changes of immunoglobulin G N-glycosylation in blood lipids and dyslipidaemia. J Transl Med. 2018; 16(1):235. doi: 10.1186/s12967-018-1616-2

18. Eleftherios P.D. Cancer biomarkers: can we turn recent failures into success? J Natl Cancer Inst. 2010; 102(19):1462-1467. https://doi.org/10.1093/jnci/djq306

19. Knezevic A., Gornik O., Polasek O., Pucic M., Redzic I., Novokmet M., et al. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. Glycobiology 2010; 20(8):959-69. https://doi.org/10.1093/glycob/cwq051

20. Ding N., Nie H., Sun X., Sun W., Qu Y., Liu X., et al. Human serum N-glycan profiles are age and sex dependent. Age Ageing 201; 40(5):568-575. https://doi.org/10.1093/ageing/afr084

21. Yu X., Wang Y., Kristic J. Dong J., Chu X., Ge S., et al. Profiling IgG *N*-glycans as potential biomarker of chronological and biological ages: A community-based study in a Han Chinese population. Medicine (Baltimore) 2016; 95(28):e4112. doi: 10.1097/MD.0000000000004112

22. Kamiyama T., Yokoo H., Furukawa J., Kurogochi M., Togashi T., Miura N., et al. Identification of novel serum biomarkers of hepatocellular carcinoma using glycomic analysis. Hepatology 2013; 57(6):2314-2325. https://doi.org/10.1002/hep.26262

23. Inafuku S., Noda K., Amano M., Ohashi T., Yoshizawa C., Saito W., et al. Alteration of *N*-Glycan profiles in diabetic retinopathy. Invest Ophthalmol Vis Sci. 2015; 56(9):5316-5322. doi: 10.1167/iovs.15-16747

24. Kita Y., Miura Y., Furukawa J., Nakano M., Shinohara Y., Ohno M., et al. Quantitative glycomics of human whole serum glycoprotein based on the standardized protocol for

liberating N-glycans. Mol Cell Proteomics 2007; 6(8):1437-1445. https://doi.org/10.1074/mcp.T600063-MCP200

25. Furukawa J., Shinohara Y., Kuramoto H., Miura Y., Shimaoka H., Kurogochi M., et al. Comprehensive approach to structural and functional glycomics based on chemoselective glycoblotting and sequential tag conversion. Anal Chem. 2008 Feb 15;80(4):1094-1101. doi: 10.1021/ac702124d PMID:18205388.

26. Hirose K., Amano M., Hashimoto R., Lee Y.C., Nishimura S. Insight into glycan diversity and evolutionary lineage based on comparative avio-N-glycomics and sialic acid analysis of 88 egg whites of Gallonaserae. Biochemistry 2011; 50(21):4757-4774. doi: 10.1021/bi101940x

27. Rehan I.F., Ueda K., Mitani T., Amano M., Hinou H., Ohashi., et al. Large-scale glycomics of livestock: Discovery of highly sensitive serum biomarkers indicating an environmental stress affecting immune responses and productivity of Holstein Dairy Cows. J. Agric. Food Chem. 2015; 63(48):10578-10590. doi: 10.1021/acs.jafc.5b04304

28. Chou H.H., Takematsu H., Diaz S., Iber J., Nickerson E., Wright K.L., et al. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. Proc Natl Acad Sci USA. 1998; 95(20):11751-11756.

29. Ohtsubo K., Marth J.D. Glycosylation in cellular mechanisms of health and disease. Cell 2006; 126(5):855-867. https://doi.org/10.1016/j.cell.2006.08.019

30. Pinho S.S., Reis C.A. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer 2015; 15(9):540-555. doi: 10.1038/nrc3982

31. Clerc F., Reiding K.R., Jansen B.C., Kammeijer G.S., Bondt A., Wuhrer M. Human plasma protein N-glycosylation. Glycoconj J. 2016; 33(3):309-343. doi: 10.1007/s10719-015-9626-2

32. Anderson N.L., Anderson N.G. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002; 1(11):845-867. https://doi.org/10.1074/mcp.R200007-MCP200.

33. Knezević A., Polasek O., Gornik O., Rudan I., Campbell H., Hayward C., et al. Variability, heritability and environmental determinants of human plasma N-glycome. J Proteome Res. 2009; 8(2):694-701. doi: 10.1021/pr800737u

34. Pucić M., Knezević A., Vidic J., Adamczyk B., Novokmet M., Polasek O., et al. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Mol Cell Proteomics 2011; 10(10):M111.010090. https://doi.org/10.1074/mcp.M111.010090

35. Bieberich E. Synthesis, processing, and function of N-glycans in N-glycoproteins. Adv Neurobiol. 2014; 9:47-70. doi: 10.1007/978-1-4939-1154-7_3

36. Menni C., Keser T., Mangino M., Bell J.T., Erte I., Akmačić I., et al. Glycosylation of immunoglobulin G: Role of genetic and epigenetic influences. PLoS One 2013; 8(12):e82558. https://doi.org/10.1371/journal.pone.0082558

37. Nairn A.V., York W.S., Harris K., Hall E.M., Pierce J.M., Moremen K.W. Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. J Biol Chem. 2008; 283(25):17298-17313. doi: 10.1074/jbc.M801964200

38. Tahmasebi H., Trajcevski K., Higgins V., Adeli K. Influence of ethnicity on population reference values for biochemical markers. Crit Rev Clin Lab Sci. 2018; 55(5):359-375. https://doi.org/10.1080/10408363.2018.1476455

39. Ellman N., Keswell D., Collins M., Tootla M., Goedecke J.H. Ethnic differences in the association between lipid metabolism genes and lipid levels in black and white South African women. Atherosclerosis 201; 240(2): 311-317. https://doi.org/10.1016/j.atherosclerosis.2015.03.027

40. Hatakeyama S., Amano M., Tobisawa Y., Yoneyama T., Tsuchiya N., Habuchi T., et al. Serum N-glycan alteration associated with renal cell carcinoma detected by high-throughput glycan analysis. J Urol. 2014; 191(3):805-813. https://doi.org/10.1016/j.juro.2013.10.052

41. Ishibashi Y., Tobisawa Y., Hatakeyama S., Ohashi T., Tanaka M., Narita S., et al. Serum Tri- and tetra-antennary N-glycan is a potential predictive biomarker for castration-resistant prostate cancer. Prostate 2014; 74(15):1521-1529. https://doi.org/10.1002/pros.22869

42. Wang B. Molecular mechanism underlying sialic acid as an essential nutrient for brain development and cognition. Adv Nutr. 2012; 3(3):465S-472S. https://doi.org/10.3945/an.112.001875

43. Varki A. Sialic acids in human health and disease. Trends Mol Med. 2008; 14(8):351-360. https://doi.org/10.1016/j.molmed.2008.06.002

44. Varki A., Gagneux P. Multifarious roles of sialic acids in immunity. Ann N Y Acad Sci. 2012; 1253:16-36. https://doi.org/10.1111/j.1749-6632.2012.06517.x

45. Apweiler R., Hermjakob H., Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1999; 1473(1):4-8. https://doi.org/10.1016/S0304-4165(99)00165-8

46. Arnold J.N., Wormald M.R., Sim R.B., Rudd P.M., Dwek R.A. The Impact of glycosylation on the biological function and structure of human immunoglobulins. Annu Rev Immunol. 2007; 25:21-50. doi: 10.1146/annurev.immunol.25.022106.141702

47. Kevin D.S., Jennifer B., Gerardine M.S., Anthony M.M. Alpha-1-Acid Glycoprotein (AGP) as a potential biomarker for breast cancer. InTechOpen 2012. http://dx.doi.org/10.5772/48177.

48. Yogesh K.V., Kamiyama T., Ohyama C., Yoneyama T., Nouso K., Kimura S. Synthetic glycopeptides as a designated standard in focused glycoproteomics to discover serum cancer biomarkers. Med. Chem. Commun. 2018; 9:1351-1358. doi: 10.1039/C8MD00162F.

49. Terashima M., Amano M., Onodera T., Nishimura S., Iwasaki N. Quantitative glycomics monitoring of induced pluripotent- and embryonic stem cells during neuronal differentiation. Stem Cell Res. 2014; 13(3):454-464. https://doi.org/10.1016/j.scr.2014.10.006

50. Gizaw S.T., Ohashi T., Tanaka M., Hinou H., Nishimura S. Glycoblotting method allows for rapid and efficient glycome profiling of human Alzheimer's disease brain, serum and cerebrospinal fluid towards potential biomarker discovery. Biochim Biophys Acta 2016; 1860(8):1716-1727. https://doi.org/10.1016/j.bbagen.2016.03.009

Chapter 5

Concluding Remarks

Despite significant advances in cancer health care system, breast cancer has continued being the most prevalent and deadliest cancer in women. Since current diagnostic methods have invasive nature and suffer from lack of sensitivity and specificity to predict the disease at its early stage, there is a pressing need for novel serological biomarker discovery and thus to improve breast cancer diagnosis and treatment. Apart from the cancer genome, deciphering posttranslational protein glycosylation alterations has been of critical importance as a promising target for discovering novel diagnostic and therapeutic agents.

As emphasized in chapter 2 and chapter 3, I have reproducibly quantified *N*-glycans from whole serum and from IgG fraction of breast cancer patients and matched controls by using a glycoblotting-assisted MALDI-TOF/MS-based analysis, aiming at identification of new biomarkers for early detection of breast cancer. In this comprehensive work, I have found glycan structures abundantly expressed in serum of the breast cancer patients, with strong diagnostic potential primarily demonstrated in the early stage (stages I and II). The IgG-focused *N*-glycomics result further revealed patient associated increase in IgG fucosylation and agalactosylation that could specifically discriminate stage-II patients from controls. These particular alterations in IgG glycoforms have been reported in association with a decrease in immunosuppressive effect of IgG, allowing the tumor cells to escape from the immune system.
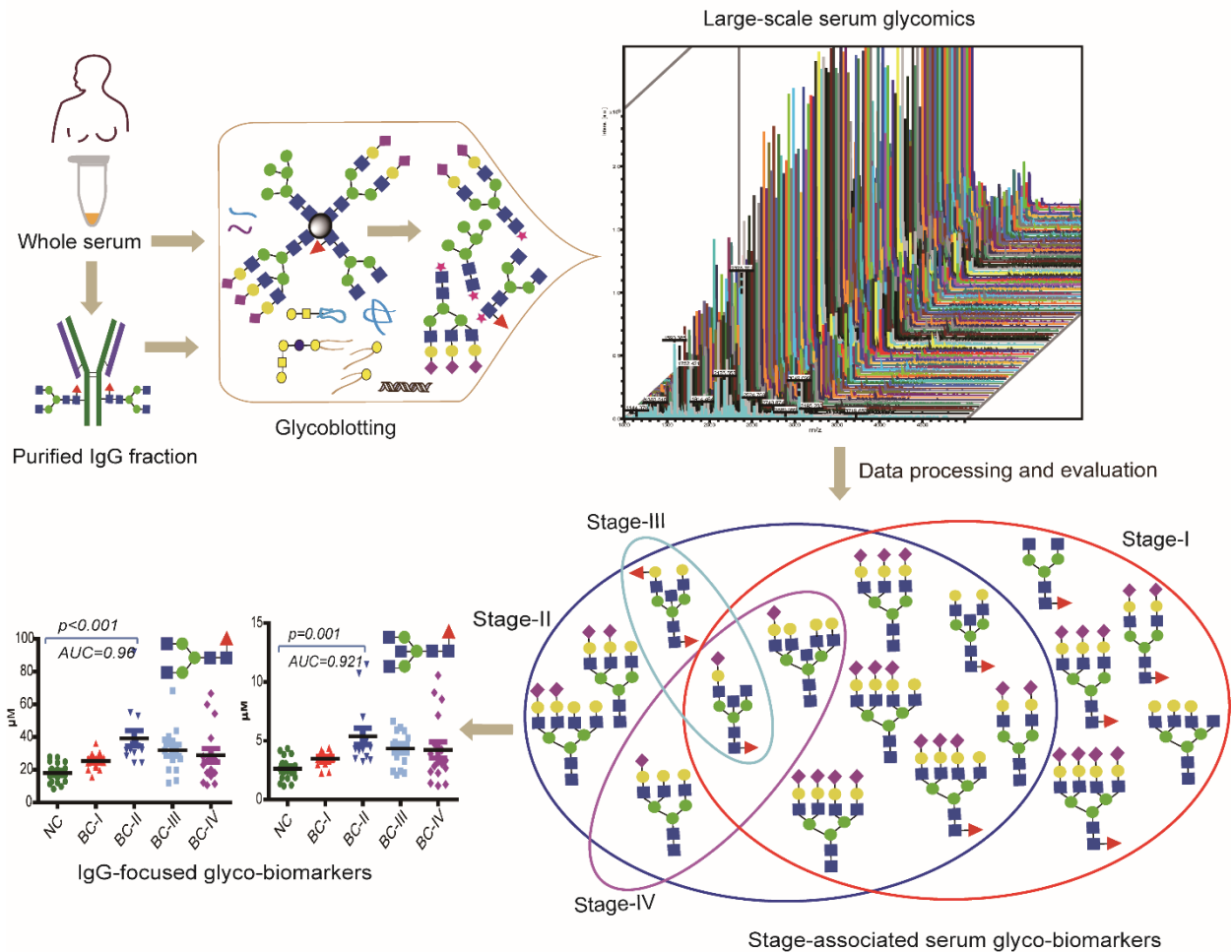
Breast cancer stage specific serum and IgG *N*-glyco-biomarker candidates identified by glycoblotting-assisted MALDI-TOF/MS analysis

Using glycobloting-assisted MALDI-TOF/MS- and HPLC-based quantitative approaches, I further extended my research (discussed in chapter 4) to investigate inter-ethnic serum *N*-glycome and sialic acid variations among US origin control, Japanese, Indian, and Ethiopian healthy volunteers in association with the identified glycan biomarkers, aiming at noting the influence of confounding factors like ethnic variation in glyco-biomarker discovery research. The results provide very informative evidence for the great impact of ethnic difference on the human serum *N*-glycome and sialic acid profile. Marked up-regulation of several glycans and free Neu5Ac was observed mainly in the Ethiopian ethnic group compared

to the other ethnicities. Surprisingly, some of the *N*-glycans greatly elevated in the healthy Ethiopian subjects have been identified as sensitive serum biomarkers of various cancers, suggesting to critically follow ethnic matching during biomarker discovery researches and in this context, to reconsider previously suggested glyco-biomarkers.

In conclusion, apart from addressing serum and IgG *N*-glycosylation signatures of native black population for the first time, this study revealed novel candidate glyco-biomarkers markedly associated with early stage breast cancer and indicated the influence of ethnic variation on drastically affecting the human glycome profile.

Finally, I clearly proved the potential of serum and IgG *N*-glycomics profiling for non-invasive and early stage cancer biomarker discovery, as well as the need to strictly consider ethnic matching during population based glyco-biomarker discovery research.

Acknowledgements