Proteomics and *N*-glycome analyses reveal insights into the pathogenesis of the molecular subtypes of breast cancer

Ling Yen Lee

A thesis submitted for the degree of

Doctor of Philosophy

Department of Chemistry and Biomolecular Sciences

Macquarie University, Sydney

February 2015

TABLE OF CONTENT

SUMMARY 4

DECLARATION 6

ACKNOWLEDGEMENTS 7

PUBLICATION AND CONTRIBUTION STATEMENT 8

ABBREVIATIONS 9

CHAPTER 1 13

INTRODUCTION 13

1.1 Breast cancer 14
1.1.1 Breast cancer incidence and mortality 14
1.1.2 Anatomy of the breast 14
1.1.3 Types of breast cancer 15
1.1.4 Molecular subtypes of breast cancer 15

1.2 Proteomics 18
1.2.1 Proteomics – a brief overview 18
1.2.2 Sample preparation 19
1.2.2.1 Gel-based separation of proteins 21
1.2.2.2 Gel-free separation of peptides 22
1.2.2.3 Lectin affinity chromatography 23
1.2.3 LC-MS/MS-based protein detection of peptides 25
1.2.4 Label-assisted and label-free mass spectrometry-based protein quantitation 26
1.2.4.1 Label-assisted methods 26
1.2.4.2 Label-free quantitation 28
1.2.5 Bioinformatics tool for protein identification 30
1.2.6 Proteomics in breast cancer biomarker discovery 33
1.2.7 Functional analysis of proteomics – pathway analysis and interaction networks 38

1.3 Protein glycosylation 42
1.3.1 Protein glycosylation – a brief overview 42
1.3.1.1 Biosynthesis and the endoplasmic reticulum (ER)-Golgi secretory pathway 45
1.3.2 Characterization of protein *N*-glycosylation 47
1.3.2.1 LC-MS/MS based structural analysis of the N-glycome 49
1.3.3 Protein *N*-glycosylation changes in breast cancer 54
1.3.3.1 Sialylation 56
1.3.3.2 Fucosylation 58

1.3.3.3 Branching and bisecting GlcNAc 60
1.3.3.4 High mannose N-glycans 61

1.4 Aims of the thesis 62

Publication I - Cell surface protein glycosylation in cancer (Review) 64

CHAPTER 2 87

FUNCTIONAL ANALYSIS OF PROTEOME CHANGES IN BREAST CANCER 87

2.1 Introduction 88

2.2 Materials and methods 90
2.2.1 Cell cultures and sample preparation under serum-free conditions 90
2.2.2 Gel electrophoresis of subcellular proteomes and in-gel digestion 91
2.2.3 LC-MS/MS-based proteomics 92
2.2.4 Protein identification 93
2.2.5 Label-free quantitation using normalized spectral abundance factor 93
2.2.6 Statistical analysis and bioinformatics 94

2.3 Results 94
2.3.1 Optimization of cultured cells for proteomics analysis 94
2.3.2 Identification of secreted and membrane proteins in HMEC, MCF7, SKBR3 and MDA231 96
2.3.3 Global biological and functional analyses of secreted and membrane proteins 101
2.3.4 Comparative analysis of secreted and membrane proteins differentially expressed between normal and breast cancer cells 103
2.3.5 Differential expression of subtype-specific proteins in breast cancer 113
2.3.6 Proteomics-based clustering of tumorigenic and breast cancer subtypes 117

2.4 Discussion 118

2.5 Conclusion 125

CHAPTER 3 127

STRUCTURAL ANALYSIS OF *N*-GLYCOME CHANGES IN BREAST CANCER 127

Part 1 129

Publication II - Comprehensive *N*-glycome profiling of cultured human epithelial breast cells identifies unique secretome *N*-glycosylation signatures enabling tumorigenic subtype classification 129

Part 2 143

***N*-glycome analysis of membrane proteins from a panel of breast epithelial cell lines 143**

3.1 Introduction **144**

3.2 Materials and Methods **144**

3.3 Results **145**

3.3.1 Comparative analyses of *N*-glycan sub-structures on membrane proteins from tumorigenic and non-tumorigenic breast cells 145

3.3.2 Global comparison between secreted and membrane *N*-glycans in the panel of breast epithelial cell lines 152

3.4 Discussion **154**

3.5 Conclusion **156**

CHAPTER 4 **157**

Publication III - Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants **157**

CHAPTER 5 **171**

Publication IV - An optimized approach for enrichment of glycoproteins using native multi-lectin affinity chromatography **171**

CHAPTER 6 **181**

GENERAL DISCUSSION **181**

6.1 Thesis summary **182**

6.2 Future directions **191**

6.3 Conclusion **193**

APPENDICES **195**

REFERENCES **343**

Summary

Breast cancer is the most common type of cancer affecting women worldwide. It is the leading cause of cancer-related death among females and its incidence rate is rising sharply. Significant molecular heterogeneity exists within breast cancer, which consequently leads to the formation of multiple molecular subtypes of the disease. In an effort to address the challenges associated with establishing reliable markers predictive of breast cancer and to develop effective drug therapies, the major aim of this thesis is to achieve an improved understanding of the molecular mechanisms and pathway deregulation in the breast cancer pathology.

The studies described in this thesis applied high throughput proteomics and glycomics analyses, which allowed parallel global protein and *N*-glycan comparisons, respectively, to be made to define discriminatory patterns that correlated with the molecular heterogeneity observed in breast cancer. Specifically, comparative proteomics and glycomics of secreted and membrane fractions from a panel of breast cancer cell lines corresponding to three common breast cancer subtypes including luminal A, HER2-enriched and basal B subtypes, were performed using non-tumorigenic human mammary epithelial cells (HMEC) as a normal healthy reference. The distinctive subcellular proteome and glycome signatures unique to the individual cancer subtypes were functionally evaluated by utilizing a range of bioinformatics-assisted pathway analysis tools to gain insights into regulatory mechanisms underlying the normal and tumorigenic cellular processes.

The combination of structural and functional proteomics yielded consistent molecular themes involved in the pathogenesis of breast cancer. In addition, distinctive molecular features associated with each subtypes were present. In the first study of its kind, comprehensive analysis of the secreted *N*-glycome of a panel of breast epithelial cells investigated the involvement of protein *N*-glycosylation in breast cancer. The causative and/or effector roles of aberrant *N*-

glycosylation in breast tumorigenesis were evident as strongly supported by the presence of tumor-promoting *N*-glycan determinants in the secreted and membrane fractions of breast cancer cells. Significantly, unique secretome *N*-glycosylation signatures enabled breast cancer subtype classification.

Subcellular-specific *N*-glycosylation was found to be a universal cellular feature not only limited to epithelial breast cancer cells and was mechanistically explained by the differential solvent accessibility to the asparagines residues forming the *N*-glycosylation sites. Having mapped this relationship between spatial accessibility and *N*-glycan processing of glycoproteins is important to allow us to understand the expression and (de)regulation of glycoepitopes in breast cancer.

In recognizing the importance of investigating intact glycopeptides to integrate the information from the obtained breast cancer cellular proteome and glycome and obtain site-specific information of protein *N*-glycosylation of breast cancer cells in future work, a multi-lectin affinity chromatography platform for cancer-specific glycoprotein enrichment directly from whole cell lysates was developed and optimized, which will serve as a useful tool in glycoproteomics.

In conclusion, this thesis provides the most detailed picture of the proteome and *N*-glycome deregulation in multiple breast cancer subtypes to date, which yields valuable insight into the multiple mechanisms associated with the pathophysiological changes in breast cancer. This molecular insight forms an important knowledge platform from which the emerging field of glycoproteomics promise to yield an even higher definition of the tumor-specific protein modifications and, as a consequence, eventually allow us to develop targeted molecular therapeutics and diagnostics tools to benefit the growing number of women affected by the disease.

Declaration

I certify that the work presented in this thesis titled "Proteomics and *N*-glycome analyses reveal insights into the pathogenesis of the molecular subtypes of breast cancer" has not previously been submitted, in either whole or part, for the purposes of obtaining any other degree to any other university or institution other than Macquarie University. Unless otherwise acknowledged, this work was carried out by the author.

Ling Yen Lee

41328655

20 February 2015

Acknowledgements

The thesis could not have been completed without the help and support of my family, friends and colleagues. First of all, I would like to thank my "original" supervisors Professor Hancock and Dr Susan Fanayan for providing me with the great opportunity to work on this project. I am grateful to Susan for her advice, encouragement and support in many ways during challenging times in the past four years. My appreciation goes to Prof Hancock and his group at Northeastern University during my 2-month lab visit, especially to Dr Marina Hincapie, who has provided much guidance on the work there. It has been a great pleasure to be able to work with my current supervisor, Dr Morten Thaysen-Andersen, who has contributed much of his time and valuable knowledge to this thesis. I look forward to work with him in the next project. I would also like to thank my associate supervisors, Prof Mark Baker and Prof Nicki Packer for giving me the opportunity to join their groups' meetings and having their constructive feedback on my work. Thanks to the only member in my group, Manveen as well as students and postdocs from both Baker's and Glyco (Nicki's) groups for sharing their "proteo-" and "glyco-knowledge". It has also been a pleasure to work briefly with Dr Fei Liu who shared her insights and knowledge on cancers.

I would like to acknowledge the technical expertise given by APAF staff including Dylan, Dana and Ardeshir; the assistance provided by Nicole and Debra for fluorescence microscopy work; administrative support from Catherine and Michelle; academic support from Prof Joanne Jamie for designating a writing room to stressful final year students like me. I must also thank Macquarie University for providing me the scholarship (MQRES) to pursue my research.

Most importantly, I would like to thank my parents for their support and understanding.

Publication and Contribution Statement

This thesis contains material that has been published, and the percentage contribution in each of the publications by the candidate (Lee, L.Y or Lee, L.) are as follows:

Publication I (Co-first author)
Christiansen, M. N., Chik, J., **Lee, L.,** Anugraham, M., Abrahams, J. L., Packer, N. H. (2013) Cell surface protein glycosylation in cancer. *Proteomics*. 14: 525-546.

Concept - 15%; Writing - 20%; **Total (Average) - 17.5%**

Publication II (First author)
Lee, L.Y., Thaysen-Andersen, M., Baker, M. S., Packer, N. H., Hancock, W. S. and 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. (In press)

Concept - 50%; Data collection - 100%; Analysis - 80%; Writing - 50%; **Total (Average) - 70%**

Publication III (First author)
Lee, L.Y., Lin, Chi-Hung., Fanayan, S., Packer, N.H. and Thaysen-Andersen, M. Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants. *Frontiers in Immunology*, 2014. **5**.

Concept - 40%; Data collection - 80%; Analysis - 80%; Writing - 40%; **Total (Average) - 60**%

Publication IV (First author)
Lee, L. Y., Hincapie, M., Packer, N., Baker, M. S., Hancock, W. S. and Fanayan, S. (2012) An optimized approach for enrichment of glycoproteins from cell culture lysates using native multilectin affinity chromatography. *J Sep Sci,* 35**,** 2445-52.

Concept - 40%; Data collection - 100%; Analysis - 80%; Writing - 50%; **Total (Average) - 68%**

Abbreviations

1DE	One-dimensional electrophoresis
2AB	2-aminobenzamide
2DE	Two-dimensional electrophoresis
ACN	Acetonitrile
ADCC	Antibody-dependent cellular cytotoxicity
Asn	Asparagine
BM	Basement membrane
C18	Octadecylsilyl
CID	Collision induced dissociation
CM	Conditioned media
Con A	Concanavalin A
DCIS	Ductal carcinoma *in situ*
DIGE	Differential in-gel electrophoresis
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EIC	Extracted ion chromatogram
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
ER	Endoplasmic reticulum
ESI	Electrospray ionization
FBS	Fetal bovine serum
FDR	False discovery rate
Fuc	Fucose
FUT	Fucosyltransferases
GAG	Glycosaminoglycans
Gal	Galactose
GalNAc	*N*-acetyl galactosamine
Glc	Glucose
GlcNAc	N-acetyl glucosamine
GnT-III	*N*-acetylglucosamintransferase III
GnT-V	*N*-acetylglucosaminyltransferase V
GO	Gene ontology
GPCR	G protein-coupled receptor
GPM	Global Proteome Machine

HER2	Human epidermal growth receptor 2
HILIC	Hydrophilic-interaction liquid chromatoigraphy
HMEC	Human mammary epithelial cells
HPLC	High performance liquid chromatography
ICAT	Isotope-coded affinity tags
IDC	Invasive ductal carcinoma
IEF	Isoelectric focusing
IHC	Immunohistochemistry
ILC	Invasive lobular carcinoma
iTRAQ	Isobaric tags for relative and absolute quantitation
JAC	Jacalin
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LacdiNAc	*N*,*N*'-diacetyllactosamine
LacNAc	*N*-Acetyl-D-lactosamine
$Le^{X/Y}$	Lewis X/Y
LC	Liquid chromatography
LCA	Lens culinaris agglutinin
LDS	Lithium dodecyl sulfate
MALDI	Matrix-assisted laser desorption/ionization
Man	Mannose
MGAT	Mannoside *N*-acetylglucosaminyltransferase
M-LAC	Multi-lectin affinity chromatography
MudPIT	Multidimensional protein identification technology
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
NaCl	Sodium Chloride
NAF	Nipple aspirate fluid
NCBI	National Centre for Biotechnology Information
NeuAc	*N*-acetylneuraminic acid
NeuGc	*N*-glycolylneuraminic acid
NOS	Not otherwise specified
NSAF	Normalized spectral abundance factor
OST	Oligosaccharyltransferase
PCA	Principal component analysis
PGC	Porous graphitized carbon
PBS	Phosphate buffered saline
PHA-L	Phaseolus vulgaris leucoagglutinin

PNA	Peanut agglutinin
PNGase	Peptide -*N*-glycosidase
PPD	Plasma Proteome Database
PPM	Parts per million
PR	Progesterone receptor
PSM	Peptide spectrum match
PTM	Post-translational modification
RP	Reversed phase
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELDI	Surface-enhanced laser desorption/ionization
SFM	Serum-free media
SILAC	Stable isotope labeling by/with amino acids in cell culture
SLAC	Serial lectin affinity chromatography
$SLe^{x/a}$	Sialyl Lewis x/a antigen
SNA	Sambucus nigra lectin
SpC	Spectral count
ST3 Gal	β-galactoside alpha-2,3-sialyltransferase
ST6 Gal	β-Beta galactoside alpha 2,6 sialyltransferase
TIF	Tissue interstitial fluid
TMT	Tandem mass tags
TOF	Time of flight
WGA	Wheat germ agglutinin

There are no facts, only interpretations.

Friedrich Nietzsche

CHAPTER 1

INTRODUCTION

1.1 Breast cancer

1.1.1 Breast cancer incidence and mortality

With one in eight women predicted to develop breast cancer in their lifetime, breast cancer is the most commonly diagnosed cancer affecting Australian women [1]. In 2011, the disease accounted for 15.6% of all female cancer deaths, making it the second leading cause of cancer-related deaths in Australian women. The incidence rate is rising steadily with 13,567 new cases diagnosed in Australia in 2008. Approximately 15,270 new cases will be detected in Australia in 2014 and this figure is estimated to increase by about 13% by 2020 [2].

1.1.2 Anatomy of the breast

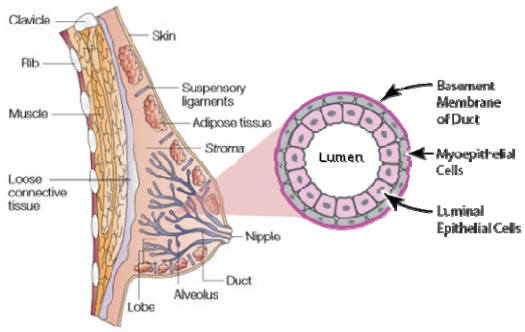


Figure 1.1 Anatomy of the mammary gland. (Adapted from Ali *et al.* 2002 [3] and http://www.breastcancer.com)

The mammary gland is a complex tissue composed of a series of milk-producing lobes connected to the lactiferous ducts that converge near the nipple. Together, these form a branching ductal network that is embedded in a mass of fibrous connective tissues, adipose tissues and extracellular matrix collectively known as the mammary stroma. The breast is thus comprised of a diverse array of cell types, although the majority belong to two types of differentiated epithelial cells found within the epithelium ductal network – an inner layer of polarized luminal epithelial cells facing the ductal lumen surrounded by an outer layer of myoepithelial cells (Figure 1.1). The myoepithelial cells affect the differentiation, polarity, proliferation and migration of the adjoining luminal epithelial cells [4]. In addition, they secrete major structural proteins such as laminin and

collagen that contribute to the formation of the basement membrane, which is a physical barrier separating the ductal epithelium structures and the stroma [5].

1.1.3 Types of breast cancer

Breast cancer is a highly diverse disease with more than a dozen histopathological variants defined by the World Health Organization (WHO) [6]. Breast tumors that are still confined within the ducts or lobules are known as ductal or lobular carcinoma *in situ*, respectively. When lesions have breached the basement membrane or spread to the lymph nodes, they are generally known as invasive breast cancer. The majority of invasive breast cancers are of epithelial origin known as carcinomas. Sarcomas which arise from the stroma in the breast are rare. The two dominant types of invasive breast cancers are invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). Up to 80% of breast lesions are IDC, often described as not otherwise specified (NOS), making it the most common form of breast cancer while 10% to 15% are represented by ILC. Both of these carcinomas show distinguishable molecular and genetic features [7, 8]. The remaining carcinomas are rare and include the inflammatory breast cancer, Paget's disease and IDC variants such as tubular, medullary, mucinous and papillary carcinomas [9].

1.1.4 Molecular subtypes of breast cancer

Following initial diagnosis, breast cancer is often categorized according to established classification schemes to determine the prognosis of the disease and more importantly, to aid in selection of the most appropriate treatment for the individual breast cancer patient [10]. The classification schemes are heavily based on pathological examination of breast tumors which group them into histopathological types, tumor grades and stages. In addition, breast tumors are assessed for expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth receptor 2 (HER2). The hormone receptors (ER and PR) are assayed by

immunohistochemistry (IHC) while HER2 status is confirmed by fluorescent *in situ* hybridization.

Table 1.1 Summary of characteristics of breast cancer subtypes.

Molecular subtype	Gene expression[1]	Prevalence	Relative Survival	Therapeutic options
Luminal A	ER+, PR+, HER2-, low Ki-67	~40%	Longest	Hormonal therapy
Luminal B	ER+ and/or PR+, HER2-, high Ki-67	~20%	Decreased	Hormonal therapy
HER2-enriched	ER-, PR-, HER2+	15-20%	Decreased	Trastuzumab
Basal-like	ER-, PR-, HER2-, EGFR+, CK5/CK17		Shortest	Surgery/Chemo-therapy
Claudin-low	ER-, PR-, HER2-, claudin genes	10-15%	Decreased	Surgery/Chemo-therapy

[1]ER+/-, ER positive/negative; PR+/-, PR positive/negative, HER2+/-, HER2 positive/negative, EGFR+, EGFR positive

In recent years, there has been an increased emphasis to use molecular approaches to improve breast cancer diagnostics due to the limited clinical utility of the conventional classification schemes. The receptor status of ER, PR and HER2 has prognostic value in predicting efficacy of targeted hormone and cytotoxic drug treatment against these receptors. However, the predictive value has insufficient specificity and sensitivity, and is inadequate for newly developed targeted therapies. Moreover, these traditional classification schemes alone are unable to capture the genetic diversity that is invariably present within the largest IDCs NOS group. The seminal work by Perou *et al*, revealed that distinctive molecular features associated with IDC-classified breast tumors, such as differential expression of the three receptors (ER, PR and HER2) and proliferative genes, such as Ki-67, could be used to segregate tumors into various intrinsic subtypes [7]. The four subtypes identified were known as luminal A (ER/PR positive, HER2 negative), luminal B (ER positive and/or PR positive, HER2 positive), HER2-enriched (ER/PR negative, HER2 positive) and basal-like (ER/PR/HER2 negative, also known as triple-negative) (Table 1.1). Subsequent gene profile-centric investigations have reproducibly observed similar trends showing that these key molecular features are conserved among breast cancers [11-14].

More recently, a novel subtype known as claudin-low that shares some characteristics of the basal-like subtype has been described [15].

Molecular subtyping has great prognostic value since each subtype has unique survival outcomes. The division of luminal subtypes into luminal A and B is of clinical interest because despite both groups being positive for ER, they have different prognosis with luminal A having a better prognosis than luminal B. Similarly, significantly worse prognosis is observed in HER2-enriched and basal-like subtypes. Subtype classification also influences the therapeutic options and serves to predict treatment response, in particular for tumors that are highly responsive or non-responsive to hormonal or targeted drug therapies. Patients with ER+ breast cancer are treated with drugs such as Tamoxifen that blocks ER activity while those with HER2+ benefit from anti-HER2 drugs such as Trastuzumab. In contrast, patients with breast tumors that lack the three receptors, i.e. ER, PR and HER2, are not expected to respond to these targeted treatments and may be more suitable to undergo surgery and chemotherapy.

Although the gene expression-based stratification of breast cancer has led to better insights into the biological diversity of breast cancers, the overall underlying molecular mechanism(s) in breast tumorigenesis, including those associated with the more aggressive basal-like breast tumors, are still poorly understood. It is widely accepted that gene transcription does not necessary correlate with the expression of gene products (i.e. proteins), which are the key mediators of cellular processes. Various spliced protein variants are known to exist for a single gene, and proteins undergo a wide range of post-translational modifications (PTMs) such as glycosylation, phosphorylation or methylation which dramatically can affect their biological functions.

Studying protein expression and the associated PTMs is therefore crucial to understand how these molecular effectors of function regulate key biological events during malignant breast transformation.

1.2 Proteomics

1.2.1 Proteomics – a brief overview

The proteome is defined as the entire complement of proteins, produced by the genome of a particular cell or tissue, at a specified time, space and condition [16]. Proteomics encompasses the structural analysis of proteins involving aspects such as protein identification, protein abundance measurements as well as both qualitative and quantitative characterization of any PTMs associated with the expressed proteome. The development of conjugated liquid chromatography (LC) and advanced mass spectrometry (MS) technologies have been pivotal in aiding the rapid advancement of proteomics by allowing for high throughput and ultra-sensitive protein identification and quantitation. The LC interfaced with tandem MS (MS/MS) is currently the mainstay technology for proteomics-based studies, which are undertaken via either a global or targeted approach of the proteome being investigated. Global proteomic analysis entails the identification and quantitation of all proteins in a given sample, in contrast to targeted proteomics, which investigate a relatively small group of proteins under various conditions [17]. By combining appropriate upstream methodologies such as sample preparation and fractionation/enrichment and downstream computational tools, proteomics is a powerful analytical approach to allow parallel global proteome profiling and identify distinct protein expression patterns in tumorigenesis, so as to discover cancer biomarkers and gain insights into molecular perturbations. A typical global proteomics workflow to find differences between the proteomes of a disease and healthy reference (control) samples is illustrated in Figure 1.2. The following sections will describe the global proteomics workflow that incorporates various

proteomics technologies used for proteome profiling and their applications in breast cancer research.

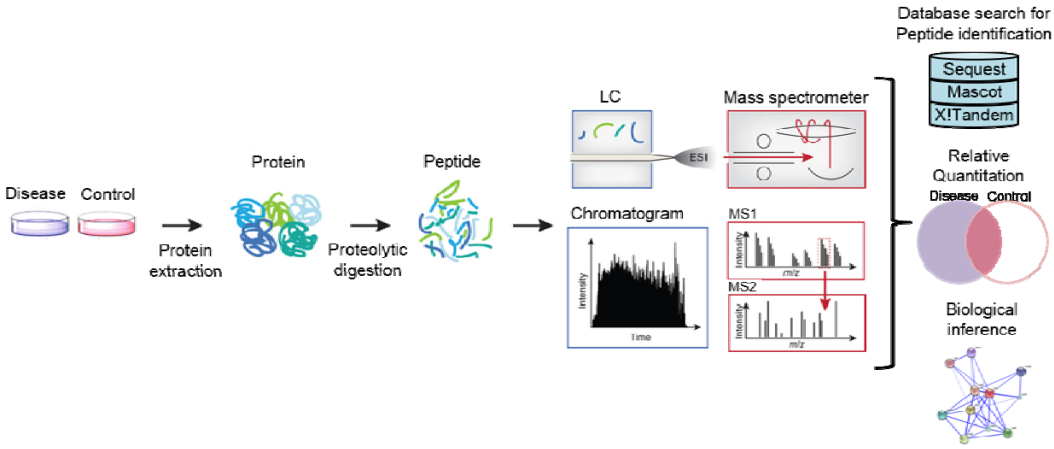


Figure 1.2 A Schematic representation of a typical proteomics workflow involving the comparative investigation of disease and healthy control samples. In summary, the samples are digested and applied to LC-MS/MS after which the data are analysed for proteins identification, relative quantification and biological significance. (Modified from Meissner and Mann 2014 [18])

1.2.2 Sample preparation

In the last two decades, technological advancements in MS-based proteomics have facilitated the rapid, accurate and highly sensitive analysis of proteins at a relatively low cost. However, despite these advances, proteomics analyses remain challenging due to the complexity of samples investigated and the extensive dynamic molar range of proteins present within the proteome. To overcome these issues, appropriate implementation of upstream methodologies are needed. This includes sample enrichment or separation/fractionation steps to enhance the dynamic range and depth of analysis [19]. In plasma or serum samples displaying an extreme dynamic range of protein concentration, depletion strategies are often employed to remove the high abundance proteins in order to detect those of lower abundance [20]. Protein glycosylation is a biologically significant PTM with more than 50% of proteins considered to be glycosylated [21].

In recent years, the field of glycoproteomics have emerged to focus on the analysis of this major class of biomolecules (glycoproteins) [22], which have the potential to serve as cancer biomarkers. Several separation tools have been developed to enrich for glycoproteins to decrease

sample complexity thus facilitating more efficient analysis of glycoproteins. Enrichment technologies include the use of lectins [23], boronic acid [24] and hydrazine chemistry [25] to capture glycoproteins from complex mixtures. Lectins bind to glycoproteins via recognition of specific glycoepitopes attached to the proteins. Such property is exploited to identify altered glycoforms on tumor samples, for example, a two-step fractionation strategy combining serum depletion with the enrichment of glycoproteins using multi-lectin affinity chromatography (M-LAC) has been developed for biomarker discovery studies [26]. M-LAC is the subject of investigation presented in Chapter 5. Boronic acid form reversible covalent complexes with the *cis*-diols present in monosaccharides of glycan residues on proteins in an alkaline/acidic aqueous solution and have been shown to successfully isolate and identify glycoproteins from complex protein mixtures [27, 28]. Similarly targeting the *cis*-diols found in glycosylated proteins, the hydrazide chemistry-based selectively enrichment of glycoproteins has been widely applied in proteomics research [29-31] since it was first developed by Zhang *et al* [25]. Although this method may be easily integrated into LC-MS/MS workflows, the irreversible glycoprotein attachment to hydrazide beads limit the downstream analysis of the intact glycoprotein relative to the non-covalent and reversible covalent lectin and boronic acid enrichment strategies, respectively.

Very often, reduction in the sample complexity can be easily achieved by sample separation by gel-based or gel-free approaches. Gel-based methods are capable of separating proteins, whereas gel-free methods in proteomics are typically used to separate peptides after digestion and may be conjugated directed to the LC-MS/MS analysis as described below. Such strategies are now a standard step incorporated prior to MS/MS analysis to enhance the protein identification.

1.2.2.1 Gel-based separation of proteins

By exploiting the physicochemical properties of proteins, one-dimensional electrophoresis (1DE) on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) resolves proteins according to their molecular weight (MW) while two-dimensional electrophoresis (2DE) separates proteins first based on isoelectric focusing (IEF) followed by MW on SDS-PAGE, thus allowing detection of protein charge-isoforms. The gel is then fixed, stained and imaged to visualize the protein bands or spots where they are subsequently excised and proteolytically digested into peptides. The enzymatic release of peptides from proteins for MS characterization is commonly referred to as "bottom-up" protein identification, an approach which fundamentally defines shotgun proteomics [32]. This contrasts the "top-down" strategy where intact proteins or large protein fragments are directly analyzed (Figure 1.3). The "top down" strategy preserves the biological organization within the protein but due to the associated analytical challenges including, but not limited to, limited throughput and sensitivity is best suited for the emerging field of mechanistic biology for studying single proteins or simple protein mixtures. The "bottom up" method, therefore, remains the preferred approach for most current proteomics research [33].

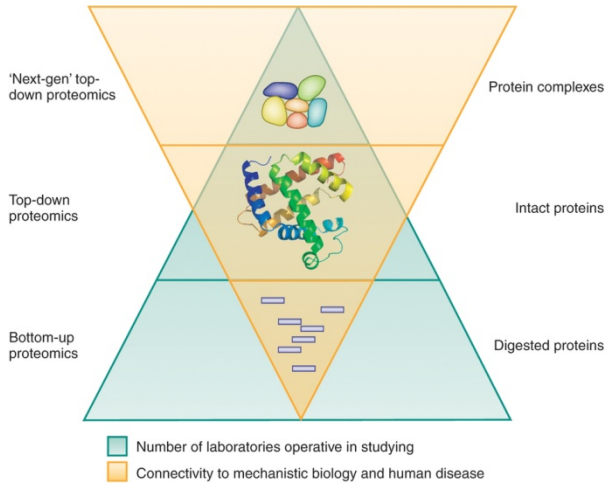


Figure 1.3 The "top-down" vs "bottom-up" approach in proteomics research. (Adapted from Compton *et al*, 2012 [33])

A related alternative to 2DE gel based proteomics is known as differential in-gel electrophoresis (DIGE). DIGE was introduced to improve the reproducibility of 2DE, mainly as a result of gel to gel variations [34]. DIGE allows the labelling of proteins with up to three different fluorescent cyanine dyes and pooling of the labelled proteins, followed by 2D separation to quantitatively compare for differential expression [35]. This strategy has been widely adopted in breast cancer biomarker studies [36-39]. However, taking into account overall cost, time, reproducibility and recovery rate, the 1DE SDS-PAGE is becoming the fractionation method of choice for many global LC-MS-based proteomics applications [40, 41].

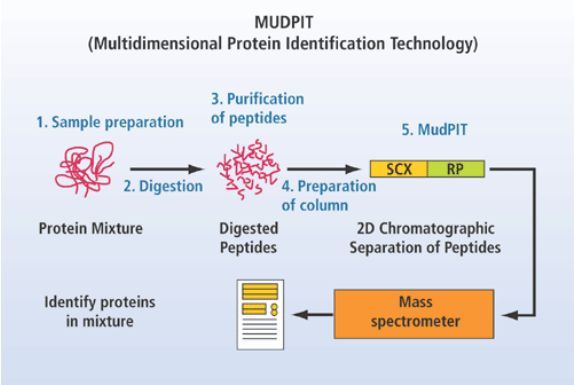


Figure 1.4 A schematic diagram of a MudPIT experiment.
(Source: http://www. dddmag.com/articles/2007/10/got-mudpit)

1.2.2.2 Gel-free separation of peptides

Ongoing concerns regarding limitations of gel-based methods such as biased detection of certain classes of proteins and increased sample handling led to the development of an innovative gel-free technique for protein separation known as multidimensional protein identification technology (MudPIT) [42]. The seminal work in the early days of modern proteomics impressively described the identification of more than 5,000 peptides mapped to around 1,500 proteins from the yeast proteome, many of which were of low abundance. The technology has been widely applied to study global proteomic changes in cancer [43-45], where a complex mixture of proteins is first digested and applied to a strong cation exchange resin (SCX) chromatography column, followed by reversed phase (RP) chromatography, prior to MS analysis

(Figure 1.4). The coupling of ion exchange and RP chromatography to MS represent a typical mode of orthogonal 2D LC analysis of proteins where peptides are separated using two different mechanisms to ensure maximum peptide coverage, thus enhancing protein identification. Limitations of SCX chromatography including low peptide resolution, reduced sample recovery and the need for sample desalting [46] have led to the development of other orthogonal methods such as strong anion exchange (SAX)/RP [46] and high pH-low pH RP systems [47] showing improved separation efficiency.

The choice of either gel-based or non-gel-based method will depend on several factors including time, cost, sample type and complexity. A brief comparison of the analytical advantages and limitations of both approaches is listed in **Table 1.2.**

Table 1.2 Comparison between gel-based and non-gel-based protein separation for LC-MS/MS analysis

Advantages	Disadvantages
1D/2D GEL	
▪ Ability to identify novel proteins ▪ Separates protein modifications (2D) ▪ Good resolving power (2D) ▪ Less time required (1D)	▪ Biased towards certain classes of proteins ▪ Limited reproducibility (2D) ▪ Limited dynamic range (1D) ▪ Limited sensitivity but improved for DIGE
NON-GEL BASED 2D LC-MS/MS	
▪ Ability to identify novel proteins ▪ Improved separation efficiency and proteome coverage ▪ Less sample handling ▪ Less biased than gel-based for certain protein classes	▪ No visualization of separated proteins ▪ Time consuming for increased SCX/SAX fractionation ▪ Less flexibility in setup

1.2.2.3 Lectin affinity chromatography

As products of one of the most common PTMs, glycoproteins constitute a major class of biomolecules with significant roles in pathological processes including various human cancers. The glycoproteome, i.e. the entire complement of cellular glycoprotein expression, of tumor cells is therefore an attractive source to mine for potential biomarkers. For comprehensive coverage of

the glycoproteome in a complex mixture such as serum, the wide dynamic molar range of proteins often necessitates enrichment prior to their analysis. Plants lectins have the ability to bind to specific glycan epitopes and more than 60 of them, which recognize a diverse range of glycan structures are commercially available. Table 1.3 shows a partial list of commonly used lectins and their known specificity including concanavalin A (Con A), jacalin (JAC) and wheat germ agglutinin (WGA), *Aleuria aurantia* lectin (AAL), *Lens culinaris* agglutinin (LCA), *Sambucus nigra* lectin (SNA), peanut agglutinin (PNA), and *Phaseolus vulgaris* leucoagglutinin (PHA-L). Some of these lectins such as Con A, WGA and JAC display a broader glycan specificity, which is useful to capture a wider range of glycoproteins while others such as LCA, SNA and PHA-L have narrower glycan selectivity and can be used for more targeted glycoform enrichment.

Table 1.3 Commonly used lectins and their glycan specificities (Adapted from Fanayan *et al*, 2012 [23])

Lectin	Specificity
Concanavalin A (Con A)	High-mannose type, branched α-mannosidic structures
Wheat germ agglutinin (WGA)	*N*-acetylglucosamine; chitobiose (sialic acid)
Jacalin (JAC)	Galactosyl (b-1,3) *N*-acetylgalactosamine (O-glycoproteins)
Sambucus nigra lectin (SNA, EBL)	Sialic acid attached to terminal galactose in (a-2,6)
Peanut agglutinin (PNA)	Galactosyl (b-1,3) *N*-acetylgalactosamin (T-Antigen)
Lens culinaris agglutinin (LCA)	α-Linked mannose residues
Phaseolus vulgaris leucoagglutinin (PHA-L)	Tri/tetra-antennary complex-type *N*-glycan
Aleuria aurantia lectin (AAL)	Fucose linked (a -1,6) to *N*-acetylglucosamine; fucose linked (a -1,3) to *N*-acetyllactosamine

Several modes of lectin affinity chromatography workflows have been established to isolate cancer-associated glycoproteins from complex biological samples. Using single lectin affinity chromatography, lectins with narrow selectivity were shown to enrich a small group of

glycoproteins with 3-fold or more change in concentration between normal and breast cancer patient plasma [48]. Varieties of lectins can be used consecutively in an approach known as serial lectin affinity chromatography (SLAC); The specificity of JAC for O-glycans was shown to be increased by first using Con A to remove high mannose type *N*-glycans before application of the flowthrough fraction to the JAC column [49]. On the other hand, using multiple lectins that recognize different glycan motifs in a single column can increase the range of glycoproteins isolated simultaneously [50]. This approach is known as multiple lectin affinity chromatography (M-LAC). This method is the subject of investigated detailed in Chapter 5 where lectins with broad glycan specificities were used i.e. Con A, JAC and WGA with the aim to capture large and complex subsets of proteins with different glycoforms for further analysis.

1.2.3 LC-MS/MS-based protein detection of peptides

Both gel-based and gel-free separation shotgun proteomic approaches have allowed for high throughput characterization of complex mixtures of proteins when coupled with advanced LC-MS/MS [51-53]. The LC serves to chemically separate the peptides and is usually achieved on the basis of differential peptide hydrophobicity using RP column packed with octadecylsilyl (C18) stationary phases. Bound peptides are progressively eluted with increasing gradient of hydrophobic organic solvents such as acetonitrile (ACN) which are then subjected to ionization in the MS. The development of the two most popular ionization sources for MS, namely, matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) has won their inventors the Nobel Prize in Chemistry in 2002 [54]. Both types are known as relative soft ionization methods (Figure 1.5). MALDI involves embedding the analyte molecules (i.e. peptides) in an organic matrix which become ionized when the matrix absorbs energy from the laser. The exact mechanism of desorption and ionization is still unclear [55]. ESI utilizes electrical voltage and heat to transfer ions from liquid to the gas phase generating peptide precursor ions; which are then separated according to their mass to charge ratios (m/z) for further fragmentation using,

for example, collision induced dissociation (CID) to generate MS/MS spectra [56]. Although both methods of ionization are routinely used in proteomics, ESI typically produces a range of multiply charged ions that can be detected in ideal m/z ranges of all common types of mass analyzers for biomolecular analysis thereby enabling efficient identification and characterization of large biologically important macromolecules such as peptides, proteins, nucleic acids and carbohydrates [57]. On the other hand, the combination of MALDI and time of flight (TOF)-MS in applications such as peptide mass fingerprinting demonstrates the usefulness of MALDI as a rapid and sensitive analytical tool for protein identification.

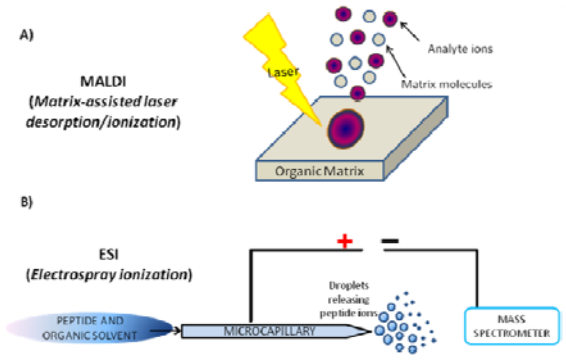


Figure 1.5 The two major ionization methods for modern MS of biomolecules are (a) MALDI and (b) ESI. (Adapted from Lucio *et al*, 2013 [58])

1.2.4 Label-assisted and label-free mass spectrometry-based protein quantitation

The ability to accurately quantify protein expression in comparative studies represents an important but challenging task enabling determination of proteins that may play key biological roles in disease development and potential biomarkers or drug targets. Currently, MS-based relative quantitation of individual proteins from different samples can be undertaken either via the label-assisted or label-free techniques.

1.2.4.1 Label-assisted methods

To perform a typical label-assisted quantification experiment, two or more protein samples are simultaneously investigated in a single run. Stable heavy isotopes are widely used for protein

labelling (Figure 1.6b) and can be performed at either the protein or peptide level. However, it is desirable to introduce the label as early in the sample work-up as possible to allow mixing of the samples being compared and thereby avoid introduction of bias from sample preparation steps, which may skew the relative quantitation.

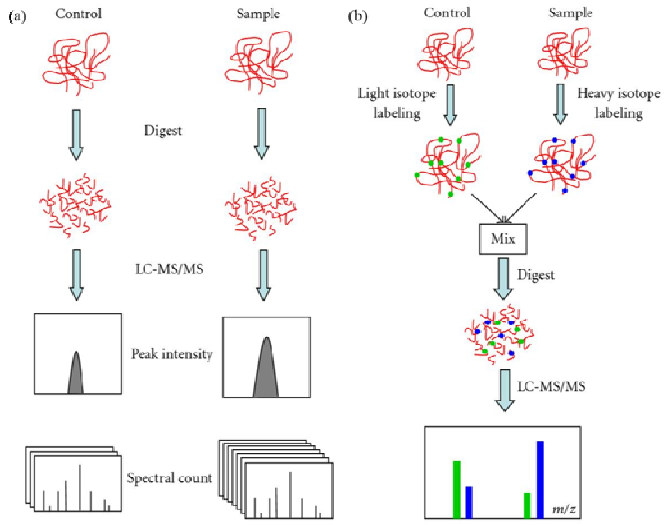


Figure 1.6 Relative quantitation can be performed with (a) label-free technique based on peak intensity or spectral count; or (b) labelled stable isotopes to generate "light" and "heavy" samples. (Adapted from Zhu *et al*, 2010 [59])

Common labelling strategies include (1) introducing stable isotope-containing amino acids in the cell culture media (SILAC) that contains $^{12}C_6$-lysine and $^{13}C_6$-arginine which are then metabolically incorporated into proteins via *de novo* protein biosynthesis [60]; (2) chemically modifying the sulfhydryl-reactive chemical group of the protein or peptide with isotope-coded affinity tag (ICAT) [61]; (3) enzymatic labelling of proteolytic fragments with ^{18}O [62]; and, (4) modifying peptides with amino-reactive isobaric labels including isobaric tags for relative and absolute quantitation (iTRAQ) [63] or tandem mass tags (TMT) [64]. As SILAC has been restricted to *in vitro* use, a modified approach has been developed to facilitate protein quantitation *in vivo*, for example, the human tissue proteome [65]. The super-SILAC strategy greatly improved quantification accuracy of various tumor tissues by analysing the combined mixtures of five SILAC-labeled cell lines with the individual tumor tissues. Differentially labelled samples, typically defined as "heavy" and "light", are pooled and analyzed together in the same LC-

MS/MS run and the quantitative difference between the samples is detected by means of the quantitative intensities of specific mass shifted signals in the MS or MS/MS.

Relative quantitation using metabolic, ICAT and enzymatic labelling are based on mass difference between differentially labelled peptides and therefore limited by the number of samples analyzed in a single run as labelling beyond 3-plex set of samples would lead to more complex mass spectra. This issue is addressed by the design of isobaric chemical tags, i.e., iTRAQ and TMT reagents, enabling multiplexed analysis of different biological samples or conditions (up to eight for iTRAQ and ten for TMT) in a single experiment. The N-termini and the lysine residues of proteolytic peptides are modified by iTRAQ or TMT tags, each containing a mass balance and a unique reporter group, which are indistinguishable in the MS. Upon peptide fragmentation, distinct low-mass reporter ions are generated and their intensity ratio measured, yielding quantitative information of proteins present in the samples. Both iTRAQ and TMT are widely applied to shotgun proteomics to quantify and identify differentially expressed proteins as disease biomarkers [66-68]. However, studies have reported an underestimation of fold changes, also known as "ratio compression", using these quantitation methods which leads to substantially less proteins being identified and quantified [69]. Strategies that addressed this limitation include employing fractionation to reduce sample complexity [70], removing co-isolating impurities that interfere with peptide elution through gas-phase purification [71] and applying computational algorithms to improve the quantitation accuracy while retaining protein coverage [69].

1.2.4.2 Label-free quantitation

.In recent years, technological advancements in the high performance (HP) LC system and high resolution/accuracy mass spectrometry have facilitated the use of label-free quantitation in numerous comparative proteomics studies. The strategy requires each sample containing unlabelled peptide mixtures to be analyzed in separate LC-MS/MS runs. The relative protein

abundance is then determined by comparing the precursor peak ion intensities or MS/MS spectral counts of the corresponding peptides across all the samples. The relative protein abundance has been reported to correlate well with the relative precursor peptide ion intensity or spectral counts of identified peptides with reproducible results [72-74]. Strong correlation has also been observed for complex protein mixtures in particularly for medium to high abundance proteins [74-76]. Precursor ion intensity is quantified by measuring the area under curve of the peptide precursor ions in the LC-MS. However, peak area can vary from run to run, even with the same sample but from two injections, extensive normalization is therefore necessary to account for such variation amongst samples and is achieved using sensitive computer algorithms to automatically align the extracted ion chromatograms prior to comparison [77].

Relative protein abundance as assessed by spectral counts relies on the sum of MS/MS spectra obtained for each identified peptide across different samples. Normalization is also required for spectral counting methods but without the need of complicated computer algorithms, for example, the simple but robust method known as the normalized spectral abundance factor (NSAF) takes the protein length into account. NSAF was shown to be able to reliably determine the quantitative changes of membrane proteins in yeast following statistical analysis [78]. Zybalilov *et al* defined the NSAF for a protein k in the formula shown below where the numerator is the spectral count (SpC) of a protein divided by the length of protein (L) and the denominator is the sum of numerator of all N proteins in the experiment. Although both quantitative methods show a high degree of correlation to protein abundance, higher reproducibility and a large dynamic range were observed with spectral counting than with precursor peptide ion intensity [76].

$$(NSAF)_k = \frac{(SpC/L)_k}{\sum_{i=1}^{N}(SpC/L)_i}$$

A major issue in label-free is the potential bias arising from sample preparation and data acquisition [79]. However, such bias can be minimized with careful sample handling and implementation of good practises for LC-MS/MS data acquisition, for example, inclusion of quality control runs and performing all analyses in a single batch in randomized order. In addition, to increase the statistical significance of the measured protein fold-change, it is necessary to perform sample analysis in multiple technical replicates (at least in triplicate). The main advantages of label-free approaches are the ability to perform and compare many samples in a single experiment without additional sample processing steps and at a relatively low cost. Label-free quantitation has been shown to have a greater and deeper proteome coverage compared to iTRAQ [80]. Taken together, label-free quantitation is particularly suited for large-scale discovery-based studies to interrogate sets of differentially expressed proteins to map unique molecular signatures associated with specific conditions. Based on these reasons and accessibility to high mass accuracy and high resolution LC-MS/MS instrumentation, the label-free quantitation method was chosen as the method of choice for the quantitative proteomics analyses in Chapter 2.

1.2.5 Bioinformatics tool for protein identification

The computational task in protein identification begins with matching the acquired MS/MS spectra to a database of theoretical spectra generated from *in silico* digestion of protein sequences (translated from protein coding regions of DNA) by specifying the appropriate parameters such as cleavage rules, possible modifications, species, precursor and fragment ion mass tolerance and charge states (Figure 1.7). The availability of complete whole-genome sequences has benefitted greatly the process of protein identification in that the protein sequence database consists of translated protein sequences from known genomic data. The curated non-redundant and publicly available SWISS-PROT database, containing the entire set of known human proteins and their predicted fragment spectra is widely used in proteomics [81], while a high-quality spectral library

with actual experimental spectra has also been constructed for more accurate spectral matching and scoring [82]. In order to validate the results of the highly automated and high-throughput process of database search, it is necessary to estimate the false discovery rate (FDR) to minimize false-positive identifications. Database search is typically repeated using identical search parameters against a decoy database, in which the protein sequences have been reversed from the true target protein database [83]. The FDR is usually pre-specified before the database search, the value is reported as the ratio of the number of matches in the decoy database to the total matches in both the target and decoy databases.

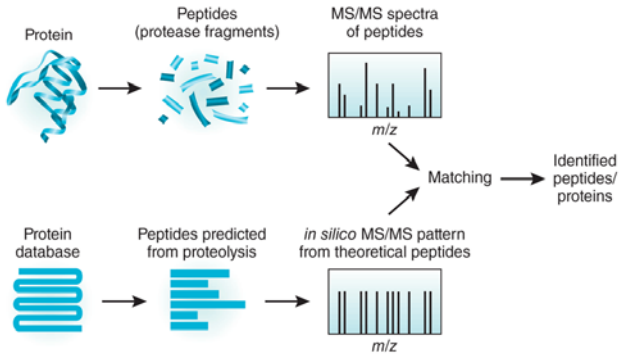


Figure 1.7 Workflows illustrating the typical approaches in database-assisted protein identification from LC-MS/MS data. (Adapted from Duncan *et al*, 2010 [84])

Confident peptide identification from complex mixtures depends heavily on the resolving power and mass accuracy of the mass spectrometer, which in turn will affect the stringency and the FDR of the database search. In typical MS/MS spectra, the predominant fragments ions observed are the b and y ions, generated from the cleavage of the polypeptide backbone. During database search, computer algorithms are ran to match experimentally-derived fragment ions against pre-defined *in silico* peptide fragmentation in the database to return a list of peptide sequences which are ranked with their probability score or FDR. Higher MS resolution allows for narrower mass tolerance of precursor and fragment ions leading to higher level of confidence in the peptide identification. Therefore MS with high resolution in excess of 20,000 and high mass accuracy below 10 parts per million (ppm), such as the orbitrap platform or triple quadrupole TOF instruments, are much sought after in proteomics-based studies.

Scores are generally given for peptide-spectrum matches (PSMs) and the highest scoring matches are used for protein assignment. Essentially the identity of the protein is inferred from the peptide sequences that match the queried MS/MS spectra during database search. Ambiguity of protein identification can arise when a common or degenerate peptides are present in multiple proteins, which therefore complicates the interpretation of proteomics data [85]. This "protein inference problem" is further compounded when a protein has only one single identifiable peptide (so-called "one-hit wonders") which naturally has a higher probability of false-positive identification compared to proteins covered by multiple identified peptides. Taken together, it becomes evident that accurate assignment of proteins assembled from identified peptides requires sophisticated statistical computations.

More than a dozen complex algorithms to assist the protein identification from LC-MS/MS data have been developed, some of which are proprietary while others are freely available. Currently, the three most commonly used are Sequest [86], Mascot [87] and X!Tandem [88]. These algorithms also address the issue of "one-hit wonders" by incorporating additional statistical tools such as PeptideProphet to validate PSMs [89]. Recently, simpler algorithms have been written to target high quality MS data obtained from the increased use of high resolution and high mass accuracy mass spectrometers such as the Orbitrap or high-end Q-TOF platforms [90, 91]. In this thesis, X!Tandem which is publicly available and run from the Global Proteome Machine (GPM) interface, is the main software tool used for protein identification.

In recent years, a number of proteomics software with a suite of analytical tools have been developed, containing pipeline features that allow comprehensive analysis of high quality MS data including protein and PTM identification as well as peptide/protein validation and quantitation. Some software are licensed by manufacturers of mass spectrometers such as Proteome Discover (Orbitrap, Thermo Scientific) and ProteinPilot (TripleTop, ABSciex), while others are standalone

proprietary software such as Scaffold ([92] and Byonic [93]). Freely available open source proteomics software tools are also widely used and include the Trans-Proteomic Pipeline[94] , Skyline [95] and MaxQuant [96].

1.2.6 Proteomics in breast cancer biomarker discovery

In the past decades, significant amount of proteomics data in the space of breast cancer has been accumulated through large scale comparative studies of breast cancer and "healthy" normal reference samples utilizing model systems such as mammalian cultured cells [45, 51, 53, 97-101], mice models [102-105] and human tumor xenografts models (in mice) [106-108], clinical biological specimen such as serum [52, 109-116] tumor tissues [98, 117-119], tissue interstitial fluid (TIF) [120, 121], nipple aspirate fluid (NAF) [122-124], cerebrospinal fluid [125], saliva [38, 126] and tear [127]. The motivation underlying most of these efforts were the anticipation that the proteins between disease and normal states provide molecular signatures or yield insight into the intracellular signalling pathways that lead to initiation and progression of breast tumors. Such knowledge may in turn identify novel biomarkers and new drug targets [128]. For example, quantitative proteomics analysis of ER-negative breast tumor cells of defined breast cancer stages identified a multi-marker signature of three proteins (isocitrate dehydrogenase 2, cellular retinoic acid-binding protein 2 and alpha-tocopherol-associated protein) that were predictive of overall breast cancer survival [98].

Well-characterized cell lines established from primary breast tumors, pleural effusions or other metastatic sites are widely used in MS-based breast cancer studies. To date more than 50 breast cancer cell lines are available to researchers; the more frequently used cell lines are shown in Table 1.4. Cell lines provide a continuous source of homogeneous cell population and therefore largely overcome the issue of cellular heterogeneity contributed by for example, the stromal, endothelial, adipose and immune cells in clinical samples [129]. The frequently-used breast cancer

cell lines such as MCF7, SKBR3 and MDA-MB-231, which are also part of the human epithelial breast cell panel investigated in this thesis, have been individually characterized at the proteome level [130-132]. The shotgun proteomics approach utilized in these studies identified numerous proteins that were previously recognized to be involved in breast cancer tumorigenesis. For example, cathepsin D, 14-3-3-sigma, antigen Ki-67 in MCF7; human receptor protein kinase, breast cancer type 1 and 2 susceptibility proteins, and N-myc proto-oncogene protein in SKBR3; breast tumor suppressor p53 and epidermal growth factor receptor in MDA-MB-231.

Table 1.4 Clinicopathological features of frequently used breast cancer cell lines*

Cell line	Subtype	ER	PR	HER2	Source	Tumor type
184A1	B	–	NA	–	RM	NA
BT20	A	–	–	–	PT	AC
BT474	L	+	+	+	PT	IDC
BT549	B	–	–	–	PT	IDC
HS578T	B	–	–	–	PT	C Sar
MCF7	L	+	+	–	PE	Met AC
MCF10A	B	–	–	–	RM	F
MDA157	B	–	–	–	PE	Med C
MDA231	B	–	–	–	PE	Met AC
MDA453	L	–	–	+	PE	Met C
MDA468	A	–	–	–	PE	Met AC
SKBR3	L	–	–	+	PE	Met AC
T47D	L	+	+	–	PE	IDC

A = Basal A subtype; AC = adenocarcinoma; B = Basal B subtype; C Sar = carcinoma sarcoma; F = fibrocystic disease; IDC = invasive ductal carcinoma; L = Luminal subtype; Med C = medullary carcinoma, Met AC = metastatic adenocarcinoma; Met C =metastatic carcinoma; NA = not available; PE = pleural effusion; PT = primary tumor; RM= reduction mammoplasty. (*Extracted from Kao *et al*, 2009 [133])

The major concern whether breast cancer cells are representative of the molecular diversity observed in breast tumors were addressed through profiling of various breast cancer cell lines, which mirrored the luminal-basal subtype distinction established in true breast tumors after surgery [129, 133, 134]. However, researchers need to take into consideration the limitations associated with using cell lines, including their genomic instability, risk of cross-contamination and intra-laboratory cell line heterogeneity during the experimental design and data interpretation [135].

The global expression of membrane and secreted proteins of breast cancer cell lines are often the main focus as they are a rich source of potential biomarkers and drug or antibody targets. An early work using 2DE to separate membrane proteins proved to be challenging as they are not readily amenable to IEF such that very few proteins were consistently observed across the 25 malignant samples [136]. In a subsequent study, SILAC was used to investigate a pair of isogenic cell lines, accurately representing different stages of metastasis. In the same study, metastasis-related plasma membrane proteins including CD74, CD44, CD98, ecto-5-nucleotidase, integrin β1, integrin α6, annexin A2 and MUC18 were identified, some of which were validated by immunohistochemistry staining (IHC) in primary breast cancer biopsies [99]. More recently, MudPIT analysis identified more than 5,000 plasma membrane proteins extracted from a panel of breast cancer cell lines. The large amount of derived proteome knowledge revealed the correlation of the expression of plasma membrane proteins with the aberrant expression of tyrosine kinases (eg. proto-oncogene c-kit and ephrin receptor), cellular adhesion molecules (eg. CD44 and tetraspanins) and structural proteins (eg filamin A and alpha-actinin-4) [45]. Various proteomics technologies including 1DE, 2DE, 2D-DIGE and SILAC were used to capture the secretome profiles of breast cancer cells. These studies suggested a number of proteins with biomarker potential including proteasome activator complex subunit 1 and HLA class I histocompatibility antigen [100]; PDZ domain containing 1, 4-aminobutyrate aminotransferase and Pentraxin-related protein PTX3 [101]; salivary cystatins (CST1, CST2 and CST4), plasminogen activators (PLAT and PLAU) and collagen proteins (PLOD2 and COL6A1) [137]; bestrophin-3, parvabumin, barrier-to-autointegration factor and the 14-3-3 proteins [53]. A consistent observation across these studies was a significant presence of exosomes (alternatively referred to as microvesicles or microparticles in the literature) in the secretions. These organelles are secretory vesicles that carry macromolecules including proteins, nucleic acids and lipids which all may be involved in cell-cell communication [138]. Given their emerging relevance in cancer

development, they have been the focal point in recent proteomics profiling studies including those on breast cancer [139-141].

The first step for successful translation of potential biomarkers identified in *in vitro* systems into clinical applications is their detection in clinical samples such as tissue biopsies or serum of breast cancer patients. Accordingly, findings from cell cultured-based proteomics studies were validated using tissue biopsies [99, 101] or plasma [142]. However, there may be inconsistencies in the results obtained from tumor tissues as illustrated by the example of 14-3-3 sigma protein, a promising early-stage biomarker for breast cancer. Initial studies on tumor tissues showed that this signalling protein was down-regulated in a specific breast cancer subtype but this observation could not be replicated in other tumor types in a subsequent investigation [143]. Aside from tumor heterogeneity, the hormonal micro-environment of the tumor tissues can also influence protein expression, hence, affecting proteome deregulation [144].

The serum or plasma is often considered the preferred source for mining and validation of diagnostic biomarkers as it is readily obtainable directly from patients and healthy donors with minimal invasiveness. In addition, the circulating body fluid is an enriched reservoir of proteins secreted from all cells lining the blood circulation including the sites of primary tumor lesions or metastases, thus reflecting the physiological and pathological status of individuals. However, several factors present enormous analytical challenges for comprehensive serum proteome analysis. Firstly, the dynamic molar range of proteins in the serum is extreme, spanning at least 10 orders of magnitude [145]. Secondly, the serum proteome is dominated by a few high-abundant proteins including albumin which alone contributes to more than half of the total serum protein and together with at least 15-20 other high-abundant proteins make up almost 95% of the total serum protein content [146]. Thirdly, there is intra- and inter-individual variations, which lead to a large biological variation even within patient groups and various protein isoforms may be

observed in the serum [147]. Lastly, the sample collection, handling and storage conditions can lead to profound changes in the serum proteome, thereby affecting the downstream interpretation of the acquired data [148]. Immunoaffinity depletion is often applied to serum samples to remove albumin and other highly abundant serum proteins so as to benefit subsequent separation techniques such as 2DE or MudPIT. While detection of the less abundant proteins is improved with serum depletion, loss of potentially valuable low-abundant proteins has been known to occur by their binding to albumin [149]. A number of initiatives have been implemented to standardize protocols for sample collection and handling [150] and setting up resource databases containing high-confidence human plasma proteome reference sets such as PeptideAtlas [151] and Plasma Proteome Database (PPD) [152].

Extensive MS-based proteomics have been performed to compare the serum of breast cancer patients to those of normal healthy individuals to identify differentially expressed proteins. Using a three-step method (immunodepletion of abundant proteins followed by fractionation using RP-HPLC and 2DE PAGE) and the LC-MS/MS of serum samples from breast ductal carcinoma *in situ* (DCIS) patients and normal controls, one study revealed vitronectin to be a novel candidate serum marker for early detection of DCIS amongst a list of other differentially regulated proteins [116]. The elevated protein expression in DCIS samples was validated using enzyme-linked immunosorbent assay, western blot and IHC staining of tissues samples. The findings were supported by previous studies reporting similar observations [153, 154]. It is thought that vitronectin regulate proteolysis by binding to and stabilizing plasminogen activator inhibitor-1 (PAI-1) while inhibiting the activity of urokinase plasminogen activator receptor (uPAR) [155]. Both PAI-1 and uPAR have been shown to have clinical utility by their implicated roles in cancer invasion and metastasis. Thus vitronectin may be a promising biomarker candidate for the early detection of breast cancer.

Constraints in using tissues and serum as suitable sources of biomarkers in the clinic have led other researchers to seek alternative sources of biological relevant specimen including the nipple aspirate fluid (NAF) and tissue interstitial fluid (TIF). NAF breast fluid is rich in proteins, secreted by epithelial cells that line the ducts while TIF is the extracellular fluid that surrounds the breast tissues. Both are considered as attractive sources because of their proximity to the primary tumor site and the lower proteome complexity relative to serum. In a study using label-free spectral counting that compared the NAF proteomes between breast cancer and normal individuals, almost 900 non-redundant proteins were identified of which half were unique to the cancer-associated NAF, thus validating this bodily fluid to be a valuable source of breast cancer-specific biomarkers [122]. However, the results from these small sample sizes need to be validated with a larger sample cohort. Another study applied a more extensive biomarker discovery research strategy that initially identified 110 differentially regulated proteins in the TIF derived from a pair of matched breast tumor/benign tissues [120]. This was then followed by comparative proteomic analysis with the remaining 68 pairs of samples to single out a set of 26 common breast cancer-related proteins including calreticulin, cellular retinoic acid-binding protein II, chloride intracellular channel protein 1, EF-1-beta, galectin 1, peroxiredoxin-2, platelet-derived endothelial cell growth factor, protein disulfide isomerase and ubiquitin carboxyl-terminal hydrolase 5, which were validated by a tissue microarray containing 70 various grades of malignant breast carcinomas. Many of these have already been observed in other plasma- and secretome-based studies and the authors proposed that future studies will evaluate their true potential as breast cancer biomarkers.

1.2.7 Functional analysis of proteomics – pathway analysis and interaction networks

To achieve the overarching goal of identifying potential biomarkers for breast cancer diagnosis, prognosis and targets for therapy, an improved understanding of the global and integrated view of molecular mechanisms underlying breast cancer biology is essential. Studies that use high

throughput proteomics approaches, including those in this thesis, have generated a vast amount of high quality data enabling comprehensive system-wide investigation of protein deregulation during breast cancer. It is now well-accepted that cancer is a systems biology disease with multiple oncogenic proteins involved simultaneously in different cellular processes [156]. Therefore, in addition to identifying, characterizing and quantifying the proteins which are differentially expressed in biological samples, an emerging theme is to undertake a functional interpretation of the proteome-wide changes. Two strategies for functional proteome analysis that are widely adopted by an increasing number of proteomics-based studies are pathway and protein-protein network analyses.

A major challenge in functional proteome analysis is capturing biologically significant information from the large datasets. To this end, it relies heavily on the effective use of bioinformatics tools to query knowledge bases that have been established and meticulously maintained by individuals, research institutions or consortia, for example, UniProt [157], Gene Ontology (GO) [158], Kyoto Encyclopaedia of Genes and Genomes (KEGG) [159] and Reactome [160]. These database-centric resources provide integrated biological information of genes, mRNAs, proteins and other small molecules including their biological processes, components, structures or molecular interactions, either predicted or experimentally observed [161]. The most basic approach in functional proteome analysis is to categorize the identified proteins using GO terms consisting of defined descriptors that relate proteins with their biological processes, molecular functions or cellular components [158]. More in-depth analysis can be performed from the perspective of biological pathways or protein-protein networks, which seek to understand the key processes underlying the functional roles of differentially expressed proteins by statistically evaluating their relationships and interactions with one another in a given condition. Bioinformatics tools proved to be indispensable for this type of data mining as they are able to organize and reduce the complexity of large volumes of data to present a visual view of significantly important biological

patterns and relationships. Examples of freely available computational tools for functional proteome analysis include STRING [162], DAVID [163] , PANTHER [164] and Cytoscape [165] or commercially developed software such as GeneGo MetaCore (www.genego.com) and Ingenuity Pathway Analysis (IPA) (www.ingenuity.com).

In an effort to gain better mechanistic insights, proteome-based studies that explored the functional aspects of the resulting proteomes identified important oncogenic processes and protein interaction networks that are critical for cancer progression [45, 51, 100, 137, 166-168], indicating the usefulness of these approaches to interpret large data sets. Some recurring themes that emerged from these analyses include perturbations of cellular structural integrity, changes to the extracellular matrix (ECM) composition, abnormal intracellular signalling, increased cell locomotion and an activated immune system in the cancer pathophysiology. These altered processes were found to be orchestrated by changes in the expression of groups of functionally similar proteins such as cytoskeletal proteins, extracellular matrix proteins, cell surface integrins, tyrosine kinases, adhesion proteins and peptide-presenting proteins. Individual groups of proteins can interact within their own network as well as work synergistically with other protein groups to promote cancer invasion and metastasis.

Increasing evidence suggests that dramatic changes in the extracellular matrix (ECM) composition play a significant role during successive stages of breast cancers from the initial appearance over progression to metastasis [169]. Structurally, the ECM is composed of two cellular and biochemically distinct components, the basement membrane (BM), which forms a physical barrier separating the epithelium or endothelium from the stroma; and the interstitial matrix, which is mainly made up of the stromal cells [5]. Laminin, entactin, type IV collagen and heparin sulphate proteoglycan (perlecan), secreted from the epithelial, endothelial and stromal cells are found in the BM while the interstitial matrix is composed of a mesh of fibrillar collagen,

glycosaminoglycans (GAGs), glycoproteins such as fibronectin, thrombospondin, tenascin and tissue inhibitor metalloproteinase. The BM has two important functions: (1) it serves as an anchorage for the epithelium through binding to the transmembrane integrins or non-integrin protein such as dystroglycan, both of which act as a linkage between the ECM and cytoskeleton within the cells [170], (2) it induces epithelial cell polarity and differentiation, mediated through the integrins, to regulate the development and homeostasis of epithelial tissues [171]. Similarly, the interstitial matrix is critically involved in cellular communication by regulating the activity of growth factors by means of binding to them thereby limiting their diffusion. Thus the ECM components serve to provide structural support to tissues and modulate biochemical signals to influence cellular behaviour such as proliferation, polarity adhesion, migration, polarity and migration through its interaction with cellular receptors, primarily the integrins.

The ECM-cell interactions are highly dynamic, with multiple regulation and feedback mechanisms to keep the cellular activities under tight control. Evidence from *in vitro* studies demonstrated that fibroblasts in the stroma can become activated, secrete various growth factors and ECM proteins, and as a consequence initiate carcinogenesis by autocrine signalling [172]. Aberrant ECM remodelling that leads to the degradation of ECM components, particularly in the BM, represents an essential step for tumour invasion and metastasis (Figure 1.8). Several studies have shown that the down-regulation of BM components, such as the laminins and type IV collagen [173-175]; and over-expression of ECM degrading enzymes such as the matrix metalloproteinases, are associated with breast tumorigenesis [176, 177]. The loss of epithelial anchorage and polarity allows the cells to gain mobility, breaching the BM to invade the dense interstitial matrix and acquiring mesenchymal-like characteristics; a model known as epithelial mesenchymal transition (EMT), which describes the progression of cancer development into metastasis [178]. This process is generally accompanied by a number of deregulated events including altered expression of cytosekeletal proteins, cell adhesion molecules such as cadherins,

integrins, membrane-associated tyrosine kinases, growth factors and cytokines. Cadherin switching, i.e loss of E-cadherin and over-expression of N-cadherin on tumor cell surface, and altered expressions of catenins, which are found in cadherins complexes, is well-documented in epithelial carcinomas such as breast cancer [179-181]. Although EMT is well studied, the molecular events that initiate these processes are still poorly understood but may involve deregulation of pathways associated with protein modification, i.e. phosphorylation and glycosylation.

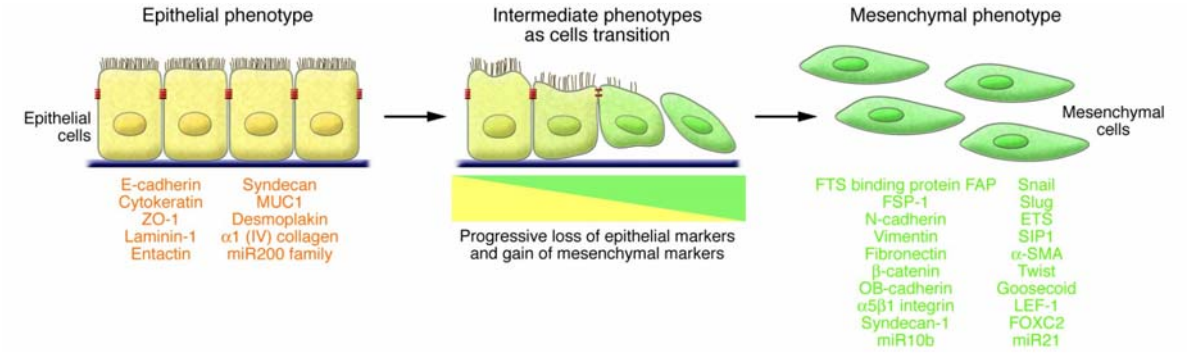


Figure 1.8 Schematic diagram showing the process of epithelial-mesenchymal transition. (Adapted from Kalluri *et al*, 2009 [182])

1.3 Protein glycosylation

1.3.1 Protein glycosylation – a brief overview

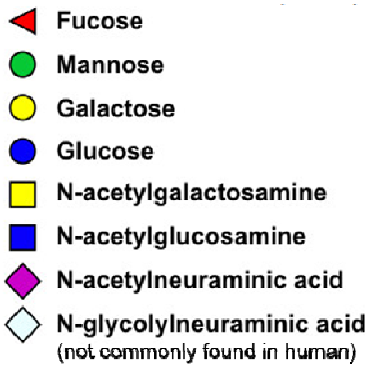


Figure 1.9 Symbol nomenclature for the representation of glycans as proposed by the Consortium of Functional Glycomics (CFG). Only the most common monosaccharide building blocks for mammalian glycans are shown.

Glycosylated proteins ubiquitously decorate human cell surfaces and are major components of the extracellular matrix. Protein glycosylation is a universal phenomenon that occurs in all forms of life ranging from most basic prokaryotic cells to the complex multicellular eukaryotic cellular systems where it generates an array of glycoproteins [183]. The attached glycans displayed high

structural diversity, which is rather remarkable considering the limited number of monosaccharide building blocks utilized for mammalian glycan synthesis including fucose (Fuc), mannose (Man), galactose (Gal), glucose (Glc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), *N*-acetylneuraminic acid (NeuAc) and *N*-glycolylneuraminic acid (NeuGc) (Figure 1.9). The often partial glycan occupancy of various glycosylation sites on the polypeptide backbones, also known as protein macro-heterogeneity, in combination with the extensive micro-heterogeneity caused by variation in the glycan length (number of building blocks), the monosaccharide compositions, topology/branching and linkage types dramatically increase the diversity of the glycosylated proteome, which is considered to be essential to facilitate the diverse functional roles of glycoproteins [184].

Several types of protein glycosylation are known in human but the two most common types involve glycans enzymatically attached to the protein via either *N*- or O-glycosidic linkages. In protein *N*-glycosylation, an *N*-glycan precursor is added via a reducing-end *N*-acetylglucosamine (GlcNAc) residue to asparagine (Asn) residues on polypeptides found within a consensus peptide sequences or "Asn-sequons" displaying Asn-X-Serine/Threonine (Ser/Thr), where X can be any amino acid residue except for proline [185]. However, not all predicted Asn-sequons are glycosylated, indicating that consensus sequences alone do not solely dictate *N*-glycosylation and that additional primary structure features or conformational requirements may be needed to promote *N*-glycosylation [186, 187]. *O*-glycosylation involves the attachment of an *N*-acetylgalactosamine (GalNAc) residue to either Ser or Thr residues on the protein backbone. So far, no recognition motifs or sequons have been identified and it remains as such unclear why certain Ser or Thr residues are *O*-glycosylated whilst others are not. Both *N*- and *O*-glycosylations are prevalent on membrane and secretory (non-mucin) proteins while *O*-glycosylation is also commonly found on large viscous cysteine-rich glycoproteins known as mucins. Mucins are expressed in large quantities on many epithelial surfaces of the body, including the

gastrointestinal and respiratory tract and by the salivary and sweat glands [188]. We have established a robust MS-based workflow for profiling and characterization of *N*-glycans released from various biological samples and the focus in this thesis is exclusively on protein *N*-glycosylation.

Despite the fact that a plethora of *N*-glycoforms decorate mammalian proteins, all *N*-glycans are synthesized using the same biosynthetic machinery. As such, all maturing glycoproteins traffic a common pathway known as secretory pathway. Starting with the addition of a common glycan precursor of which only the outer domains gets modified by truncation and extension reactions, all *N*-glycosylation share a common tri-mannose chitobiose core ($Man_3GlcNAc_2$) [189]. The extension from the two non-reducing end mannose residues of the chitobiose core [Manα(1,3) and Manα(1,6)] by the addition of various monosaccharide residues generates an assortment of *N*-glycan structures, which can be classified into three major classes: high-mannose, hybrid and complex (Figure 1.10). Paucimannose (truncated chitobiose *N*-glycan cores) is a more unusual mammalian *N*-glycan type, but widely expressed in plants and invertebrates [190, 191]. This type of *N*-glycosylation has gained much attention in recent years for its association with pathophysiological conditions such as inflammation and cancer [192].

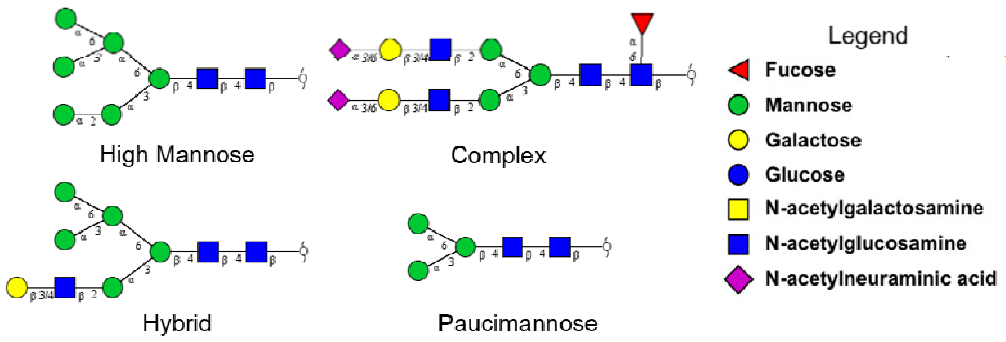


Figure 1.10 Types of *N*-glycans and their linkages

1.3.1.1 Biosynthesis and the endoplasmic reticulum (ER)-Golgi secretory pathway

Extensive metabolic studies in mammalian cell cultures and in yeast have unravelled this seemingly complicated process of *N*-glycan biosynthesis into four distinct stages [193]: (1) the synthesis of lipid-linked glycan precursors, a highly conserved process among all eukaryotes, (2) *en bloc* transfer of the glycan precursors to the Asn-sequons of protein acceptors, (3) early monosaccharide trimming in endoplasmic reticulum (ER), and (4) further processing of the *N*-glycans in the *cis*-, *medial*- and *trans*-Golgi network (Figure 1.11). As these series of steps are tightly coupled to the synthesis of secretory proteins, the process is also known as ER-Golgi secretory pathway.

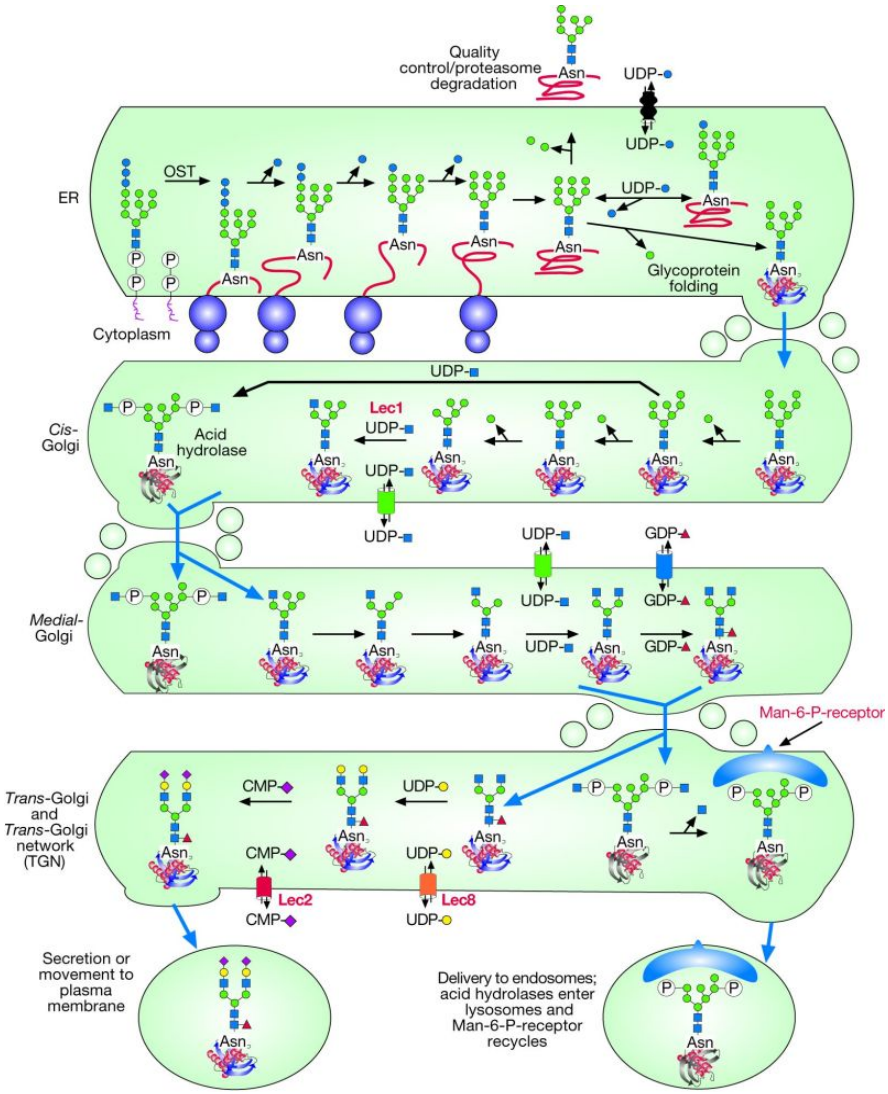


Figure 1.11 Initial synthesis, processing and maturation of human *N*-glycoproteins in the secretory pathway (Adapted from Varki *et al*, 2009 [194])

The assembly of the glycan precursors on the lipid dolichol phosphate carrier begins on the cytosolic face of the ER culminating with the lipid-linked 14-monosaccharide complex glycan precursor with the composition $Glc_3Man_9GlcNAc_2$ (Figure 1.12). The translocation to the luminal side of the ER is mediated by an ER "flippase" enzyme [195]. Secretory, membrane-bound, ER-, Golgi- or endosome-residing proteins are targeted to the secretory pathway by their signal or signal-anchor sequences [196]. The co-translational protein glycosylation modification is initiated by the transfer of the entire oligomannose precursor onto selected Asn-sequons of newly-synthesized polypeptides entering the ER, a process facilitated by the multisubunit enzyme oligosaccharyltransferase (OST). Variable occupancies at the individual Asn-glycosylation sites of glycoproteins give rise to the macro-heterogeneity of glycoforms [197]. Local sequence and topological constraints may influence site-specific occupancy; however there is incomplete understanding of the factors controlling the glycosylation efficiency [187].

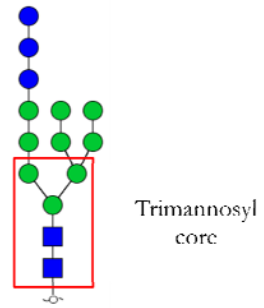


Figure 1.12 *N*-glycan precursor containing the trimannosyl core structure (box).

In the ER, the outer glucose residues of the *N*-glycan precursor are sequentially removed by interactions with the ER chaperones (i.e. calnexin and calreticulin) and glycosidase enzymes (α-glucosidase I and II, ERα(1,2)-mannosidase) to ensure correct folding of the glycoproteins before leaving the ER and into the Golgi apparatus. Trimming continues in the *cis*-Golgi by a series of Golgi-resident α-mannosidases, which removes more mannose residues until a key intermediate ($Man_5GlcNAc_2$) is formed in the *medial*-Golgi. The arrays of glycosyltransferases localized in the *medial*- and *trans*-Golgi act upon this intermediate in a step-wise manner yielding hybrid- or complex-type structures containing up to four antennae extending from the two α-mannoses of

the chitobiose core. The series of enzymatic remodelling eventually lead to the "maturation" of *N*-glycans, which primarily involves the addition of sialic acid, fucose, galactose, *N*-acetylgalactosamine to the non-reducing end (antennas), which in addition may receive another layer of structural complexity by the infrequent addition of sulphate and phosphate to particular monosaccharide residues. Not all high mannose *N*-glycans that enter the Golgi are fully processed but may terminate at any given point in the glycosylation machinery. This introduces an extensive glycan heterogeneity resulting in glycosylated proteins displaying for example, varying numbers of mannose residues ($Man_{5-9}GlcNAc_2$). Thus, multiple *N*-glycan types often appear at a single *N*-glycosylation site due to competing enzymatic reactions giving rise to protein micro-heterogeneity [197]. Several factors can affect the differential processing of *N*-glycans including trafficking rates along the ER-Golgi secretory pathway, the availability of sugar donors and abundance/activity and localization of the modifying glycosyltransferases [198]. These factors are well controlled during cellular homeostasis and growth, development and differentiation and often unique in the individual cell and tissues types, giving rise to cell- and tissue-specific *N*-glycosylation [199]. In addition, cellular systems may further fine tune these expression patterns through the feature of protein or site-specific *N*-glycosylation to express unique sets of glycoforms on individual proteins [200]. The interesting observation of subcellular-specific *N*-glycosylation on secreted and membrane glycoproteins was investigated in this thesis (Chapter 4).

1.3.2 Characterization of protein *N*-glycosylation

In the past few decades, it has become increasingly evident that aberrant protein glycosylation is intimately associated with numerous pathological conditions including many human cancers [201], congenital disorders [202], inflammation [203], diabetes [204] and neurodegenerative diseases [205]. This has prompted many biochemists and glycobiologists to investigate the glycome, which is defined as the entire set of glycans displayed in a specified "system" such as a

cell or organism at a given time under a given condition. The system-wide analysis of the glycome complements the molecular studies of other "omes" including the genome, transcriptome, lipidome, proteome and metabolome. Glycomics research is currently expanding rapidly, covering many aspects of scientific research from basic science and fundamental biology over therapeutic areas to the development and refinement of state-of-the-art analytical technologies with fusion to neighbouring analytical disciplines such as proteomics and glycoproteomics. In cancer research where the attention is focused on identifying unique expression patterns of *N*-glycans, two major objectives have emerged. Firstly, detecting *N*-glycan changes associated with cancer may lead to the identification of candidate *N*-glycan biomarkers of sufficiently high sensitivity and specificity for early diagnosis and monitoring of cancer progression. Secondly, functional glycomics studies may provide insights into the significance of *N*-glycosylation in cellular functions during tumorigenesis. However, the lack of a direct synthesis template or "blue-print" as for the protein equivalent and the structural heterogeneity of *N*-glycans have posed a significant challenge for the identification and structural characterization of *N*-glycans [206] (Figure 1.13).

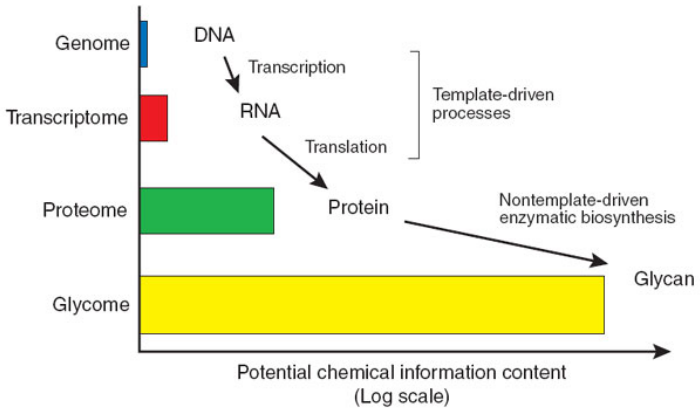


Figure 1.13 The challenges in glycomics research. Unlike the transcriptome and proteome which are based on a genetically encoded template, the glycome is built via nontemplate-driven processes as a secondary gene-product. In addition, the encoded information is significantly enhanced as it flows from the genome to the glycome. (Adapted from Turnbull *et al*, 2007[206])

Until recently, the field of glycomics research lagged far behind the genomics and proteomics disciplines, but recent advances have improved the analytical capabilities to allow for streamlined analysis of the glycome. As such as, the development of robust and relatively high-throughput analytical platforms integrating the use of powerful mass spectrometry has allowed larger-scale characterization of *N*-glycosylation profiles [207], thus aiding to establish its place amongst the other "omics" fields. In recent years, the need for an integrative understanding of the glycoproteome has placed more emphasis on glycoproteomics, which has a significantly higher level of complexity relative to the proteome and glycome alone.

1.3.2.1 LC-MS/MS based structural analysis of the *N*-glycome

By global analysis of *N*-glycans released from mixtures of glycoproteins, *N*-glycome profiling is an approach to capture the *N*-glycosylation status of a biological event. However, this method suffers from a loss of information on the protein origin of the released glycans including site occupancy. In contrast, site-specific glycoprofiling of glycopeptides retains vital information of the carrier protein identity [22]. Such an approach, which is defined as glycoproteomics when preformed on the system-wide level, is often necessary to get a better and more exact understanding of the functional roles of protein glycans.

The liberation of *N*-glycans from glycoproteins is achieved enzymatically by *N*-glycosidase F (PNGase F) treatment, which specifically hydrolyses the amide bond between *N*-glycan and the Asn residue, converting the Asn to an aspartic acid residue in the process. PNGase F is effective on virtually all types of *N*-glycans of the mammalian type but does not release *N*-glycans having chitobiose cores containing α(1,3) linked fucosylation, which are common features of plant *N*-glycosylation. Instead, PNGase A is used to release all plant *N*-glycans [208]. Alternatively, *N*-glycans can be chemically removed by for example hydrazinolysis or *β*-elimination. However, one major drawback is the significant degradation of the protein component in these chemical

reactions. *β*-elimination is frequently used for *O*-glycan release due to limited availability of enzymes for complete *O*-deglycosylation [209].

Historically, *N*-glycomics research focused on identifying monosaccharide compositions of *N*-glycans in the *N*-glycome. With advances in instrumentation for glycan separation, detection and characterization, various *N*-glycan analytical methods have been developed and optimized enabling now complete structural characterization and quantification of *N*-glycan species within a glycome population. Three major approaches have been described and are routinely used for both structural analysis and quantitative glycomics [209] i.e. (1) Reductive amination where the reducing end of glycans is derivatized (and reduced) by labelling with a functional group followed by HPLC analysis and fluorescence detection [210, 211], (2) permethylation of *N*-glycans followed typically by MALDI-MS analysis in the positive polarity mode [212, 213], and (3) glycans can be left underivatized. The reduced *N*-glycans are separated and detected by porous graphitized carbon PGC-LC-ESI-MS in the negative polarity mode [209] (Figure 1.14). This approach has been utilized for the *N*-glycan analysis in this thesis.

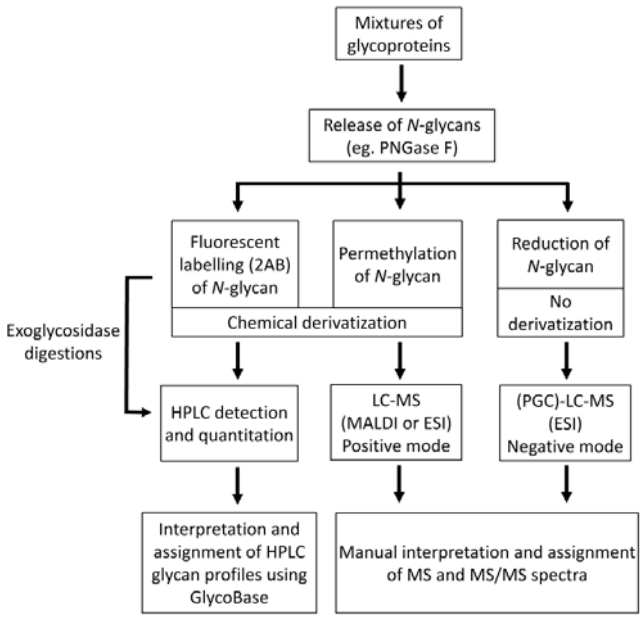


Figure 1.14 The three main approaches in *N*-glycan profiling and characterization.

The advantages and disadvantages associated with each of these methods are presented in Table 1.5 and these are briefly discussed below. For a complex mixture of *N*-glycans, the practice of fluorescent labelling of *N*-glycans at the reducing terminus offers several advantages. Amongst various fluorescent molecules useful for such so-called reductive amination reactions, 2-aminobenzamide (2-AB) is the most widely used fluorophore. The fluorescent tag facilitates and enhances the *N*-glycan detection limit and thereby the sensitivity on HPLC-fluorescence with sensitivity limits in the low femtomole range [214]. In addition, the fluorescence intensity measured from the stoichiometric labelling of all free *N*-glycans in the pool allows for accurate *N*-glycan quantitation [215]. This method is often accompanied by sequential treatment with an array of specific exoglycosidases, which remove specific terminal monosaccharide residues, generating a series of related glycan species from which the glycan sequence can be deduced [216]. However, this validation process is labour intensive as it requires repeated exoglycosidase digestions and reanalysis by HPLC to be performed. A comprehensive database (GlycoBase), containing chromatographic data generated from these reactions has been established to facilitate higher throughput of the structural identification of *N*-glycans [217].

Without doubt, advent of modern MS has revolutionized glycomics research and the approach of using permethylated *N*-glycans analyzed by MALDI-TOF-MS or ESI-MS in positive mode has contributed to this impetus. Permethylation, which methylates all free hydroxyl and carboxyl groups, stabilizes the derivatized glycans, in particular the sialic acid residues which otherwise are prone to premature destruction in the ionization process in MS; it increases the predictability of MS fragmentation to improve interpretation of the monosaccharide sequence and branching glycans; and it allows simultaneous and quantitative analysis of acidic and neutral glycans when the MS detection is performed in the positive mode [218]. In addition, permethylated glycans can be subsequently subjected to gas chromatography MS for linkage analysis [219]. As a result, this

approach has been extensively used to profile and characterize *N*-glycans in various biological samples [212, 219-221].

Although these two glycan derivitization strategies, i.e. 2-AB labelling and permethylation greatly enhance the detection sensitivity of glycans during LC separations, they are still inadequate to resolve and identify isomeric glycans, an inevitable consequence of the extensive *N*-glycosylation micro-heterogeneity, thus limiting the in-depth and accurate characterization and quantitation of *N*-glycan isomers.

Table 1.5 Comparison of the three approaches for *N*-glycan profiling and characterization.

Advantages	Disadvantages
Fluorescent labelling of *N*-glycan (2-AB labelling)	
▪ Increased the detection sensitivity limit ▪ Allows for quantitative analysis ▪ Extensive database for structural assignment	▪ Incomplete derivatization may occur ▪ Labour-intensive work with repeated exoglycosidase digestions for validation ▪ Inadequate to resolve isomers
Permethylation of *N*-glycan	
▪ Increased the detection sensitivity limit ▪ Allows for quantitative analysis ▪ Increased predictability of MS fragmentation ▪ Allows for linkage analysis	▪ Incomplete derivatization may occur ▪ Inadequate to resolve isomers
Reduction of *N*-glycan	
▪ Forms alditol (eliminates the anomericity of reducing end of glycan) ▪ Easy to perform	▪ Slight under-derivatization may occur and loss of labile glycan PTMs during reduction

In contrast, PGC efficiently separates isomeric glycans in a reproducible manner [222]. The performance of PGC for *N*-glycan separation has been investigated and compared to those of other types of chromatography such as hydrophilic interaction liquid chromatography (HILIC) and reversed-phase chromatography (RPC) and found to have superior peak capacity allowing efficient separation of *N*-glycans [223]. When coupled to tandem MS in the negative mode, PGC provides a tool for the very detailed characterization of complex mixtures of glycans [224] with the ability to differentiate between many isobaric glycan epitopes/determinants, e.g. terminal α2,3- and α2,6-sialylation; between core α1,6- and antenna α1,2/3/4-fucosylation; between

terminal motifs such as Gal-GlcNAc (LacNAc) and GalNAc-GlcNAc (LacdiNAc); and between bisecting GlcNAc and non-bisecting GlcNAc residues [225]. The exact interaction mechanism(s) between the glycans and the hexagonal carbon atoms in the PGC stationary phase remains unclear; however, the interactions have been established at low resolution to be of mixed mode consisting partially of hydrophobic, electrostatic, and hydrophilic (dipole-dipole interactions) [226]. In order to reduce the added complexity of α and β anomers formed by the anomericity switching at reducing end of all free glycans due to their separation by the high resolving power of PGC, the reducing ends of *N*-glycans are routinely converted to free sugar alditols by simple sodium borohydride based reduction enabling single chromatographic peak detection on PGC-LC. Using this approach, several studies have successfully characterized complex mixtures of released *N*-glycans from glycoproteins extracted from various biological sources including cell lines, tissues and secreted bodily fluids such as saliva and milk [227-232]. One disadvantage of this approach is the limited availability of computational tools for high throughput data analysis of the information-rich MS/MS spectra. Thus, laborious manual *de novo* interpretation is still required for detailed assignment of structures.

Most comparative glycomics studies are based on relative quantitation of investigated glycomes, which is achieved by normalizing individual glycan structures within each glycomic profile and compared across different samples. Such comparison is commonly used although low abundant glycans may not be easily quantified if there is large variation in the whole glycan profile. Absolute quantitation of *N*-glycan can be achieved by spiking glycan samples with fluorescently-labelled glycan standards. Such measurement is desirable as each glycan quantified would be independent of the variations in the whole profile, however, this approach is currently not well established. Hence, relative glycan quantitation remains to be widely used, which also was the quantitation method applied in this study.

Notwithstanding, bioinformatics tools are clearly needed to integrate and automate interpretation of the vast amount of glycomics data being generated from these glycomics technology platforms. To meet such demands, several large scale initiatives such as Consortium for Functional Glycomics (CFG) [233], KEGG [159] and recently, the UniCarbKB [234] were set up to provide integrated resources to glycoscientists. These resources include web-based tools such as GlycoMod [235] for predicting monosaccharide compositions of glycans based on MS (precursor) data; databases such as GlycoSuiteDB which contains manually curated glycan structural information as derived from the literature [236]; KEGG-GLYCAN which maps glycan data to known molecular interactions and pathways [237] and GlycoWorkbench, which consists of a suite of software tools useful for drawing glycan structures and annotating experimentally-derived mass spectra [238]. However, there is still a lack of bioinformatics tools for the high-throughput handling of large MS/MS datasets. Interpreting the MS/MS spectra currently relies on *de novo* approaches, which is a tedious process. Sophisticated computational algorithms for database matching to experimentally-derived MS/MS spectra are currently being developed to overcome this major bottleneck in glycan analysis.

1.3.3 Protein *N*-glycosylation changes in breast cancer

We have come to understand that in protein *N*-glycosylation, the intricate organizational interplay of glycosylation enzymes including glycosidases and glycosyltransferases creates an array of highly complex and related glycan structures on proteins. It is thought that such structural diversity facilitated by the *N*-glycans on membrane-bound and secretory proteins is essential to carry out their divergent biological functions including cell proliferation, differentiation, migration, cell-cell integrity and recognition, cell-matrix and host-pathogen interactions, immune modulation and signal transduction [201]. Aberrant protein *N*-glycosylation can therefore disrupt normal cellular functions leading to lack of cellular homeostasis and pathophysiological conditions.

Early studies using cultured breast cancer cells and breast tumor tissues identified specific *N*-glycan changes that correlate well with breast tumorigenesis, thus suggesting the involvement of *N*-glycans in breast cancer [239, 240]. Before the advent of more advanced MS technologies, a variety of glycan detection methods were employed to compare the glycosylation patterns between normal and breast cancer samples. These investigations were targeted visualizing glycan epitope-changes rather than modern approaches measuring the detailed structural glycome changes in MS-based glycomics studies. Plant lectins, which have reactivity for a wide range of glycan determinants, are the most commonly exploited tool for visualization/detection of glycan epitopes and have been used in several ways to reveal differential *N*-glycoepitope expression including lectin histochemical staining [241, 242], lectin affinity chromatography [243], lectin blotting [244], and lectin array [245]. When integrated with MS platforms, the value of lectins was clearly demonstrated by the ability to identify the proteins that carry tumor-specific glycan epitopes [212]. Lectins are thus useful to isolate and visualize glycoproteins in their intact forms whereas MS may facilitate the identification of the protein carrier. Another detection method, immunohistochemical (IHC) staining, uses antibodies to target glycan-associated antigens on breast tumor tissues, which positively correlated with the increased metastatic potential and poor prognosis of breast cancer patients [239, 246]. When combined with well-designed controls, approaches using lectin and antibodies are able to provide valuable information such as glycan topology, cellular localization and relative abundance [194]. Overall, glycan epitope-detection methods (i.e. lectins and antibodies) and whole structure characterization-(i.e. HPLC, LC-MS) have identified consistent alterations in the *N*-glycan expression patterns in breast cancer. The aberrant *N*-glycosylation involves a relative increase in sialylation, fucosylation, β1-6 branching and Lewis-type epitopes such as Lewis X (Le^X), Lewis Y (Le^Y), sialyl Lewis X (sLe^X) and sialyl Lewis A (sLe^A). Differential detection using fluorescently-labelled lectin and antibody staining is generally quantified by the absolute fluorescence intensity, while in global glycan profiling, relative abundances between glycans within glycome populations are reported and the

glycoprofiles are then compared between samples. As such, an increase in a glycan determinant within a glycome population, e.g. complex type is naturally accompanied by the decrease of other glycan types, e.g. high mannose and hybrid types.

In recent years, there has been a growing interest in understanding pathology-driven glycome changes at the system-wide level, which encompasses studies to investigate the relationship between the genome, transcriptome, gene product (i.e. proteome and glycome) and glycosylation enzyme activity [247-249]. The following sections detail these molecular changes in the context of breast cancer. Cell surface protein glycosylation changes in breast cancer has also been described as part of a published review, attached at the end of Chapter 1.

1.3.3.1 Sialylation

Sialic acids belong to a large family of nine-carbon α-keto acids known collectively as nonulosonic acids. The two predominant forms of sialic acid residues in mammalian cells are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) [250] (Figure 2.15). Although Neu5Ac differs from Neu5Gc only by a single hydroxyl group, an irreversible mutation in the human gene encoding for the enzyme producing the Neu5Gc nucleotide donor have eliminated its expression in humans, thus limiting its production to non-human mammals [251]. However, minute quantities of Neu5Gc detected in normal human tissues and at somewhat higher levels in some human cancer tissue possibly by incorporation from exogenous (nutritional) sources, suggest their possible roles as cancer biomarkers [252].

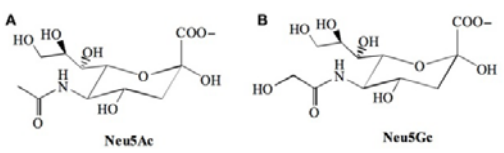


Figure 1.15 The chemical structures of Neu5Ac and Neu5Gc.

The addition of sialic acids (herein mainly Neu5Ac) to *N*-glycans of glycoproteins is catalyzed by different families of sialyltransferases, which are classified according to the carbohydrate linkage being formed from the sialic acids to the penultimate residues, e.g., β-galactoside α-2,6-sialyltransferases (ST6Gal) for α2,6-linked sialylation, β-galactoside α-2,3-sialyltransferases (ST3Gal) for α2,3-linked sialylation and α2,8-sialyltransferases (ST8Sia) for α2,8-linked sialylation (Figure 1.16). Increased expression of either α2,3- or α2,6-sialylated *N*-glycans have been observed in breast tumor tissues, breast cancer cell lines and breast cancer serum compared to healthy tissue [253-255]. In general, total sialylation is enhanced in breast malignancy although the linkage-specific expression differences remain undocumented. Sialylated Lewis epitopes including sLe^a and sLe^x were reported to be over-expressed in tissues and serum of breast carcinoma patients relative to healthy donors; their expression correlated with the increased metastatic potential of the cancer and the reduced patient survival [211, 256-259]. In contrast, the presence of the less common α2,8-polysialic acid has only been detected in the MCF7 breast cancer cell line. The two proteins known to specifically carry the α2,8-polysialic acid chains are neural cell adhesion molecule (NCAM), which is associated with nervous system development and plasticity [260], and the rat brain voltage-dependent sodium channel α subunit [261]. In an effort to better understand the mechanism of *N*-glycosylation deregulation, a transcriptomics study revealed an increased occurrence of the corresponding sialyltransferases [262]. In addition, this correlated with poor patient outcome, suggesting their clinical value as prognostic marker [263].

Sialic acids are commonly expressed on mammalian glycoconjugates i.e. glycoproteins, glycolipids and proteoglycans. Sialylation is a dominating feature on cell surface and secreted *N*-linked glycoproteins compared to the intracellular *N*-glycoproteome suggesting its involvement in extracellular biological functions [200]. Hybrid and complex type *N*-glycans of secretory nature, for example plasma glycoproteins, are often capped by sialic acid residues, serving to mask the underlying galactose residues from recognition by the liver asialoglycoprotein receptor, thus

extending the circulatory half-life [264]. The elevated expression of sLe^x in breast cancer patients and their strong correlation to advanced stages of the disease are often reported in serum-based studies [211, 259]. Serum measurements of sLe^x and CA15-3, which is currently a clinical breast tumor marker with relative low specific and sensitivity, has been suggested to improve the prognostic features (i.e. sensitivity) when monitoring breast cancer [265].

The structural and chemical properties embedded within the large nine-carbon sialic acid molecule impart the potential for generating multiple levels of diversity, allowing the sialylated glycans to mediate various significant biological roles including immune responses, cellular recognition, adhesion and signalling [22]. The exposed terminal localization of sialic acid residues of glycoconjugates is a natural disposition to interact with other biomolecules, in particular the endogenous and exogenous glycan-binding proteins such as the family of sialic acid-binding lectins known as siglecs. Siglecs are important molecules for regulating the cell-cell signalling to facilitate a functional immune response [203]. In healthy individuals, such interactions are tightly controlled to attenuate immune responses mitigating the effects of inflammation [266]. However, hyper-sialylation on cancer cells may allow such cells to escape immune surveillance [267] and may also contribute to the invasive and metastatic behaviour of cancer cells [268]. Specifically, altered sialylation displaying higher levels of α2,6-sialic acid on the cancer cell surfaces were linked to increased motility and invasive potential of breast tumor cells [254]. Although the role of altered sialylation in breast cancer is evident, there remains a lack of mechanistic understanding of their involvement in tumor metastasis.

1.3.3.2 Fucosylation

Increase in fucosylation in human cancers is well documented and may be a general glyco-phenotypic hallmark associated with malignancy [269]. Fucosylated *N*-glycans are synthesized by a wide range of human fucosyltransferases (FUT1-11) and can be broadly categorized into core-

and antenna-fucosylated glycans (Figure 1.16). The most common modification of the innermost GlcNAc residue of the *N*-glycan chitobiose core is the α1,6-linked core fucosylation catalyzed by α1,6-fucosyltransferase 8 (FUT8). The link between core fucosylation and antibody-dependent cellular cytotoxicity (ADCC) is well known. Antibodies lacking core fucosylation display a higher affinity for the Fc receptors on immune cells leading to enhanced ADCC [270]. Such modulation has important clinical implications for Trastuzumab, a monoclonal antibody used for therapeutic treatment of breast cancer patients over-expressing the human epidermal growth factor receptor 2 (HER2) [271]. In contrast, higher core fucose content has been found on the abundant serum glycoprotein α-1-proteinase inhibitor in breast cancer [272] which was supported by the detection of an up-regulated *FUT8* transcripts in breast tumor tissues [247]. The role of increased core fucosylation on glycoproteins in breast tumorigenesis remains unclear and needs further investigation.

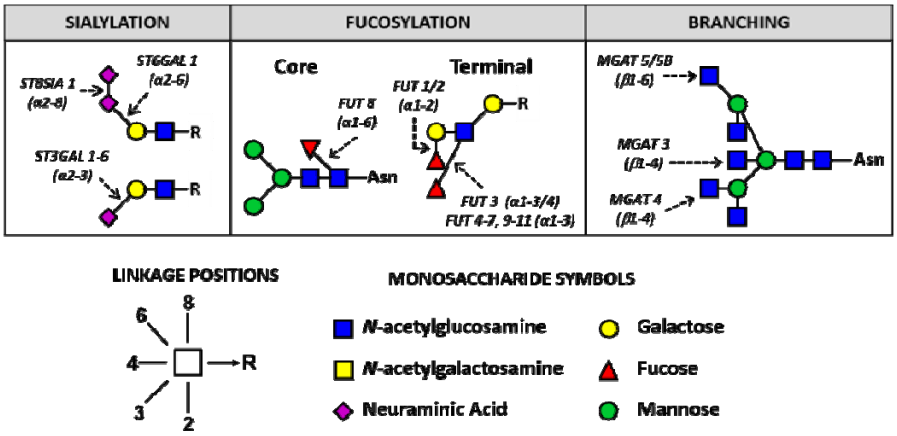


Figure 1.16 Diagram illustrating the glycosyltransferase genes involved in the *N*-glycan changes observed in breast cancer. (Adapted from Christiansen *et al*, 2013 [201])

Fucosylation may also be linked to the antenna of *N*-glycans via α1,3- and/or 4 linkages by the action of multiple fucosyltransferases including FUT1-7 and FUT9. Such additions give rise to the formation of specific Lewis blood group antigens such as $Le^{x/y}$ and $Le^{a/b}$. When α1,3 or α1,4-fucosylation occurs on antennas carrying an α2,3-sialic acid, they become part of the sialylated

and fucosylated glycoepitopes/determinants known as sLe^x and sLe^a, respectively. Many studies have demonstrated a clear pathological relationship between fucosylated Lewis antigens and the invasiveness of breast tumors [246, 273-275]. The pivotal regulatory role of fucosylation in breast cancer is manifested in the sialyl Lewis antigens as a binding ligand for E-selectins, which are expressed on endothelial cells in blood vessels [276]. Current proposition on extravasation of tumor cells at a metastatic site suggests the involvement of sLe^x antigens in a selectin-dependent manner for the adhesion and invasion similar to the mechanisms of leukocyte transendothelial migration [277]. Numerous corroborating lines of evidence are in support of this hypothesis. Reports of over-representation of sLe^x and sLe^x epitopes on metastatic breast tumor cells are well-documented [257, 258, 278] and they were shown to be critical determinants in the adhesion of tumor cells to vascular endothelium [279]. Introduction of the *FUT4* gene into the breast cancer cell line MCF7 markedly induced the expression of sLe^x and enhanced attachment of cancer cells to the endothelial cells [280]. In hormone-dependent breast cancers, the adhesion was demonstrated to mediate through E-selectin and sLe^x interactions to promote cancer metastasis [249]. Breast carcinoma-associated glycoproteins that have been identified as carriers of Lewis types determinants include CD44 [281], CD98hc [282], CD147 [283] and podocalyxin [284]. Taken together, the regulatory role of fucosylation is firmly established in breast tumorigenesis. Such knowledge could be further exploited to seek new diagnostic methods to ultimately improve the range of therapeutic options and clinical outcome of breast cancer patients.

1.3.3.3 Branching and bisecting GlcNAc

The glycosylation enzyme, *N*-acetylglucosaminyltransferase V (GnT-V), which is encoded by *MGAT5*, catalyses the formation of β1,6-GlcNAc branches on *N*-glycans (Figure 1.16). Using lectin-based methods, consistently higher levels of β1,6 branching of the cell surface *N*-glycans have been detected in breast tumor tissues when compared to normal tissues [212, 241, 242, 285].

These observations were supported by concomitant increases in the GnT-V levels [286] and its corresponding *MGAT5* [287]. Knockdown studies that dampened the expression of GnT-V inhibited cells dissolution from the ECM and the subsequent spreading of cancer cells, thus demonstrating that β1,6 branching may drive the migratory and metastatic phenotype in breast tumor cells [288, 289]. Some studies have used the strong implications of β1,6-branching in breast malignancy to propose that *N*-glycan branching may be a predictive marker for the identification of node-negative breast cancer [241, 242].

The role of β1,4-bisecting GlcNAc in breast cancer has not been extensively investigated. Bisecting GlcNAc epitopes are synthesized when a β1,4-linked GlcNAc residue is attached to the *N*-glycan core by *N*-acetylglucosamintransferase III (GnT-III). In breast tumors, the expression of the responsible *MGAT3* gene was down-regulated compared to normal breast tissues [247]. The authors suggested anti-tumor features of bisecting GlcNAcylation by reasoning that the addition of the bisecting GlcNAc to the core may prevent other types of *N*-glycan branching to form, for example the tumor associated β1,6-linked GlcNAc as described above.

1.3.3.4 High mannose *N*-glycans

Total cellular glycoproteins extracted from the human ovarian carcinoma cell line, SKOV3 predominantly displayed high-mannose type *N*-glycans [290]. Increased expression of high-mannose type *N*-glycans on cell surfaces of various tumors including breast, colorectal, lung, cervical, ovarian and lymphatic cancers have been reported [291]. These *N*-glycans were released from membrane proteins isolated using conventional ultracentrifugation method, however, a recent study showed that ultracentrifugation was inefficient in enriching for cell surface proteins, but instead predominantly capture intracellular membrane glycoproteins [292]. Evidence for this observation is substantiated in Chapter 4 where the proteomics of such preparations showed that many ER- and Golgi-residing proteins were co-purified along with plasma membrane proteins.

Therefore it cannot be ruled out that intracellular glycoproteins contributed to the high mannose-rich patterns observed in such cancer studies. Nevertheless, expression of high-mannose type *N*-glycans carrying nine mannose residues, which is normally considered to be an indicator of relative immature and intracellular *N*-glycan, were elevated in the sera of breast cancer patients and in breast tumor mice models relative to healthy references thereby indicating that these unprocessed glycoconjugates are indeed extracellular [293]. Additionally, immature high-mannose structures have been detected on cell surface glycoproteins including intracellular adhesion molecular 1 and the oncogenic form of epidermal growth factor receptor (EGFR) [294]. As is the case for many of the other glycan structures and glycoepitopes, it remains unknown whether these alterations are causing or a consequence of tumorigenesis. Thus, the exact significance of the presence of high-mannose type *N*-glycan in breast malignancy remains to be determined.

1.4 Aims of the thesis

The major aim of this thesis was to apply state-of-the-art proteomics and glycomics analytical technologies, in conjunction with sophisticated bioinformatics tools, to gain insights into the molecular alterations associated with breast cancer. The investigated samples comprised of a panel of cultured human breast epithelial cells. A total of six breast cancer cell lines were used; five were established from metastatic cells obtained by the pleural effusion and one derived from primary breast tumors. The five cancer cell lines were representative of three common breast cancer subtypes, MCF7 for luminal A, SKBR3 for HER2-enriched, MDA-MB-468 (MDA468) for basal-A, and MDA-MB-157 (MDA157), MDA-MB-231 (MDA231) and HS578T for basal-B subtypes. Two non-tumorigenic cell lines were used including the human mammary epithelial cells (HMEC), which was derived from normal breast tissues and MCF10A, an immortalized cell line originated from the mammary gland of a patient with fibrocystic disease. Minor focus was given to the development and optimization of the multi-lectin affinity chromatography (M-LAC)

methodology for fractionation/enrichment of cancer-specific glycoproteins derived from breast cancer cells. To achieve these aims, the following studies were conducted:

1. Comparative global profiling of secreted and membrane proteins extracted from four breast epithelial cell lines (HMEC, MCF7, SKBR3 and MDA231) to identify differentially expressed and unique proteins in breast cancer cells. Global and subtype-specific functional analyses were performed on the subset of significantly regulated proteins to elucidate cancer-related pathways and protein-protein interaction networks.
2. Global *N*-glycan profiling and structural characterization of secreted and membrane fractions of six breast epithelial cell lines (HMEC, MCF7, SKBR3, MDA157, MDA231 and HS578T). Global and subtype-specific comparative analyses were carried out to identify differentially expressed *N*-glycan determinants.
3. Performing systematic investigation of subcellular-specific *N*-glycosylation of cell surface, secreted and microsomal fractions extracted from three breast epithelial cells (MCF10A, MCF7 and MDA468) by using a combination of structural knowledge, computational and analytical tools.
4. Optimizing the multi-lectin affinity chromatography platform, comprising of a combination of three lectins (Con A, WGA and Jac), to enrich for glycoproteins from the cell lysates of MCF7 with the aim for future application to target tumor-specific glycoepitopes present in complex biological samples.

Publication I - Cell surface protein glycosylation in cancer (Review)

Pages 65-86 (Publication 1) of this thesis have been removed as they contain published material under copyright. Removed contents published as:

Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L. and Packer, N. H. (2014), Cell surface protein glycosylation in cancer. *Proteomics*, vol. 14, no. 4-5, pp. 525-546. https://doi.org/10.1002/pmic.201300387

CHAPTER 2

FUNCTIONAL ANALYSIS OF PROTEOME CHANGES IN BREAST CANCER

2.1 Introduction

Molecular profiling of breast cancer has successfully characterized the disease into various subtypes with distinctive pathological features and clinical outcomes [7, 295]. The remarkable heterogeneity of breast cancer tumors underscores the importance of identifying molecular signatures specific to each subtype to aid the development of targeted therapies. Tumorigenic transformation in breast epithelial cells has correlated well with protein expression changes [296]. Global differential protein analysis, i.e. mapping and comparing all proteins between different samples, could unravel protein signatures that determine the biological and functional characteristics associated with each breast cancer subtype, providing us with better insights into the underlying molecular mechanisms involved. However, such function-based global analysis approach to examine different breast cancer subtypes has not been widely undertaken by those investigating the proteome-wide changes in various breast cancer cells [51, 99, 101].

Both secreted and membrane proteins are involved in key biological processes such as cell-cell communication, transportation of molecules, enzymatic activities, cellular adhesion and immune response. During malignant transformation, cells secrete various effector molecules into the extracellular space that promote cellular migration, invasion, adhesion and matrix degradation [297]. More than 50% of membrane proteins are potential drug targets [298]. Therefore, analysing the secreted and membrane proteomes is a promising approach to identify potential cancer biomarkers and drug targets.

In this study, we utilized a shotgun proteomics method to investigate the subcellular specific proteomes of four cultured breast epithelial cells. Three well-characterized breast cancer cell lines established from the pleural effusion representative of each breast cancer subtypes were selected,

namely, MCF7 for luminal A subtype, SKBR3 for HER2-enriched subtype and MDA231 for basal B subtype. The primary human mammary epithelial cells (HMEC) served as a normal reference for comparison with the three breast cancer cell lines.

Although comparative analysis performed between paired tumor and normal non-tumorigenic tissues may represent the most suitable comparison for understanding tumorigenesis and for downstream clinical applications, tissue heterogeneity in terms of cellular and molecular composition remains a significant challenge. Hence, cultured cells were used in this study as they constitute a homogeneous population of epithelial cells, which allow for interrogation of breast epithelial cancer cell-specific proteins without contamination from other cell types such as stromal cells, adipocytes, endothelial cells or immune cells In addition, breast cancer cell lines recapitulate the subtype classification observed *in vivo* in breast tumors, making them suitable models for studying breast cancer subtypes [134]. In total, we identified more than 3,000 secreted and membrane proteins from the conditioned media and the enriched membrane of four breast epithelial cells. Accordingly, we subjected these identified proteins to a system-wide functional analysis to gain insight into the molecular events underlying breast tumorigenesis. The functional analysis of differentially expressed proteins in the three breast cancer cell lines revealed common functional features involved in breast cancer biology, including abnormal activities associated with the proteasomes, translation initiation factors, cytoskeletal proteins and in the extracellular matrix (ECM). Functional analysis of the proteins specific to each of the three breast cancer subtypes revealed that the G protein-coupled receptor GPCR signalling pathway was activated in the three breast cancer subtypes, but that this pathway involved different sets of proteins. Importantly, this approach identified a number of proteins that were central to the altered biological processes or pathways and which could serve as potential cancer biomarkers or targets for future cancer therapy.

2.2 Materials and methods

2.2.1 Cell cultures and sample preparation under serum-free conditions

Human mammary epithelial cells (HMEC) were purchased from Lonza (CC-2551, Walkersville, MD). Human breast cancer cell lines MCF7, SKBR3 and MDA231 were obtained from American Type Culture Collection (Manassas, VA). HMEC was grown in HuMEC Ready Media (Invitrogen, CA). The other three cell lines were grown in RPMI (Sigma, MO) supplemented with 5% FBS (Invitrogen, CA), 10 mM glutamine (Invitrogen, CA) and 10 μg/mL insulin. Cells were maintained at 37°C in 5% CO_2 for all experiments. The breast epithelial cell lines were grown in triplicates to around 80% sub-confluency and washed at least four times with ice-cold PBS to remove traces of FBS and incubated in serum-free media at 37°C in 5% CO_2 for 48 hours. Cell viability was determined by trypan blue exclusion assay after 48-hour incubation, Conditioned media (CM) containing the serum-free secreted proteins were collected, followed by centrifugation at 2,000 x g to pellet any floating cells. Supernatant was then concentrated and buffer exchanged with PBS (1x) using Amicon Ultra centrifugal filter devices with a 10,000 MW cut-off membrane (Millipore, MA). Proteins were then precipitated with acetone overnight at -20°C and stored at -80°C until further analysis. Following removal of serum-free media, cells were washed with PBS (1x) and harvested in Tris buffer containing 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA and protease inhibitors cocktail (Roche Diagnostics). The cell suspensions were ultra-sonicated (Branson Sonifier 450) on ice for 3 rounds of 10 s and centrifuged at 2,000 x g for 20 min at 4°C to remove intact cells and nuclei. The supernatant was ultra-centrifuged at 120,000 x g for 80 min after which the supernatant was discarded. The microsomal membrane pellet was washed twice with ice-cold 0.1 M sodium carbonate and resuspended in 25 mM Tris-HCl pH 7.4, 150 mM NaCl and 1% (v/v) Triton X-114. Samples were subjected to phase partitioning by incubation at 37°C for 20 min, followed by 1,000 x g centrifugation for 10 min. The upper aqueous layer was carefully removed and 9 volumes of ice-

cold acetone was added to the lower detergent phase and incubated overnight at -20°C to precipitate the proteins.

The total protein concentration of the subcellular fractions from breast cells was measured using Bradford reagent (Sigma, MO). Equal amount of total protein in each subcellular proteome was used for precipitation followed by solubilization in NuPAGE LDS sample buffer to prepare protein samples for gel electrophoresis.

2.2.2 Gel electrophoresis of subcellular proteomes and in-gel digestion

Proteins were separated using 1D gel electrophoresis (SDS-PAGE) and each lane was sliced in eight fractions which, after in-gel digestion, were analyzed by LC-MS. This analytical approach (GeLC-MS) was chosen to provide an additional orthogonal platform for protein separation, which has been shown to achieve in-depth protein identification [299]. SDS-PAGE also has the advantages to improve protein solubility and to render samples more compatible with LC-MS by the option for removing salts, buffers and detergents in the in-gel digestion step.

Approximately 50 μg of membrane proteins and 20 μg of secreted proteins were reduced with 50 mM of dithiothreitol for 10 min at 70°C and alkylated with 125 mM iodoacetamide in the dark at room temperature for 30 min. Each sample (10 μL), in NuPAGE LDS buffer, was loaded on 4-12% Bis-Tris PAGE gel (Invitrogen) and electrophoresis was performed at 200 V for 50 min. After separation of proteins, the gel was fixed in 40% (v/v) ethanol and 10% (v/v) acetic acid for at least 2 hours and stained overnight with Coomassie Blue G250, and destained with Milli-Q water (Millipore).

To perform in-gel trypsin digestion each lane was cut into 8 segments of equal size. Each segment was further sliced into 1 mm smaller pieces and placed in a 96-well plate. The gel pieces were destained with 50% (v/v) ACN in 50 mM ammonium bicarbonate until they became clear. They were then dehydrated in 100% ACN and dried. Trypsin (sequencing grade Modified, Promega) was added at a weight ratio of 1:30 to digest the proteins overnight at 37°C. The next day, the tryptic peptide mixtures were collected and two more extractions were performed with 2% (v/v) formic acid in 50% (v/v) ACN and 50 mM ammonium bicarbonate. All three extracted fractions were combined and the solution was dried by vacuum centrifugation. Tryptic peptides were acidified in 10 μL 0.1% (v/v) formic acid and desalted. Briefly, C18 tips were washed three time with 20 μL 100% ACN, three times with 20 μL 50% (v/v) ACN in 0.1% formic acid, and equilibrated with 50 μL 0.1% (v/v) formic acid. After sample loading, tips were washed three times with 20 μL 0.1% formic acid. Peptides were eluted by 20 μL 60% (v/v) ACN in 0.1% formic acid and 20 μL 90% (v/v) ACN in 0.1% formic acid and dried. The samples were at stored at -80°C until used for LC-MS/MS analysis.

2.2.3 LC-MS/MS-based proteomics

Triplicates LC-MS/MS injections from all peptide mixtures derived from extracted membrane and secreted proteomes of the breast epithelial cell lines were performed using a Q-Exactive Orbitrap (ThermoFisher) mass spectrometer. Tryptic peptide mixtures in 0.1% (v/v) formic acid were loaded onto an in-house packed RP column (2.7 μm Halo C18 resins, 100 mm x 75 μm). Separation of peptides was performed over 60-min gradient with the first 50 min linear gradient increasing from 0-50% in solvent B (0.1% (v/v) aqueous formic acid in ACN) and up to 85% in solvent B for the next 2 min and maintained at 85% for 8 min. The flow rate was set at 300 nl/min. The nanoLC system was connected directly to the nanoESI source of the mass spectrometer. MS and MS/MS spectra were acquired with resolution of 35,000 in the positive

polarity mode and over the range of m/z 350 – 2000. Automated peak recognition, dynamic exclusion, and tandem MS of the top 10 most intense precursor ions were performed using Xcalibur v2.2 (ThermoFisher). Yeast enolase was routinely used between samples as quality control.

2.2.4 Protein identification

Raw LC-MS/MS data files were converted to MGF format using Proteome Discoverer (v2.0) and searched against SwissProt protein database (*Homo sapiens*, 20,279 reviewed entries) using the global proteome machine (GPM, Cyclone version). The following criteria were used during the search: carbamidomethylation of cysteine residues was set as a fixed modification and oxidation of methionine and deamidation of asparagine and glutamine residues were used as variable modifications. Mass tolerances of 10 ppm and 0.02 Da were selected for precursor and MS/MS fragment ions, respectively, with a maximum of two missed trypsin cleavages.

Scaffold (v4.2.1, Proteome Software) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability assigned by the Protein Prophet algorithm incorporated in the software. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide homology were grouped into clusters of protein families. Proteins were annotated using GO terms from NCBI.

2.2.5 Label-free quantitation using normalized spectral abundance factor

Normalized spectral abundance factor (NSAF) was calculated based on the following formula: NSAF = (Spc/L)/$\sum$(Spc/L), where Spc refers to the spectral count (number of non-redundant

peptide identifications for a given protein) and L is the length of the protein in amino acid residues [300]. Protein identifications were only included in NSAF data analysis if a given protein were covered by a minimum of two peptides in at least one of the three technical replicates and contained at least a total of four spectral counts across all replicates. For comparative analysis using fold change as a measure for protein regulation, only proteins that were present in both of the compared samples were included. The fold change of a protein was calculated by the ratio of its NSAF across different samples.

2.2.6 Statistical analysis and bioinformatics

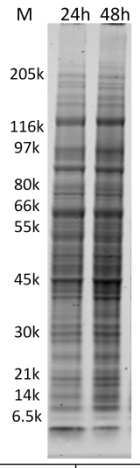
Statistical analyses were conducted using SPSS software (v22). One-way ANOVA analyses were performed for proteins displaying a minimum of a three-fold change between each of the cancer cells and the HMEC reference followed by post-hoc Dunnett's test. All P values were adjusted taking into account the multiple comparisons made and reported as multiplicity adjusted P values, where a value of less than 0.05 was regarded as statistically significant. GO annotation and functional analyses were performed using the open source program Cytoscape (v3.1.1) http://www.cytoscape.org/). Statistical test for enrichment or depletion was based on a two-tailed hypergeometric test and corrected for multiple testing using Bonferroni [301]. Protein interaction networks were performed using STRING (v9.1) (http://string-db.org/). Hierarchical clustering analysis was performed using an in-house program written in R.

2.3 Results

2.3.1 Optimization of cultured cells for proteomics analysis

In order to profile secreted proteins in the conditioned media of the four breast epithelial cell lines, it was essential to use serum-free media (SFM) to ensure no exogenous proteins from the FBS were included in the analysis of the secretome. Prior to SFM incubation, cells were adapted to growth in media containing a reduced amount of FBS (< 5% v/v) so as to minimize

deleterious effects on cell growth induced by rapid serum starvation. The confluency of cultured cells was checked and at ~70% sub-confluency the media were removed and cells washed at least three times with PBS before replacing the cells in SFM. Preliminary studies using conditioned media (CM) of MCF7 cells collected after 24 hours and 48 hours post-SFM incubation showed no significant difference in the protein expression patterns as evaluated by the protein patterns on SDS-PAGE between the two time points. Increased total cell counts and higher protein levels were observed in the 48-hour CM (Figure 2.1). Cell viability as measured by trypan blue exclusion remained above 90% for both time points indicating minimal cell death. The 48-hour SFM incubation time was thus chosen to maximize the protein concentration in the CM. After the CM was collected, cells were harvested and membrane proteins (microsomal fraction) extracted using ultracentrifugation followed by Triton X-114 phase partitioning. Both subcellular fractions (i.e. secreted and microsome) were fractionated on SDS-PAGE and each lane cut into eight equal-sized fractions. The gel fractions were trypsinized and the resulting tryptic peptides were analyzed using LC-MS/MS to obtain the secreted and membrane subcellular proteomes of each of the investigated cultured breast cells.



	24hr	48hr
Total cell count	3.7×10^6	6.5×10^6
Viability	96%	92%
Protein amt (µg)	115	187

Figure 2.1 SDS-PAGE gel of MCF7 CM proteins sampled after 24 and 48 hours of incubation in SFM. Total cell counts, viability and protein amounts were also measured at the two time points.

2.3.2 Identification of secreted and membrane proteins in HMEC, MCF7, SKBR3 and MDA231

Shotgun proteomics is a powerful analytical tool for system-wide proteome analysis, depicted in Figure 2.2. We applied this approach to investigate the global cellular profiles of secreted and membrane proteins in four breast epithelial cell lines. HMEC is a non-tumorigenic breast epithelial cell line while the other three breast cancer cell lines represent the three common breast cancer subtypes, namely, luminal A (MCF7), HER2-enriched (SKBR3) and basal B subtype (MDA231). Three technical replicates of the individual cell lines were performed. By applying strict criteria for peptide and protein identification (see Section 2.2.4 – Materials and Methods for details), we confidently identified a total of 1,755 and 2,063 non-redundant proteins in the secreted and the membrane fractions of the four cell lines, respectively (Table 2.1). The confidence of the protein identifications as measured by the false discovery rate (FDR) was less 1% for all samples. Of the proteins identified in the two subcellular proteomes, i.e. secreted and membrane fractions, 34% to 42% have at least five unique peptides in each of the replicates.

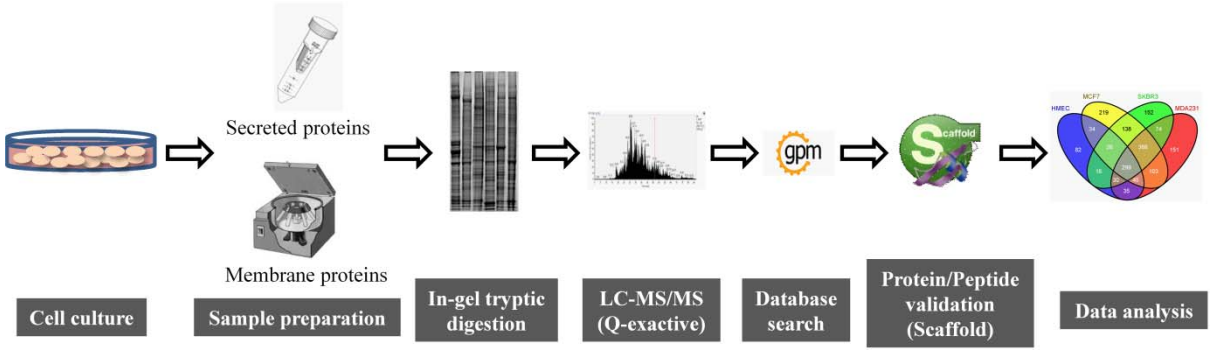


Figure 2.2 Workflow illustrating the sample preparation and the proteomic analyses of secreted and membrane protein fractions from the investigated breast epithelial cell lines.

Table 2.1(a) Summary of the number of proteins identified in the secreted subcellular proteome.

Cell line	Number of proteins identified with two or more unique peptides (UP)*				Total number of non-redundant proteins	FDR (%)
	2 UP	3 UP	4 UP	≥ 5 UP		
HMEC	114	109	78	232	558	0.49
MCF7	267	228	163	561	1219	0.53
SKBR3	261	174	166	487	1088	0.57
MDA231	243	203	165	483	1094	0.49

Table 2.1(b) Summary of the number of proteins identified in the membrane subcellular proteome.

Cell line	Number of proteins identified with two or more unique peptides (UP)*				Total number of non-redundant proteins	FDR (%)
	2 UP	3 UP	4 UP	≥ 5 UP		
HMEC	211	180	121	286	798	0.49
MCF7	343	283	237	673	1536	0.53
SKBR3	266	226	172	340	1004	0.57
MDA231	306	239	171	500	1215	0.49

* Unique peptides based on 95% confidence

Amongst all samples, the non-tumorigenic cell line HMEC has the lowest number of secreted and membrane proteins identified whereas MCF7 has the highest number of proteins identified in both subcellular fractions. In each cell line, with the exception of SKBR3, more proteins were detected in the membrane relative to the secreted fraction. Figure 2.3 shows the proteome overlap for the two subcellular proteome fractions. The same proteins that were detected across the four different cell lines, herein termed "common proteins", comprised 17% (292) and 20% (423) of the total secreted and membrane proteome, respectively.

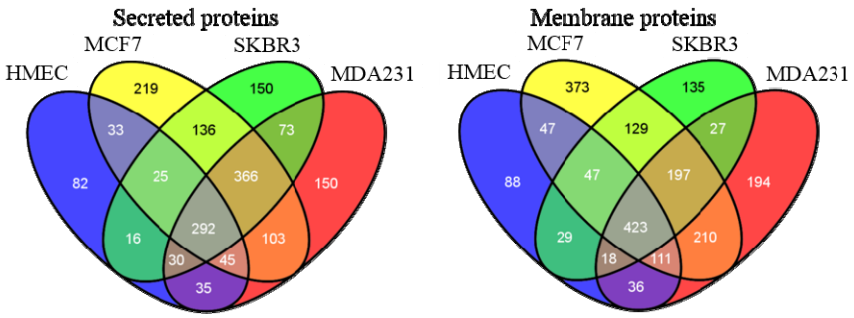


Figure 2.3 Four-way Venn diagrams showing the number of secreted and membrane proteins identified across the four breast epithelial cell lines.

In total, the secreted and membrane subcellular proteomes of the four cell lines generated 3,052 non-redundant proteins. A global comparison between the total secreted and membrane proteomes showed that they shared around 25% proteins, that is, 766 common proteins were observed in the two subcellular fractions (Figure 2.4). The common proteins ranged between 14% for HMEC to 22% for MDA231, indicating significant heterogeneity between the investigated cell lines for these common proteins. A summary of the identified proteins in the

secreted and membrane fractions of each cell line, along with their gene names, information on signal peptide, transmembrane information, exosome and breast cancer-specificity can be found in the Appendix 1 and 2.

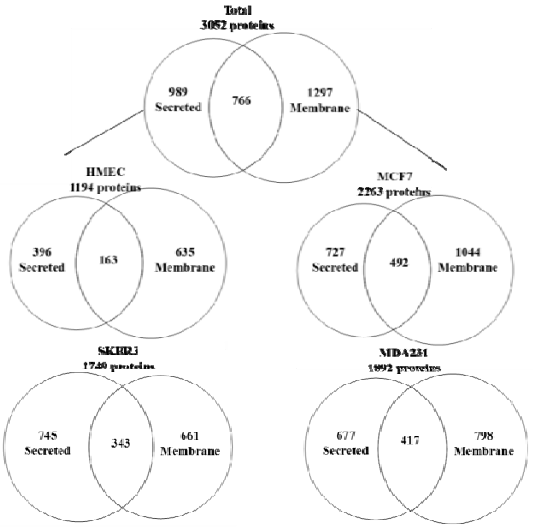


Figure 2.4 Number of common and unique proteins in the secreted and membrane subcellular fractions in all four breast epithelial cell lines (top) and separately in each of the cell lines (bottom).

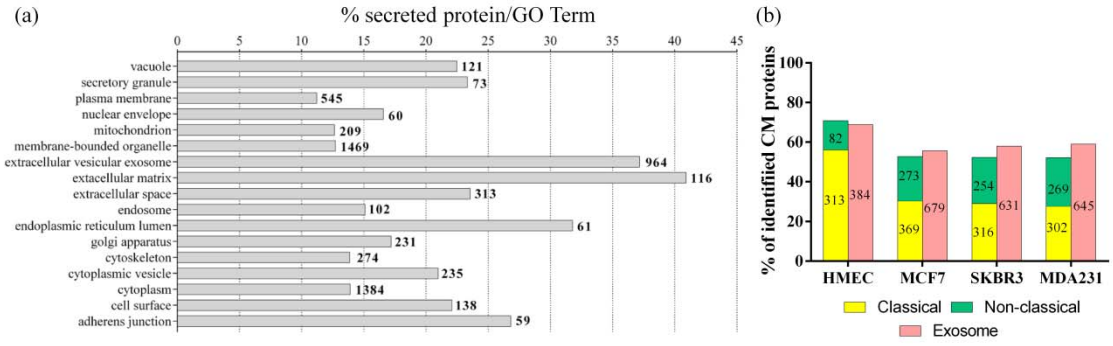


Figure 2.5 (a) Distribution of GO terms of proteins identified in the secreted fraction. (b) Distribution of secretion mechanisms used by secreted proteins i.e. secretion through classical (yellow) or non-classical pathways predicted either by SecretomeP (green) or via exosomes (light red).

Although cells were progressively adapted to grow with reduced FBS to minimize cell stress, the presence of ER stress was evaluated by assessing the expression of ER stress markers in the secreted proteome. Proteins indicative of ER stress include 78 kDa glucose-regulated protein precursor (GRP78), homocysteine-induced endoplasmic reticulum protein (HERP), endoplasmic reticulum resident protein 72 (ERP72), tryptophan--tRNA ligase (WARS), 52 kDa repressor of

the inhibitor of the protein kinase (P58IPK) and ER degradation enhancer mannosidase alpha-like 1 (EDEM) [302]. Of these, GRP78 and WARS were detected to be more abundant in the cancer cell lines (MCF7, SKBR3 and MDA231) relative to levels in the normal HMEC cell line (Table 2.2). WARS was not observed in HMEC; negligible in SKBR3 and moderately low in MCF7 and MDA231. The over-expression of GRP78 in tumor cells is well-documented and together with the lack of other ER stress indicators suggested that the elevated GRP78 levels in the three breast cancer cell lines maybe a cancer-associated feature rather than due to cell stress.

Table 2.2 Relative abundance of GRP78 and WARS expressed in the secretome of cultured breast cells.

Cell line	GRP78		WARS	
	Av. spectra count	Fold change[1]	Av. spectra count	Fold change[1]
HMEC	25.7	1.0	0	N/A
MCF7	121.3	2.7	12	High (infinite)
SKBR3	88	2.4	1.3	High (infinite)
MDA231	108.7	2.9	9.3	High (infinite)

[1] Fold change relative to HMEC, NA = not applicable

It is anticipated that the proteins identified in the secreted fraction (i.e. culture medium) are either secreted or shed from the cell surface of the cultured cells. Based on GO term, 313 and 116 proteins were classified to be located in the extracellular space and extracellular matrix, respectively, while around 79% (1,384) of the these proteins were assigned as cytoplasmic proteins (Figure 2.5a). These data suggested that between 7-18% of proteins observed in the secreted fraction were proteins actually secreted from the cells. Given that cell viability was more than 90%, it is unlikely that cytoplasmic proteins were significantly released into the extracellular space due to cell death. In the classical secretion pathway in mammalian cells, proteins that are destined for the cell surface and secretion into the extracellular environment are targeted to the secretory pathway by a signal peptides and/or transmembrane domain. Based on the prediction tool SignalP v4.1 [303] and the curated information provided by UniProt, 503 proteins in the secreted fraction were predicted to contain signal sequences and an additional 106 proteins were

further predicted to have transmembrane domains. Hence, in total 609 of the 1,755 identified proteins (~35%) in the secreted fraction could be classified as secreted proteins. There is increasing evidence that, in addition to the well characterized classical secretory pathway, proteins that lack a signal peptide can be transported to the extracellular space via various non-classical pathways, which are independent of the ER-Golgi route [304]. Of the few types known, secretions through exosomes are the most well studied [305]. Using a sequence-based non-classical protein secretion prediction tool**,** SecretomeP v2.0 [306], we found 82, 254, 269 and 273 proteins in HMEC, MCF7, SKBR3 and MDA231, respectively, predicted to reach the cell exterior via the exosome route. Based on these numbers alone, this indicated a three-fold increase of the use of the non-classical secretion pathways in breast cancer cells relative to non-tumorigenic cells. Interestingly, GO term defined 964 proteins in the secreted fraction as extracellular vesicular exosomes, which are membrane vesicles secreted by cells. Exosomes are also referred to as microvesicles, microparticles, ectosomes and by other terms in the literature. At the same time, 1,182 proteins in the secreted fraction were mapped to human-derived exosome in the Vesiclepedia database (http://microvesicles.org/) [307], which contains experimentally profiled proteins released from the exosomes. Considering the total set of identified proteins secreted via the classical pathways and, the non-classical pathways including secretions through exosomes, our analysis verified that at least 60% of proteins in the secreted fraction were of secretory nature across the four breast epithelial cell lines (Figure 2.5b).

On the other hand, GO annotation of the proteins in the membrane fraction revealed that although 570 proteins were classified as being associated with the plasma membrane, a significant proportion were associated with intracellular membrane organelles, derived from the mitochondrion, endoplasmic reticulum (ER), Golgi apparatus, endosome and lysosome (Figure 2.6). Although membrane proteins were extracted from the crude cell lysate under high-speed centrifugation, intracellular organelles were co-purified along with the plasma membrane proteins

in the microsome pellet. It is likely that integral membrane proteins were enriched during phase partitioning using Triton X-114 including those derived from the ER, Golgi apparatus, mitochondria and nucleus, in addition to the proteins from the plasma membrane [200].

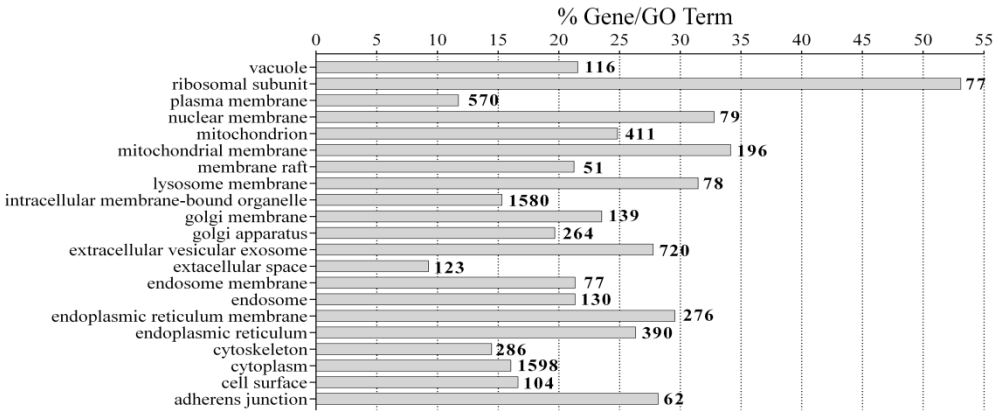


Figure 2.6 GO term classification of the subcellular locations of proteins identified in the membrane fraction. Only GO terms significantly associated with the proteins are shown ($P < 0.05$).

2.3.3 Global biological and functional analyses of secreted and membrane proteins

The proteins aberrantly secreted or shed by cultured cancer cells into the extracellular environment may mirror those of tumor cells released into the blood circulation. Over 80% of the secreted and 70% of the membrane proteins from the breast cell lines were found to map to the proteome data in the PPD (http://www.plasmaproteome.org/) [152], which contains proteins reported in plasma and serum. Cluster analysis using GO term representation for biological processes of the identified proteins revealed as expected that the membrane and secreted proteins have considerable unique biological functions assigned to them (Figure 2.7). While the membrane proteins were largely involved in different intracellular metabolic and transportation processes (i.e. macromolecule localization), secreted proteins have varied roles including biological adhesion, wound healing, cell migration, response to organic substances and antigen processing and presentation of peptide antigen. Together this indicates that the roles of secreted proteins in general are centered on the interaction with other external molecules. When secreted proteins were further interrogated using Gene-to-Systems Breast Cancer Database

(http://www.itb.cnr.it/ breast cancer) around 30% of secreted proteins from the three breast cancer cell lines matched to breast cancer-associated proteins in the database, almost twice as much as those secreted by the non-tumorigenic cells (Table 2.3). However, this difference was not observed for membrane proteins.

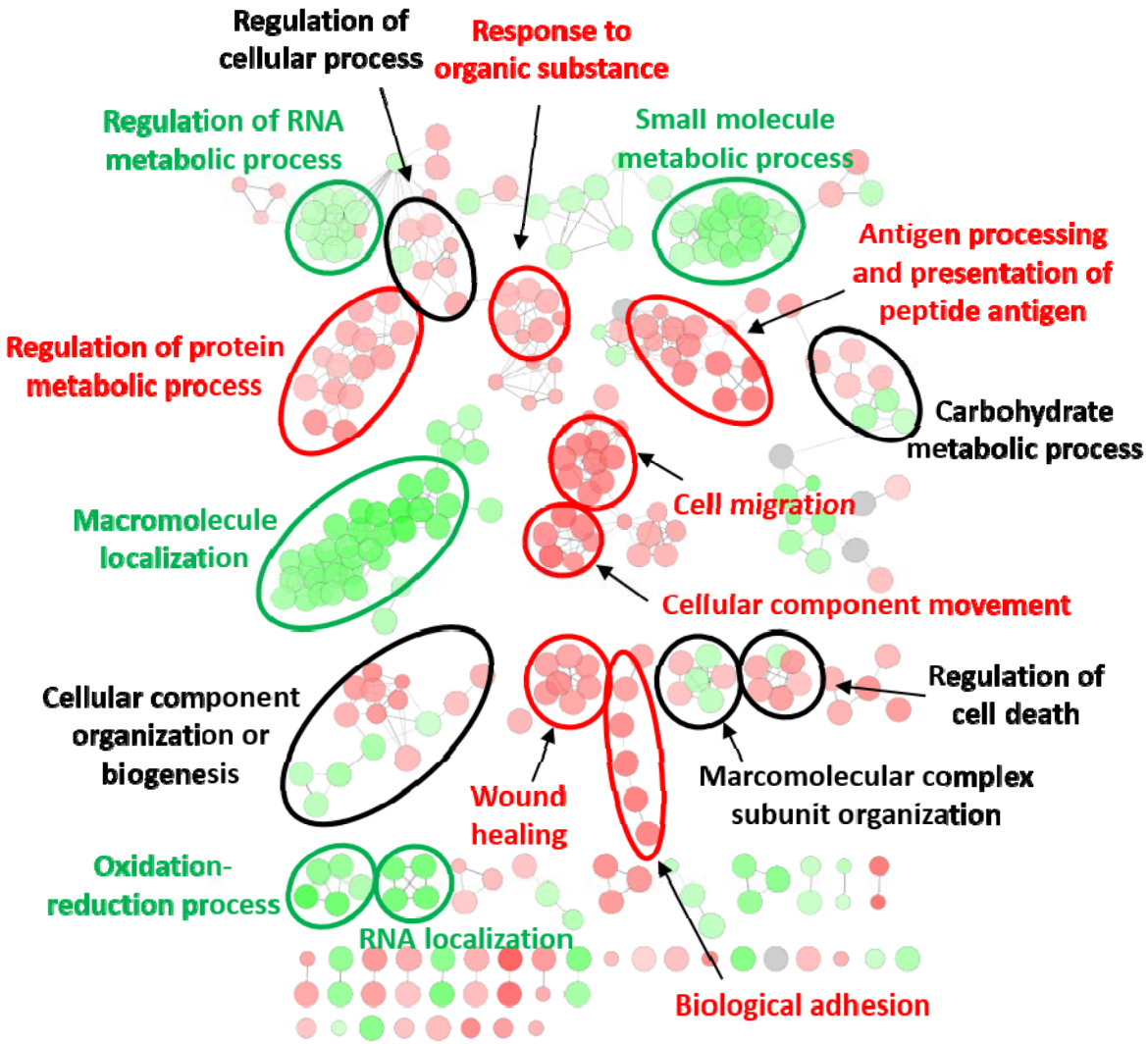


Figure 2.7 Cluster analysis of identified secreted and membrane proteins from the investigated breast cell lines and their classification according to their biological processes based on their significant association with GO terms, $P < 0.05$. Each circle represents a GO term. An over-representation of secreted proteins associated with a specific GO term is indicated in red and an over-representation of membrane proteins appears in green. When there is no over-representation of either sub-proteome, it appears in grey. Selected clusters of GO terms are highlighted by the representative GO term of the cluster.

Table 2.3 Number of secreted and membrane proteins found in the breast cancer database*

Cell lines	Secreted protein		Membrane protein	
	Number#	%#	Number#	%#
HMEC	98	17.6	198	24.8
MCF7	370	30.4	304	19.8
SKBR3	329	30.2	209	20.8
MDA231	332	30.3	273	22.5

* Gene-to-Systems Breast Cancer Database (http://www.itb.cnr.it/ breast cancer)
# Total proteins identified in the subcellular proteome

2.3.4 Comparative analysis of secreted and membrane proteins differentially expressed between normal and breast cancer cells

Label-free spectral counting of the identified proteins was used to determine the protein deregulation in the investigated cancerous breast cell lines relative to the non-cancerous cells as measured by a fold change in protein expression. As described explicitly in the introduction, spectral counting is a robust quantitative method for LC-MS/MS based proteomics data that has been shown to reliably and accurately yield a measure for the relative protein abundance between samples [231, 308-310]. The relative abundances of the proteins were expressed as normalized spectral abundance factors (NSAFs) where raw data are logarithmic transformed. Natural log transformation was applied to the raw intensity data to normalize the distribution and allow for significance testing as well as quantitation of proteins of both high and low abundance (Figure 2.8).

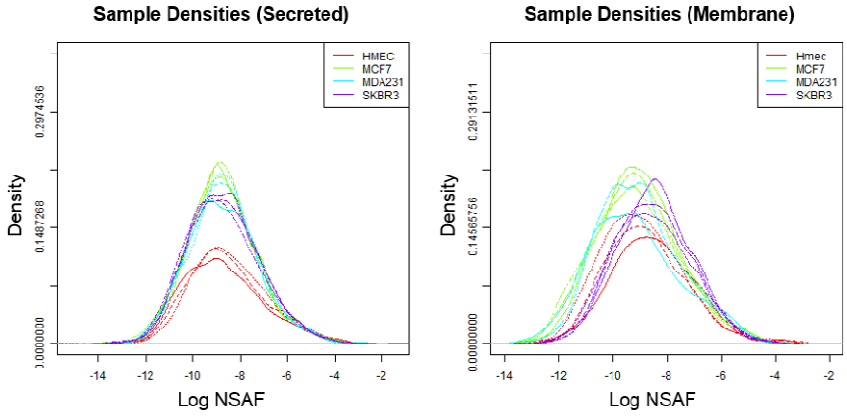


Figure 2.8 Logarithmic (natural log) transformed NSAF of secreted and membrane proteins derived from the four breast epithelial cell lines yield sample densities with normal distributions which allow for statistical comparative analyses of high and low abundant proteins between samples. See insert for colour

coding (the triplicate LC-MS/MS analyses are indicated in same colour but with full, half broken and broken lines).

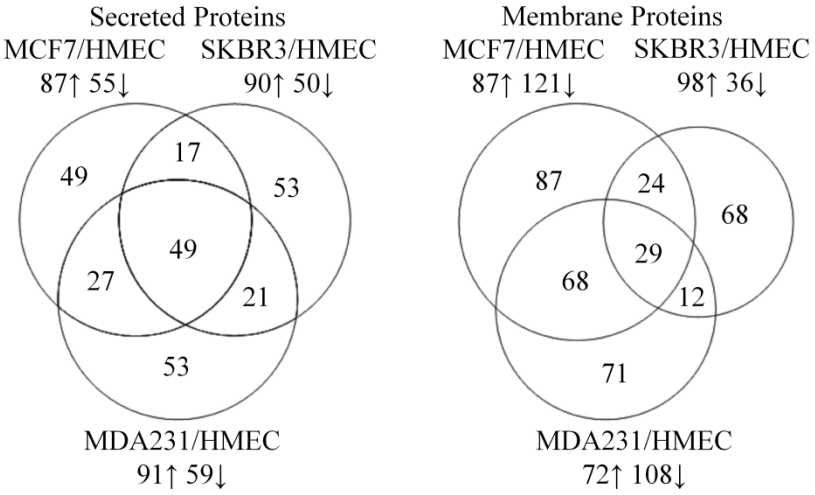


Figure 2.9 Venn diagrams showing the numbers of differentially expressed proteins between the individual breast cancer cell lines relative to the non-tumorigenic cell line (HMEC) for the subcellular proteomes (secreted, left; membrane, right).

In total, 269 secreted proteins and 360 membrane proteins displayed a threefold or greater ($P < 0.05$) regulation in the expression level for the three malignant breast cancer cell lines compared to HMEC. For the proteins identified in the secreted fraction, 142 proteins were in addition differentially expressed between MCF7/HMEC, 140 between SKBR3/HMEC; and 150 between MDA231/HMEC. For membrane proteins, 208 proteins were differentially expressed between MCF/HMEC, 134 between SKBR3/HMEC and 180 between MDA231/HMEC (Figure 2.9). A total of 49 secreted and 29 membrane proteins were found to be differentially expressed in all three breast cancer cell lines relative to the non-tumorigenic breast cell line (Table 2.4). The majority of these proteins are regulated in the same directions in all breast cancer subtypes with the exception of clusterin (CLU, secreted), heterogeneous nuclear ribonucleoproteins A2/B1 (HNRNPA2B1, membrane), pyruvate dehydrogenase E1 component subunit beta, mitochondrial (PDHB, membrane) and HLA class I histocompatibility antigen, B-41 alpha chain (HLA-B, membrane). In addition, another nine secreted and eight membrane proteins that were mutually expressed in two of the three cell lines were also differently regulated between the different cancer subtypes (Table 2.5). Notable examples include vimentin (VIM), extracellular matrix

protein 1 (ECM1), neuronal cell adhesion molecule (NRCAM) and annexin A1/A2 (ANXA1/2). The top ten exclusive proteins as measured by NSAF in both subcellular fractions of each breast cancer lines are listed in Table 2.6.

Table 2.4 List of 49 secreted and 29 membrane proteins that were differentially expressed in all three breast cancer cell lines relative to HMEC.

Secreted proteins differentially expressed in all three breast cancer cell lines			
Gene	**Protein Name**	**Average FC**	**BC-specific**
ACTR2	Actin-related protein 2	2.65	No
ALDOA	Fructose-bisphosphate aldolase A	2.59	No
BSG	Isoform 2 of Basigin	2.41	No
C3	Complement C3	-5.32	No
CAND1	Cullin-associated NEDD8-dissociated protein 1	2.89	No
CAPZB	Isoform 2 of F-actin-capping protein subunit beta	1.90	No
CCT8	T-complex protein 1 subunit theta	2.61	No
CLTC	Clathrin heavy chain 1*	3.71	Yes
CLU	Clusterin	See Table 2.5	Yes
COL12A1	Collagen alpha-1(XII) chain	-2.11	No
COPB1	Coatomer subunit beta	2.43	No
EEF1D	Elongation factor 1-delta	3.22	Yes
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	-4.14	No
ERP29	Endoplasmic reticulum resident protein 29	1.66	No
FASN	Fatty acid synthase*	4.33	Yes
FLNA	Filamin-A	2.30	Yes
FN1	Fibronectin	-4.25	No
GM2A	Ganglioside GM2 activator	-3.94	Yes
GNB1	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit	1.85	No
GPI	Glucose-6-phosphate isomerase	3.01	Yes
HIST1H4A	Histone H4	2.32	No
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	2.12	No
HRNR	Hornerin	1.97	No
HSP90AA5P	Cluster of Putative heat shock protein HSP 90-alpha A5	2.52	No
HSPA1A	Cluster of Heat shock 70 kDa protein 1A/1B	2.37	Yes
HSPG2	Basement membrane-specific heparan sulfate proteoglycan	-3.01	Yes
HYOU1	Hypoxia up-regulated protein 1	2.56	Yes
IQGAP1	Ras GTPase-activating-like protein IQGAP1	3.67	No
LMNA	Prelamin-A/C	2.75	Yes
LTBP1	Cluster of Latent-transforming growth factor beta-binding	-3.43	No
NCL	Nucleolin	2.40	Yes
NUCB1	Nucleobindin-1	-2.37	No
PA2G4	Proliferation-associated protein 2G4	2.33	No
PDCD6IP	Programmed cell death 6-interacting protein	1.93	No
PGK1	Phosphoglycerate kinase 1	3.05	Yes
PLEC	Cluster of Plectin	2.25	No
PPA1	Inorganic pyrophosphatase	2.58	No
PSMB5	Proteasome subunit beta type-5	2.03	Yes
RAB14	Ras-related protein Rab-14	2.72	No
RPS27A	Ubiquitin-40S ribosomal protein S27a	2.73	No
SDF4	45 kDa calcium-binding protein	-2.16	No
SFN	14-3-3 protein sigma	-2.37	Yes

SPTAN1	Fodrin alpha chain*	3.93	No
TFRC	Transferrin receptor protein 1	3.89	Yes
TGFBI	Transforming growth factor-beta-induced protein ig-h3	-2.61	No
TLN1	Talin-1	3.62	No
TXNDC17	Thioredoxin domain-containing protein 17	3.17	No
VASN	Vasorin	2.87	No
WDR1	WD repeat-containing protein 1	2.04	No

Membrane proteins differentially expressed in all three breast cancer cell lines

Gene	Protein Name	Avg FC	BC-specific
AHNAK	Neuroblast differentiation-associated protein AHNAK	3.71	No
ATL3	Atlastin-3	3.01	No
BCAP31	B-cell receptor-associated protein 31	3.65	No
CDIPT	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	2.21	No
CLTC	Clathrin heavy chain 1*	2.83	Yes
DSP	Desmoplakin	-2.95	No
FASN	Fatty acid synthase*	2.61	Yes
FLII	Isoform 2 of Protein flightless-1 homolog	-2.14	Yes
H2BFS	Histone H2B type F-S	2.30	No
HLA-B	HLA class I histocompatibility antigen, B-41 alpha chain	See Table 2.5	No
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	See Table 2.5	No
IMMT	Mitochondrial inner membrane protein	1.96	No
KRT14	Keratin, type I cytoskeletal 14	-2.90	Yes
KRT18	Keratin, type I cytoskeletal 18	3.42	Yes
KRT9	Keratin, type I cytoskeletal 9	-2.19	No
LMAN2	Vesicular integral-membrane protein VIP36	2.00	No
LMNB1	Lamin-B1	2.13	No
LPCAT1	Lysophosphatidylcholine acyltransferase 1	2.47	No
MYO1B	Unconventional myosin-Ib	-3.00	No
NUMA1	Nuclear mitotic apparatus protein 1	3.31	Yes
PDHB	Pyruvate dehydrogenase E1 component subunit beta,	See table 2.5	No
RAB2A	Ras-related protein Rab-2A	1.79	No
RHOA	Transforming protein RhoA	2.55	Yes
RRBP1	Ribosome-binding protein 1	2.17	Yes
SLC25A11	Mitochondrial 2-oxoglutarate/malate carrier protein	1.79	No
SPTAN1	Fodrin alpha chain*	2.62	No
SRPR	Signal recognition particle receptor subunit alpha	2.36	Yes
TOMM40	Mitochondrial import receptor subunit TOM40 homolog	2.57	No
VAT1	Synaptic vesicle membrane protein VAT-1 homolog	-2.02	No

* Changes were observed in both the secreted and membrane proteins; FC= Log fold change; BC = breast cancer

Table 2.5 Proteins that were differently regulated (>three-fold, $P < 0.05$) in the three breast cancer cell lines relative to the normal non-tumorigenic breast cell line.

Gene	Secreted Protein	Change in expression		
		MCF 7/ HMEC	SKBR3/ HMEC	MDA231/ HMEC
ANXA1	Annexin A1	↓	AB	↑
CPA4	Carboxypeptidase E	↑	NC	↓
CLU	Clusterin	↑	↑	↓
ECM1	Extracellular matrix protein 1	↓	AB	↑
LGALS3BP	Galectin-3-binding protein	↓	↑	NC

PTPRF	Receptor-type tyrosine-protein phosphatase kappa	↑	AB	↓
NRCAM	Neuronal cell adhesion molecule	↑	AB	↓
LCN2	Neutrophil gelatinase-associated lipocalin	↓	↑	AB
FAT1	Protocadherin Fat 1	AB	↑	↓
STC1	Stanniocalcin-1	↑	↓	NC

Gene	Membrane protein	MCF 7/ HMEC	SKBR3/ HMEC	MDA231/ HMEC
RPS19	40S ribosomal protein S19	↓	↑	AB
ANXA2	Annexin A2	NC	↓	↑
ATP5B	ATP synthase subunit beta, mitochondrial	↑	NC	↓
ATP5I	ATP synthase subunit e, mitochondrial	↑	AB	↓
ADAR	Double-stranded RNA-specific adenosine deaminase	↓	↑	AB
GSTK1	Glutathione S-transferase kappa 1	NC	↑	↓
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	↑	↑	↓
HLA-B	HLA class I histocompatibility antigen, B-41 alpha chain	↓	↓	↑
PDHB	Pyruvate dehydrogenase E1 component subunit beta,	↓	↑	↓
TACSTD2	Tumor-associated calcium signal transducer 2	NC	↑	↓
VIM	Vimentin	AB	↓	↑

NC = no change; AB = absent; ↑ up-regulated; ↓ down-regulated

Table 2.6 Top 10 secreted and membrane proteins exclusively present in each subtype.

Gene	Protein Name	Cell line (Fraction)
SERPINA3	Alpha-1-antichymotrypsin	MCF7 (secreted)
GFRA1	GDNF family receptor alpha-1	MCF7 (secreted)
SERPINA5	Plasma serine protease inhibitor	MCF7 (secreted)
TFF1	Trefoil factor 1	MCF7 (secreted)
SDK1	Protein sidekick-1	MCF7 (secreted)
CLSTN2	Calsyntenin-2	MCF7 (secreted)
PCDH7	Protocadherin-7	MCF7 (secreted)
NCAM2	Neural cell adhesion molecule 2	MCF7 (secreted)
NPNT	Nephronectin	MCF7 (secreted)
BMP7	Bone morphogenetic protein 7	MCF7 (secreted)
MAOB	Amine oxidase [flavin-containing] B	MCF7 (membrane)
SLC7A2	Low affinity cationic amino acid transporter 2	MCF7 (membrane)
ABCB6	ATP-binding cassette sub-family B member 6, mitochondrial	MCF7 (membrane)
HEATR6	HEAT repeat-containing protein 6	MCF7 (membrane)
HSD17B4	Peroxisomal multifunctional enzyme type 2	MCF7 (membrane)
MAOA	Amine oxidase [flavin-containing] A	MCF7 (membrane)
RAB17	Ras-related protein Rab-17	MCF7 (membrane)
RFT1	Protein RFT1 homolog	MCF7 (membrane)
CERS2	Ceramide synthase 2	MCF7 (membrane)
HERC2	E3 ubiquitin-protein ligase HERC2	MCF7 (membrane)
ERBB2	Receptor tyrosine-protein kinase erbB-2*	SKBR3 (secreted)
SUSD2	Sushi domain-containing protein 2	SKBR3 (secreted)
FBLN2	Fibulin-2	SKBR3 (secreted)
IVL	Involucrin	SKBR3 (secreted)
EGFR	Epidermal growth factor receptor	SKBR3 (secreted)
RNF213	E3 ubiquitin-protein ligase RNF213	SKBR3 (secreted)
MUC16	Mucin-16	SKBR3 (secreted)
NAPRT1	Nicotinate phosphoribosyltransferase	SKBR3 (secreted)
CNP	2',3'-cyclic-nucleotide 3'-phosphodiesterase	SKBR3 (secreted)
CP	Ceruloplasmin	SKBR3 (secreted)

DHRS2	Dehydrogenase/reductase SDR family member 2,	SKBR3 (membrane)
KRT4	Keratin, type II cytoskeletal 4	SKBR3 (membrane)
ERBB2	Receptor tyrosine-protein kinase erbB-2*	SKBR3 (membrane)
ANXA7	Annexin A7	SKBR3 (membrane)
AHSG	Alpha-2-HS-glycoprotein	SKBR3 (membrane)
SLC35F6	Solute carrier family 35 member F6	SKBR3 (membrane)
AGRN	Isoform 6 of Agrin	SKBR3 (membrane)
EFHD1	EF-hand domain-containing protein D1	SKBR3 (membrane)
HMGN1	Non-histone chromosomal protein HMG-14	SKBR3 (membrane)
SRP14	Signal recognition particle 14 kDa protein	SKBR3 (membrane)
FLNC	Filamin-C*	MDA231 (secreted)
PTX3	Pentraxin-related protein PTX3	MDA231 (secreted)
TGFB2	Transforming growth factor beta-2	MDA231 (secreted)
CSPG4	Chondroitin sulfate proteoglycan 4	MDA231 (secreted)
LRP1	Prolow-density lipoprotein receptor-related protein 1	MDA231 (secreted)
SRGN	Serglycin	MDA231 (secreted)
MYOF	Myoferlin	MDA231 (secreted)
CFH	Complement factor H	MDA231 (secreted)
CSF1	Macrophage colony-stimulating factor 1	MDA231 (secreted)
EDIL3	EGF-like repeat and discoidin I-like domain-containing	MDA231 (secreted)
FLNC	Filamin-C*	MDA231
GNAO1	Guanine nucleotide-binding protein G(o) subunit alpha	MDA231
LBR	Lamin-B receptor	MDA231
MICAL2	Protein-methionine sulfoxide oxidase MICAL2	MDA231
SSR1	Translocon-associated protein subunit alpha	MDA231
NRP1	Neuropilin-1	MDA231
NES	Nestin	MDA231
LDHB	L-lactate dehydrogenase B chain	MDA231
RAB32	Ras-related protein Rab-32	MDA231
FMNL3	Formin-like protein 3	MDA231 (membrane)

* Changes were observed in both the secreted and membrane fractions of the same cell line.

For the functional proteome analysis, proteins that were exclusively present in all three cancer cell lines were combined with the proteins that were significantly up-regulated in the breast cancer cells relative to HMEC. This created a list of 589 breast cancer-related proteins. Similarly, proteins that were only expressed in HMEC were combined with proteins which were significantly down-regulated in the breast cancer cells relative to the HMEC, generating a list of 173 non-breast cancer-related proteins. Accessing the two groups of breast cancer- and non-breast cancer-related proteins for an enrichment of GO biological processes above the "background" distribution revealed that these abundant proteins have biological functions grouped into five major clusters (Figure 2.10). The four major biological processes associated with abundant breast cancer-related proteins were nucleobase-containing small metabolic

process, regulation of protein metabolic process, negative regulation of cell death, cell junction assembly and cellular component disassembly. On the other hand, proteins associated with hemidesmosome assembly and extracellular matrix (ECM) organization were predominantly under-represented amongst the breast cancer-related proteins. Interestingly, ECM organization is found within the cluster represented by cellular component disassembly comprising of proteins over-expressed in breast cancer.

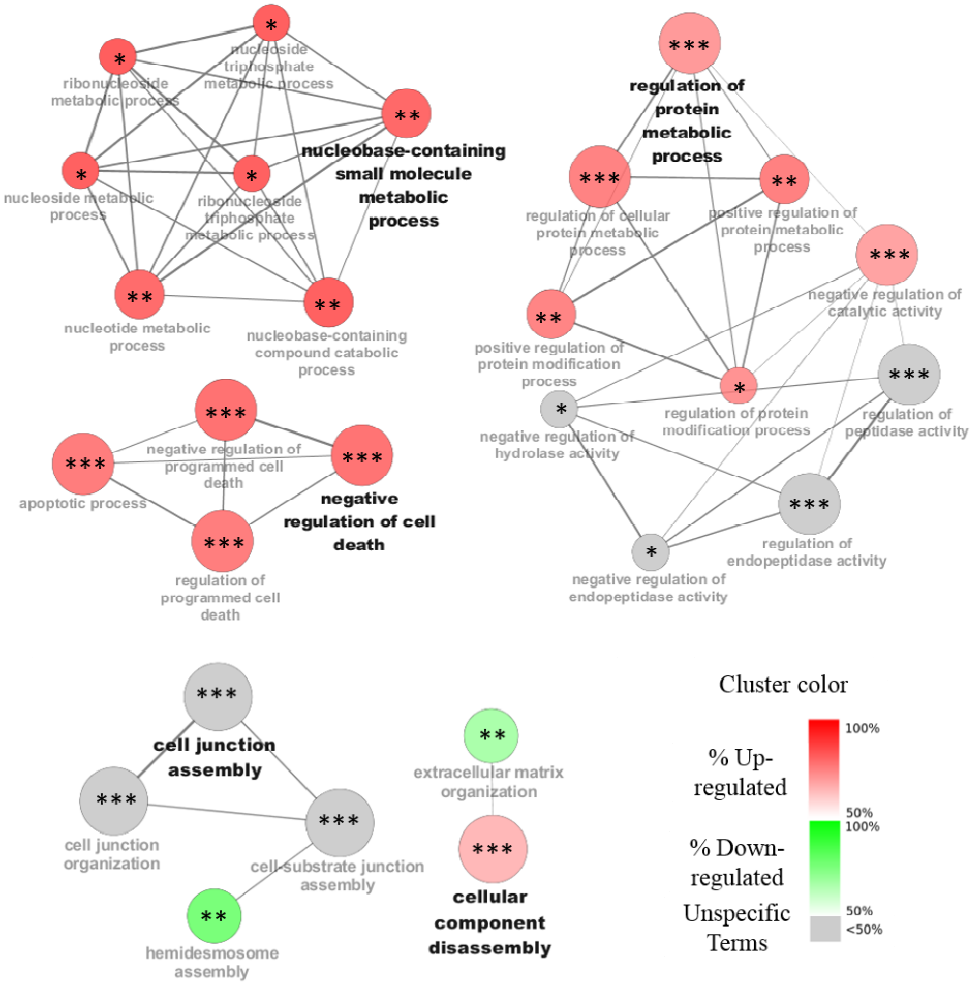


Figure 2.10 Over- (red) and under- (green) represented proteins in the investigated breast cancer cell lines relative to proteins derived from the normal non-tumorigenic breast cell line were mapped to five major clusters of biological processes. Each cluster is a network of closely related biological processes in which the one with the most number of mapped proteins is highlighted in bold. Biological processes significantly different from the normal cell line are indicated as * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0005$.

The web-based tool known as STRING (http://string-db.org/) [311] was used to further analyze protein-protein interactions that occurred in a few of the altered biological processes. In the largest cluster of regulated biological function ("The regulation of protein metabolic process"), protein interactions were observed to center around three groups of proteins (Figure 2.11). One group consisted of a dense intricate web of protein subunits of the proteasomes and proteasome activators suggesting a deregulation in the protein degradation in the breast cancer cells. Another group comprised a closely related network of translational initiator factors including eukaryotic translation initiation factor 4 gamma 1 (EIF4G1) and eukaryotic translation initiation factor 3 subunit K (EIF3K). The third group showed interactions between several breast cancer-associated proteins such as the vascular endothelial growth factor A (VEGFA), mitogen-activated protein kinase 1 (MAPK1), ras-related C3 botulinum toxin substrate 1 (RAC), 14-3-3 protein gamma (YWHAG), exportin-1 (XPO1) and GTPase NRas (NRAS).

Similarly, protein-protein interaction networks for the proteins which were found to be down-regulated in breast cancer in the clusters of "hemidesmosome assembly" and "extracellular matrix (ECM) organization", including the co-cluster of up-regulated proteins in "cellular component disassembly" (see Figure 2.10), were visualized using STRING (Figure 2.12). Protein interaction networks consisting of the highly expressed ribosomal proteins, mainly derived from the membrane fraction, and the poorly expressed secreted laminins and collagen proteins, were evident. The investigated over- and under-represented proteins mostly congregated within the same network. A few over-expressed proteins were found closely associated with the under-expressed proteins, including laminin subunit alpha-5 (LAMA5), collagen alpha-1(V) chain (COL5A1), alpha-2-macroglobulin-like protein 1 (A2M) and disintegrin and metalloproteinase domain-containing protein 9 (ADAM9).

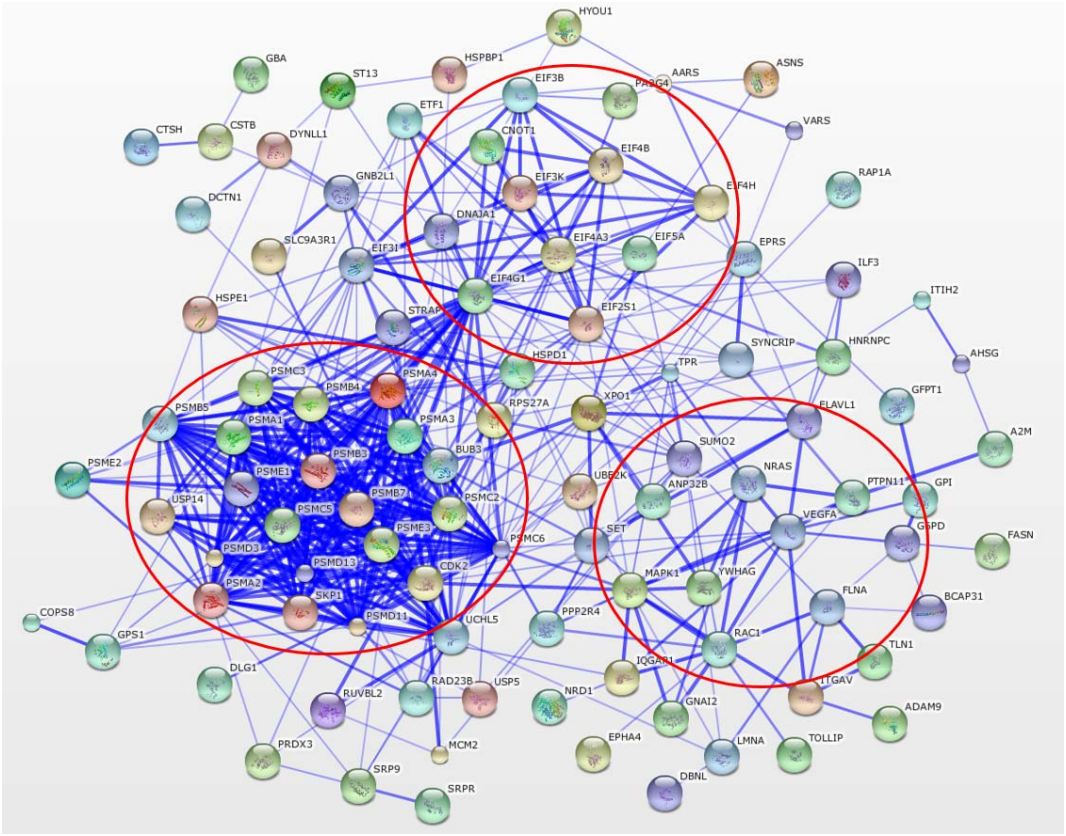


Figure 2.11 Protein-protein interaction map of proteins in the major cluster of biological processes represented by "The regulation of protein metabolic process" (see Figure 2.10 for more) shows three groups of closely associated protein groups (red circles).

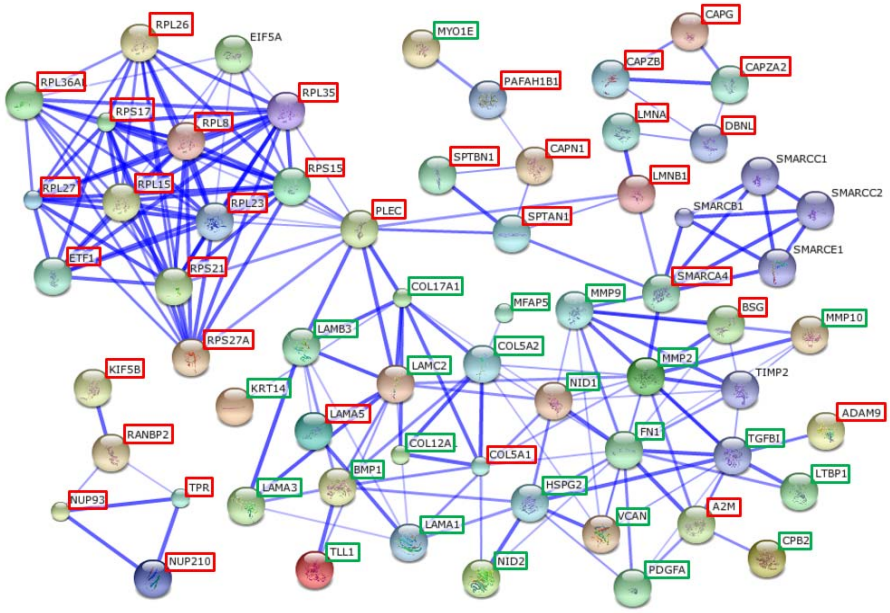


Figure 2.12 Protein-protein interaction map of proteins under-represented in breast cancer cell lines relative to normal breast cells (green box) found in the clusters of "Hemidesmosome assembly" and "Extracellular matrix organization" and over-represented proteins (red box) in the associated cluster "Cellular component disassembly". Proteins not within color boxes were either not differentially expressed or not present in the datasets.

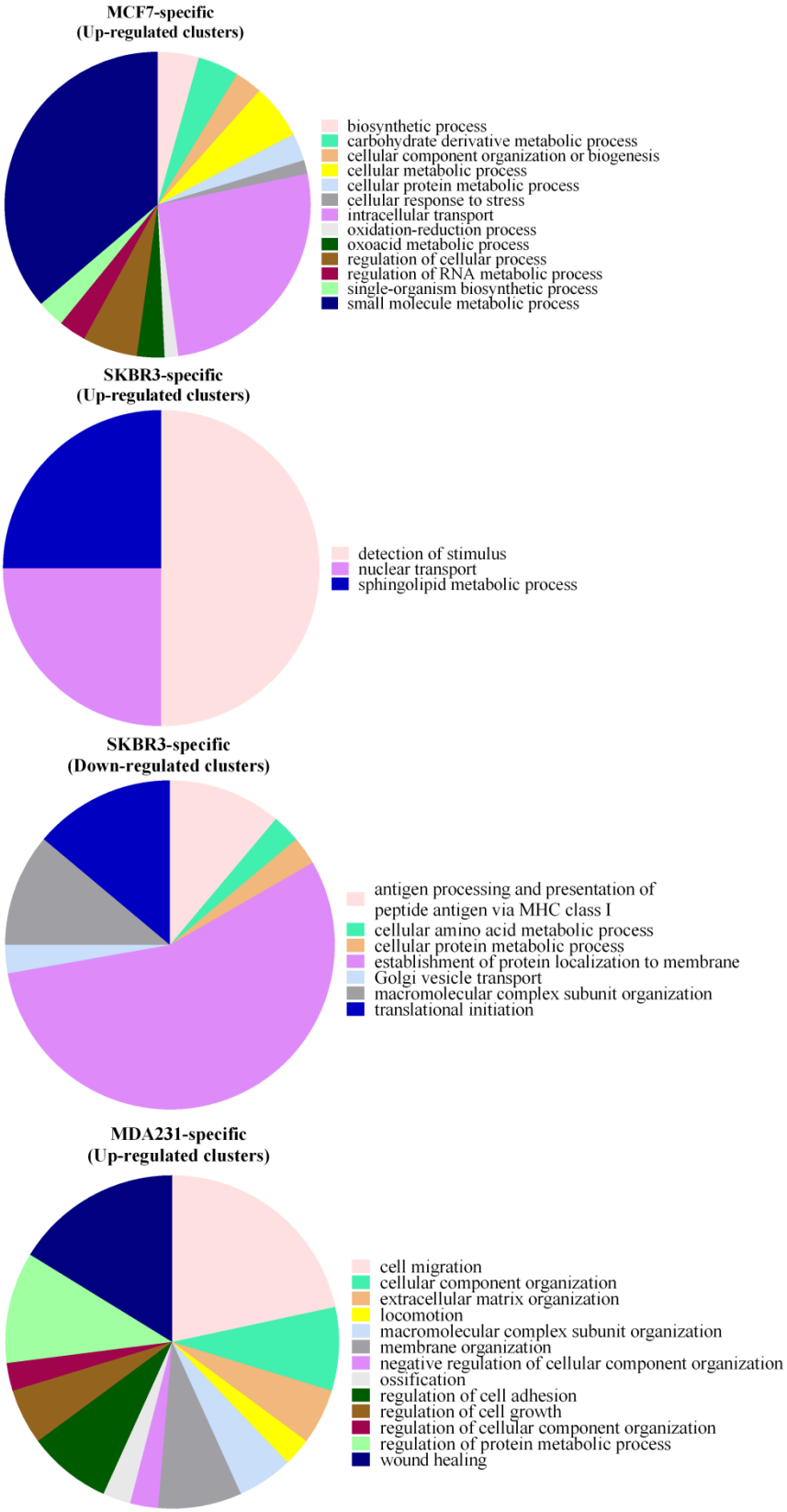


Figure 2.13 The major clusters of biological processes associated with the identified up- and down-regulated proteins in the individual breast cancer subtypes. Only data subsets yielding specific clusters of biological processes are included here. No clusters were obtained for MCF7 and MDA231 down-regulated proteins

Proteins found at the border between the networks of proteins over- and under-represented in breast cancer were of particular interest such as plectin (PLEC), matrix metalloproteinases-2/9/10 (MMP2, MMP9, MMP10), transcription activator BRG1 (SMRCA4) and basigin (BSG). In particular, plectin was observed to be a key mediator between the network of over-represented ribosomal proteins and under-presented extracellular matrix components including the laminins and collagens.

2.3.5 Differential expression of subtype-specific proteins in breast cancer

The three breast cancer cell lines studied are representative of three breast cancer subtypes, namely, MCF7 for the luminal subtype, SKBR3 for the HER2-enriched subtype and MDA231 for the triple negative subtype. Comparative proteome analysis between the "normal" reference cell line (HMEC) and each of the breast cancer cell lines indicated subtype-specific changes. Next, the subtype-specific changes at the functional level were determined by mapping the regulated protein to biological processes and pathways for better understanding of the underlying molecular mechanisms associated with each subtype. The subtype-specific analysis included proteins that were uniquely expressed and differentiated in each subtype. To improve the specificity of the analysis, redundant proteins between the cell lines were removed. In total, 1,074, 882 and 838 proteins for MCF7, SKBR3 and MDA231, respectively, were used to assess potential enrichment of GO biological process terms and perform pathway analysis. The biological processes GO terms significantly enriched ($P < 0.05$) for each subtype are shown in Figure 2.13. No enrichments were observed for under-represented proteins in MCF7 and MDA231. Over-represented proteins in each subtype were found to be involved in a wide spectrum of biological functions including various cellular metabolic processes for MCF7; nuclear transport and sphingolipid metabolic process for SKBR3; and cell migration, locomotion, regulation of cell adhesion and wound healing for MDA231. The key biological processes observed in MDA231 supports the highly invasive and metastatic nature of this breast cancer cell

line. The metabolic processes observed in MCF7 and SKBR3 may be restricted to less metastatic cells or early stages of tumorigenesis. Similarly, pathway analysis revealed increased perturbations in the extracellular matrix organization and L1CAM interactions in MDA231, both of which are known to promote cell migration and invasion (Figure 2.14). An interesting observation was that all three breast cancer subtypes shared a common up-regulated pathway, associated with G protein-coupled receptor (GPCR) signaling. Given that redundant proteins between the three subtypes were removed to increase the specificity of the analysis, the GPCR signaling in each breast cancer subtype apparently involved three different clusters of proteins, which may affect cellular transformation via different intracellular signaling mechanisms. Analysis of protein-protein interaction networks, using STRING revealed small subsets of proteins that were key integrators of the entire protein network (Figure 2.15). In MCF7, the majority of the proteins including calmodulin (CALM1), guanine nucleotide exchange factor (VAV2) and ras GTPase-activating-like protein (IQGAP1) were shown to cluster around cell division control protein 42 homolog (CDC42), a GTPase protein involved in regulating diverse signalling pathways that control cell morphology, cell migration and cell growth. In SKBR3, epidermal growth factor receptor (EGFR) is observed as the core protein in the network, associating with rho-related GTP-binding protein (RHOB) and growth factor receptor-bound protein 2 (GRB2), both of which have important regulatory roles in signal transduction. Two large and one smaller cluster of protein interaction networks were observed for MDA231. The two larger networks were focused around cyclin-dependent kinase 1 (CDK1) and lysophosphatidic acid receptor 1 (LPAR1) while the smaller cluster contained two proteins, namely, rho-associated protein kinase 2 (ROCK2) and rho-related GTP-binding protein (RHOG). Most of the proteins found in the GPCR signalling pathway of MDA231 were associated with LPAR1, including guanine nucleotide-binding protein G(o) subunit alpha (GNAO1), heme-binding protein 1 (HEBP1), annexin A1 (ANXA1) and metastasis-suppressor KiSS-1 (KISS1).

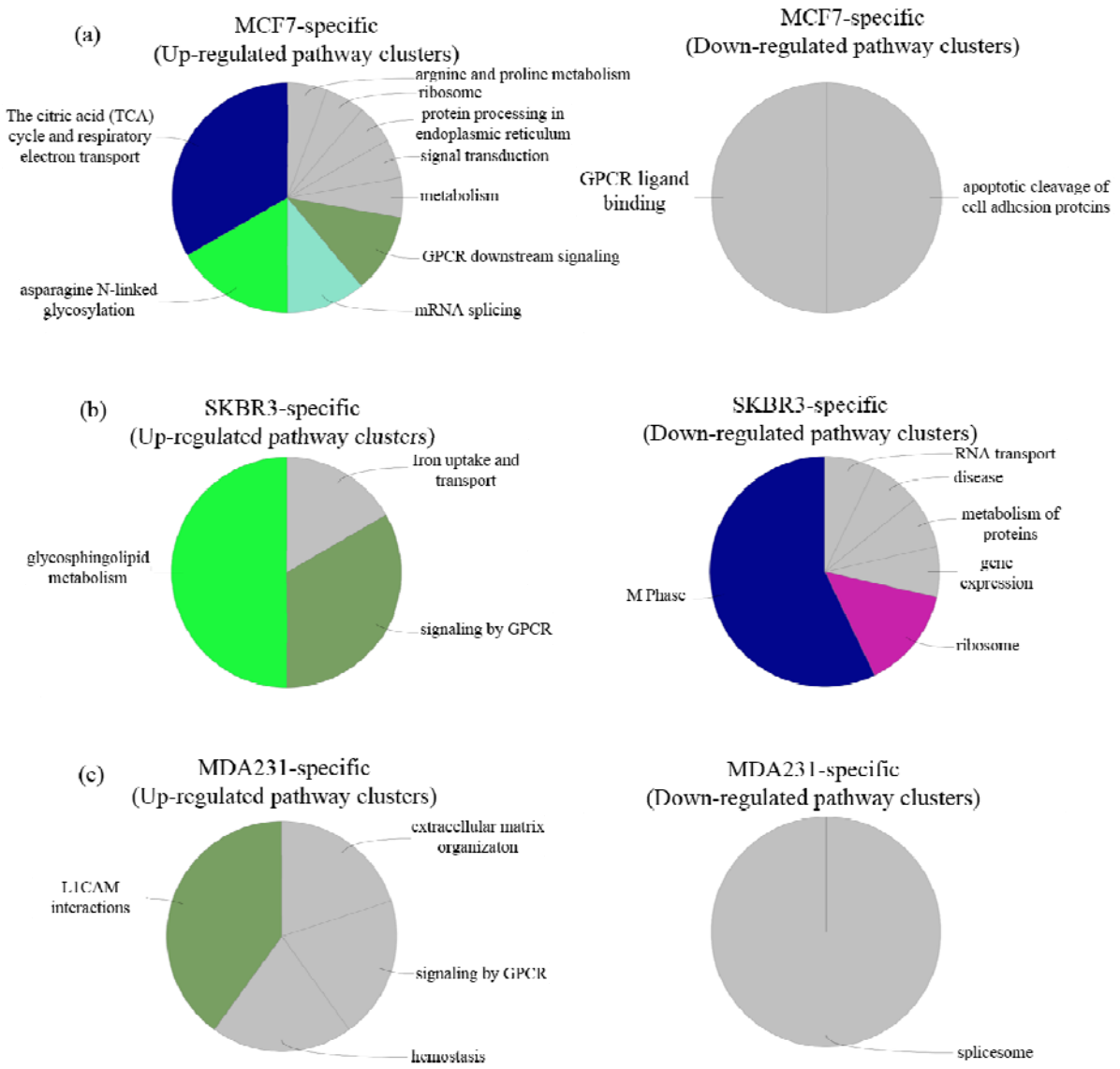


Figure 2.14 Pathway analyses of up- (left) and down- (right) regulated proteins associated with the individual breast cancer subtypes.

Further analysis was performed to interrogate the protein-protein interaction in the L1CAM interactions and extracellular matrix organization pathways observed in MDA231. The resulting protein-protein interactions featured an intricate network that centered on closely related proteins, including cell adhesion proteins (integrins and L1CAM), ECM structural proteins (laminins), basement membrane proteins (collagens) and the cell surface proteoglycan protein

syndecan-4 (SDC4) (Figure 2.16). Two proteins, VEGFA and MAPK1, with altered protein expression, involved in the regulation of cellular metabolic process, were also observed in this network.

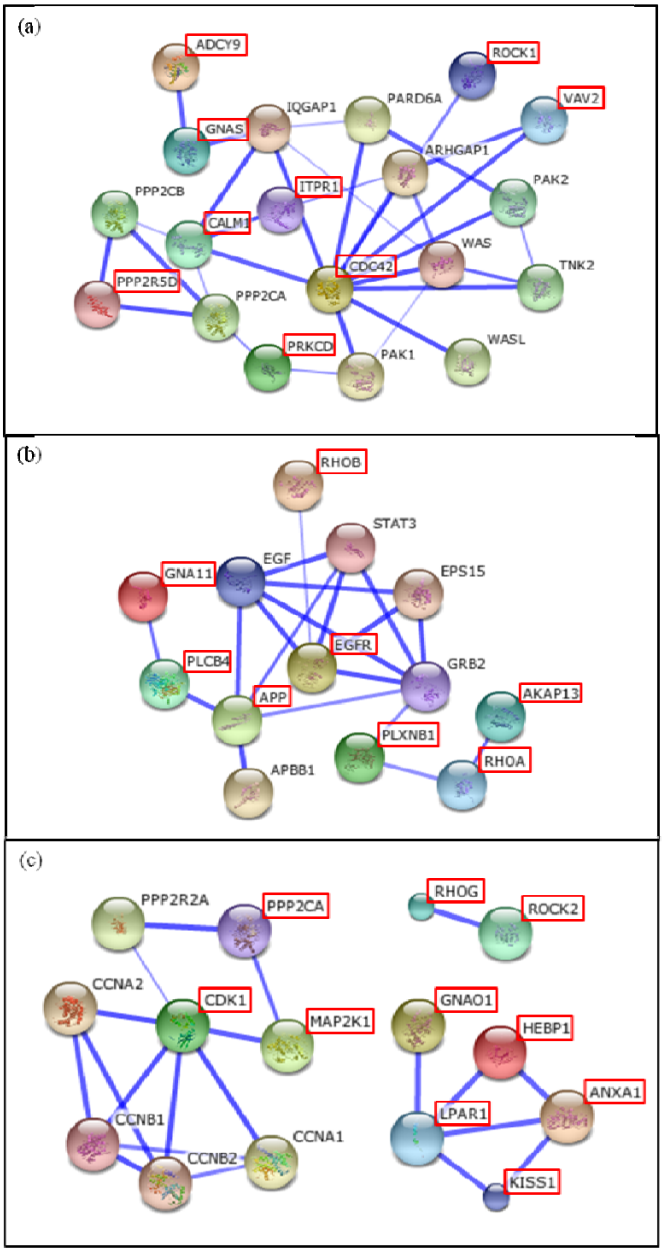


Figure 2.15 Protein-protein interaction network analysis using STRING of proteins observed in the GPCR signalling pathway specific to (a) MCF7, (b) SKBR3 and (c) MDA231. Proteins marked in red boxes are present in the specific datasets.

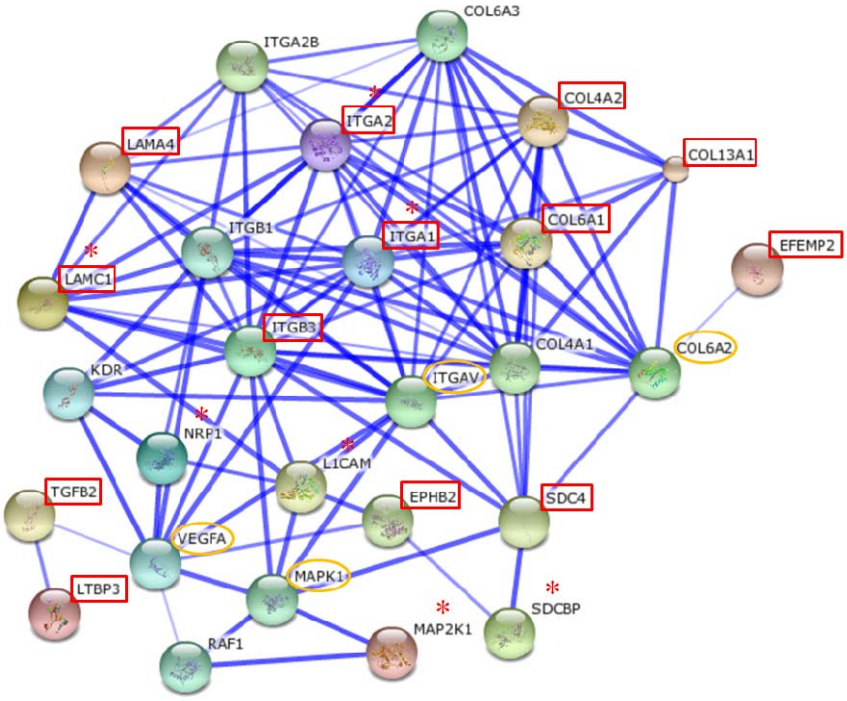


Figure 2.16 Protein-protein interaction map of over-represented proteins in MDA231 in the categories of L1CAM interaction (marked in *) and ECM organization (marked in red boxes). Proteins in yellow ovals were present in all breast cancer cell lines, but absent in HMEC.

2.3.6 Proteomics-based clustering of tumorigenic and breast cancer subtypes

Hierarchical cluster analysis and principal component analysis (PCA) were performed to evaluate the relationship of the secreted and membrane protein expression profiles with the known differences in genotype and phenotype of the four investigated breast epithelial cell lines. To achieve this, hierarchical clustering with Pearson correlation were applied to the log-transformed NSAF values of the identified proteins that were differentially regulated between the breast tumorigenic cell lines and HMEC. Two major clusters were observed in the dendrogram, with HMEC evidently well separated from the other three breast cancer cell lines albeit a better segregation (greater distance) between the two clusters were achieved for the secreted protein profiles (Figure 2.17a-b). Similar trends were observed in PCA analysis, which showed a clear division between cancer and non-cancer samples and a segregation of secreted proteins between the three breast cancer subtypes (Figure 2.17c-d).

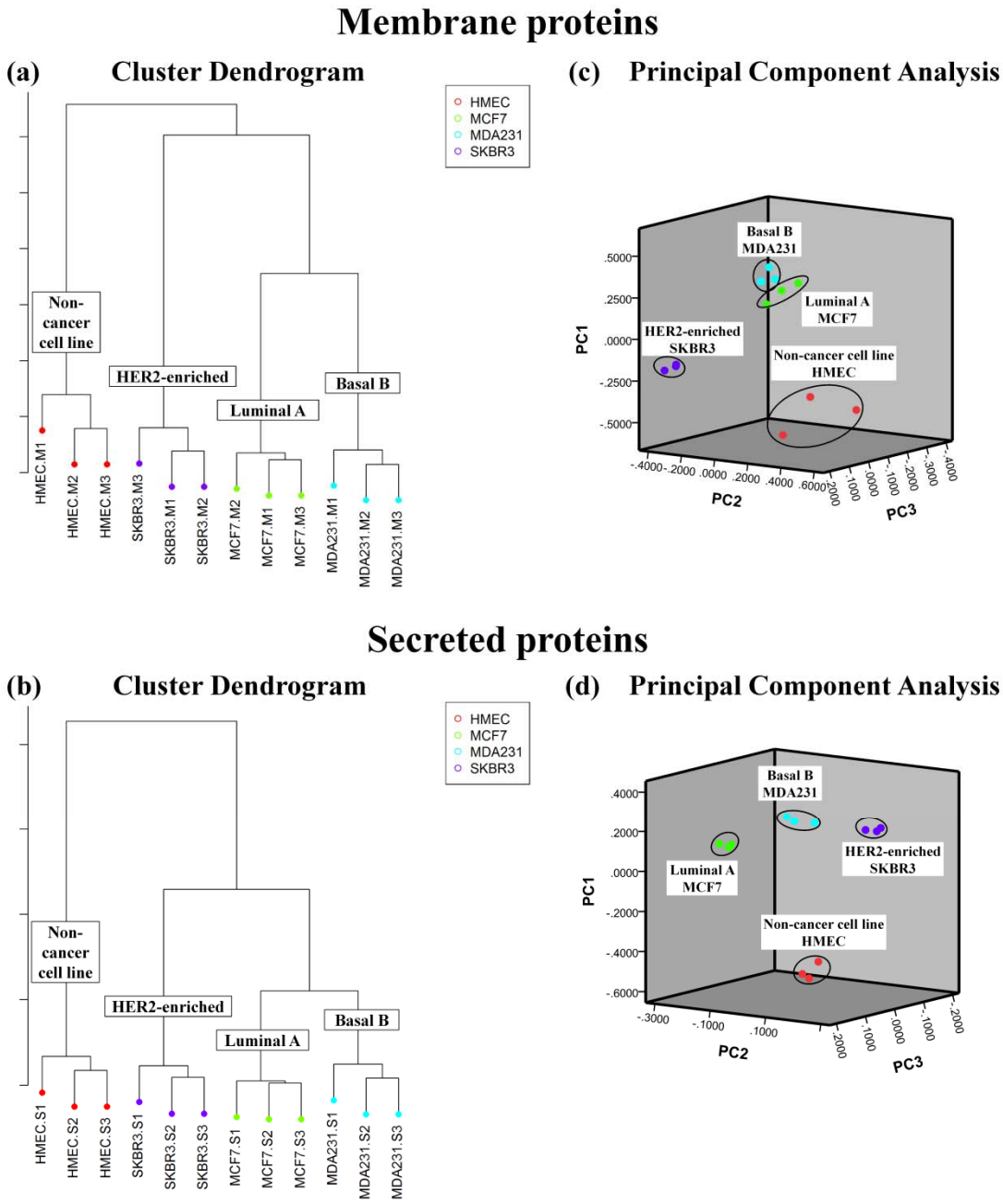


Figure 2.17 Dendrogram cluster analysis using hierarchical clustering and 3D plot of PCA of membrane protein (a, c) and secreted protein (b, d) profiles of the four epithelial breast cells investigated. PC1, principal component 1; PC2, principal component 2; PC3, principal component.

2.4 Discussion

Although considerable time and efforts have been invested in the molecular research of breast cancer, there remains a lack of definitive early-stage biomarkers for the onset and recurrence of the disease and limited drug targets for aggressive forms of breast cancer are available. Early

detection in breast cancer is crucial as it can enhance the disease prognosis and increase the survival rate of affected patients due to the availability of more effective treatment options. Breast cancer is a highly heterogeneous disease defined by multiple cellular and molecular subtypes characterized by varied clinical outcomes. It is hope that a better understanding of the underlying molecular mechanisms in breast cancer pathogenesis eventually will facilitate the development of efficient therapeutics and prognostic/diagnostic markers. In this study, a label-free quantitative LC-MS/MS based shotgun proteomics approach was applied to investigate the proteome changes in the secretory and membrane subcellular fractions of several tumorigenic (three subtypes) and non-tumorigenic breast cell lines and performed in-depth functional comparisons of the protein expression levels to identify deregulated proteins and protein networks and pathways within the individual breast cancer subtypes.

A key concern associated with investigating the secretome (secreted fraction) of cell lines is the potential contribution of intracellular proteins as a result of cell death induced by the serum-free conditions. Intuitively, longer duration of serum starvation leads to increased cellular apoptosis. Serum deprivation beyond the 48 hour incubation has been shown to result in a dramatic increase in cell lysis resulting in the release of the abundant intracellular proteins into the condition media [312]. In the same study, higher secretome contents were obtained following 48 hour incubation compared to 24 hour incubation with virtually unchanged cellular viabilities and protein profiles in excellent agreement with our observations presented here. Very often, cells stress can lead to ER stress, disturbing protein folding and leading to the activation of unfolded protein response (UPR), which could potentially modulate cellular characteristics [302]. The lack of many ER stress marker proteins suggest that there is minimal cellular stress, hence the 48 hour incubation in serum-free conditions accurately captures the "true" secretome of the cultured breast cells.

On the other hand, it became clear that the membrane fraction did not accurately reflect the cell surface proteins but enriched for most membrane bound proteins from multiple intracellular organelles or microsomes. In support of this observation, the majority of the identified membrane proteins were biologically associated with cellular metabolic processes and macromolecule localization or transportation. Additionally, this may result in the secreted protein profiles generating a clearer division between the breast cancer subtypes compared to that achieved by the membrane protein profiles.

The phase separation of integral membrane proteins using non-ionic detergent Triton X-114 was first investigated by Bordier [313] and has been shown in several studies to be effective in enriching this class of proteins [314]. However, beside cell surface proteins, the integral membrane proteome also included those derived from all other membrane e.g. ER, Golgi, nucleus and mitochondria. Moreover, many non-integral membrane proteins were identified. The sonication step during extraction may have momentarily disrupted the membranes which are quickly re-assembled thereby possibly trapping some non-integral membrane proteins within the lipid bilayer. The use of Triton X-114 is therefore not an efficient approach to isolate cell surface proteins for proteomics and glycomics studies. Cell surface specific extraction i.e. selective biotinylation of the cell surface proteins was also successfully employed to enrich and purify the cell surface proteome as has been published before [292]. This method was adopted in Chapter 4 for the isolation of cell surface proteins of selected breast cancer cell lines. Further work is needed to isolate the cell surface proteomes from the remaining cell lines for cell surface specific proteomics and glycomics analyses.

Analysis of the secretome suggests that the non-classical pathways are the major mechanisms by which the secreted proteins in breast cancer cells reach the extracellular space. Several MS-based proteomics studies of the secretome have reported the presence of many intracellular proteins,

such as cytoskeletal, ribosomal, nuclear and chaperone proteins, in the conditioned media of cultured cancer cells, in addition to the common extracellular secreted proteins [53, 101, 312, 315]. Amongst the various non-classical secretory pathways, protein secretion via exosomes has been intensely researched in recent years. Exosomal protein secretion has been associated with numerous cancers including melanoma, ovarian cancer, colorectal cancer, liver cancer and breast cancer [141, 316-319]. It is increasingly evident that proteins residing in exosomes may modulate cell-cell communication, thereby promoting cancer invasion and metastasis via various signaling mechanisms [305]. In this study, no specific isolation of exosomes was performed, yet a significant proportion of the secreted proteins were associated with exosomes. As one of the main aims of this study was to characterize the secretome in a global manner, no further analysis was performed for the exosomes. However, these findings suggest that a separate investigation of exosome secretion is warranted in breast cancer cell lines.

Both the secreted and membrane fractions are rich sources of potential protein biomarkers and drug targets. In particular, breast cancer cells were shown to secrete almost twice as many proteins known to be implicated in breast cancer, compared to non-cancer cells. To reduce false positive identification, each protein identified must be present in all three replicates of each sample and with a minimum of two unique peptides and a total minimum of five spectral counts in all replicates. The combined proteomics datasets from the secreted and membrane fractions and the multiple breast cancer subtypes identified over 3,000 non-redundant proteins, which included the protein homologs of the same family that were grouped together. Many were known to be involved in key biological roles including regulation of cell growth, cell-cell communication, cell adhesion and immune responses. Using the label-free quantification approach with strict identification criteria and a strong fold change with p-value less than 0.05, these proteins were found to be significantly regulated and by different mechanisms in the investigated breast cancer subtypes. A few themes central to breast cancer biology emerged from the comparative protein

profiling analysis; perturbations in the cellular metabolic processes, cytoskeletal organization and extracellular matrix were general alterations observed in breast cancer cell lines. Specifically, cellular hyper-activities were associated with the proteasomes, translational initiation factors and a number of proteins with diverse functions as signaling molecules (MAPK1, YWHAG, RAC1), growth factors (VEGFA), chaperones (ANP21B) and transporters (XPO1). These networks of regulated proteins were observed to be intricately connected, indicating that a combination of events including protein degradation, DNA repair, cell death and regulation of gene expression were orchestrated in breast tumorigenesis.

In the absence of subsequent verification using clinical samples, these proteins were validated *in silico* using The Human Protein Atlas (http://www.proteinatlas.org), which is a publicly available database portal where an antibody-based approach has been extensively used to explore the human proteome. Many proteasomes such as PSMC3, PSMA3, PSMB7, PSMB3, PSMD13, PSME3, USP14; initiation factors such as EIF3B and EIF4G1; XPO1 and YWHAG showed moderately to strong immunochemical staining in malignant breast tissues. Many of these proteins have well established roles in breast cancer [320, 321] and have been actively targeted in therapeutic studies [322-324]. Interestingly, inhibition of the catalytic activity of proteasomes was shown to increase anti-apoptotic processes by down-regulating the MAPK signaling pathway, which is crucial in pathological cell proliferation [325]. Over-expression of XPO1, a nuclear export protein, has been demonstrated to drive the development of breast cancer and its inhibition was shown to be a promising anti-tumor strategy to suppress the progression of invasive breast cancer [326]. However, these potential biomarker proteins have not been sufficiently validated e.g. using targeted proteome strategies such as selective reactive monitoring (SRM)/multiple reactive monitoring (MRM), which is needed for clinical utility.

The ECM and cytoskeletal organization were deregulated in the investigated breast cancer subtypes supporting the notion that alterations in these processes promote oncogenic transformation in breast cells [327-330]. The cytoskeleton and ECM proteins dynamically interact with one another to maintain the structural integrity of cells. In the tumor microenvironment, remodeling of the cytoskeleton architecture and the aberrant expression of specific ECM components underlie a process known as epithelial-mesenchymal-transition (EMT), where cells lose their epithelial polarity to acquire the migratory mesenchymal cell phenotype [182]. EMT is mediated by the re-organization of cytoskeleton components, an increase in integrin-based adhesion and a loss of expression of hemidesmosome proteins or degradation of underlying basement membrane (BM); all events will contribute to the enhanced migration and invasion of tumor cells [178, 331]. A higher expression of several cytoskeletal proteins was observed in breast cancer including various keratins, desmoplakin (DSP), filamins, spectrins and the cytoskeletal-associated PLEC. Increased integrin expression was mainly restricted to MDA231. Simultaneously, expression of hemodesmosomes as well as basement membrane proteins such as laminins and collagens, which are involved in cell-matrix adhesion were reduced. Interestingly, the intermediate filament VIM, which is an established EMT marker, was up-regulated in MDA231, down-regulated in SKBR3 and absent in MCF7. Amongst the three investigated breast cancer cell lines, MDA231 is considered to be the most invasive and metastatic cell line [332]. In effect, the majority of these proteins such as the cytoskeletal proteins, basement membrane proteins and intermediate filament proteins are part of a complex protein network described as the "integrin adhesome", i.e. large adhesion complexes at the cell interface that allow cells to detect and respond to multiple extracellular signals and consequently affecting the cell adhesion, migration and cytoskeletal organization [333]. Of this subset of proteins, PLEC and VIM may have great promise as cancer biomarkers and drug targets since their ablation indicated a modulation of the cancer cell invasion and metastasis potential by disrupting the formation of filamentous network in the ECM [334].

Another aim of this study was to identify breast cancer subtype-specific proteome changes. The analysis revealed that the GPCR signaling pathway, which occurred as a major signal transduction pathway, was mediated by different subsets of proteins in the individual breast cancer subtypes. The GPCRs constitute the largest and most diverse group of integral membrane receptors that bind to an array of external ligands including chemokines, hormones and neurotransmitters. Upon ligand binding, signals are transduced via the G proteins which are closely associated to the GPCRs and a cascade of events are then triggered leading ultimately to a specific cellular response such as gene expression [335]. Aberrant expression of components related to the GPCR signaling pathway can therefore have adverse effects on the cell growth and proliferation leading directly or indirectly to tumorigenesis. At present, 60% of current cancer drugs target the GPCRs.[336] Although our datasets did not identify any major GPCRs, several G proteins and downstream effectors, such as GTPases were well represented in the identified proteomes. Similarly, such G protein-related gene products were previously observed in the enriched plasma membrane fractions derived from several breast cancer cell lines including MCF7, SKBR3 and MDA231. [45] In this study, subtype-specific G proteins were observed including GNAS for MCF7, GNA11 for SKBR3 and GNAO1 for MDA231. These proteins have previously been linked to several other cancer types e.g. GNAS in pancreatic cancer [337], GNAO in gastric cancer [338] and GNA11 in melanoma [339]. However, their roles in breast cancer have not yet been reported. Recently, siRNA screening identified that amplification of GNAS gene locus may contribute positively to the pathogenesis of ER-positive breast cancer [340]. The analysis here showed that signaling of GPCR in MCF7 may be mediated through the activation of cdc42, a member of the Rho family of GTPases. Activation of EGFR in SKBR3 and LPAR1 in MDA231 suggested that cross-talks exists between the GPCR signaling pathway and the pathways of EGFR and LPA, respectively.

2.5 Conclusion

The high degree of interconnectivity between networks of altered proteins in breast cancer as a general pathology and within the individual breast cancer subtypes indicates that reliable breast cancer biomarkers and therapeutic targets may be discovered from improving our understanding of their functional roles and their interaction within the tumor environment. Utilizing an *in vitro* model system such as cultured breast epithelial cell lines allowed the mapping of breast cancer- and subtype-specific proteome alterations without the molecular and cellular complexity observed in tissues albeit with the potential caveat that the cell cultures may not reflect the natural *in vivo* system. Bioinformatics-assisted pathway analysis of the function and connectivity of the large proteome maps provided molecular insights into the underlying pathological mechanisms of the highly complex breast cancer biology. Crucially these semi-automated approaches were built on high quality proteome data, followed by label-free quantitative LC-MS/MS based proteomics to identify differential protein expression in the secreted and membrane protein fractions derived from the four investigated breast epithelial cell lines. Protein profile features associated with individual breast cancer subtypes were discerned. Breast cancer- and subtype-specific proteins may serve as potential cancer biomarkers and therapeutic drug targets due to their involvement in the aberrant biological processes or pathways central to breast cancer progression, invasion and metastasis.

CHAPTER 3

STRUCTURAL ANALYSIS OF *N*-GLYCOME CHANGES IN BREAST CANCER

Over half of the mammalian proteome is estimated to be glycosylated. Many important biological processes are mediated through the glycans attached to the glycoproteins. The membrane proteome and the secreted media of cultured cancer cells are a rich reservoir of glycoproteins. Analysis of N-glycans released from these proteins offers a unique opportunity to study subcellular-specific N-glycosylation changes in cancer. This chapter is made up of two parts. Part 1 is presented as a publication as the first study to investigate N-glycan changes on secreted glycoproteins from a panel of breast cancer cell lines. To our knowledge, N-glycan profiling and characterization in the secretome of breast cancer cell lines have not been systematically investigated. Part 2 focuses on N-glycome analysis of membrane proteins extracted from the same panel of cultured breast epithelial cells.

Part 1

Publication II - Comprehensive *N*-glycome profiling of cultured human epithelial breast cells identifies unique secretome *N*-glycosylation signatures enabling tumorigenic sub-type classification

Journal of proteome research

Article

pubs.acs.org/jpr

Comprehensive *N*-Glycome Profiling of Cultured Human Epithelial Breast Cells Identifies Unique Secretome *N*-Glycosylation Signatures Enabling Tumorigenic Subtype Classification

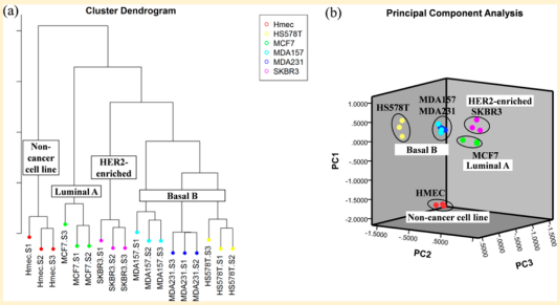
Ling Y. Lee,[†] Morten Thaysen-Andersen,[†] Mark S. Baker,[‡] Nicolle H. Packer,[†] William S. Hancock,[‡,§] and Susan Fanayan*[,†]

[†]Department of Chemistry and Biomolecular Sciences and [‡]Australian School of Advanced Medicine, Macquarie University, Sydney, NSW 2109, Australia

[§]Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States

Supporting Information

ABSTRACT: The secreted cellular sub-proteome (secretome) is a rich source of biologically active glycoproteins. *N*-Glycan profiling of secretomes of cultured cancer cells provides an opportunity to investigate the link between protein *N*-glycosylation and tumorigenesis. Utilizing carbon-LC−ESI-CID-MS/MS of protein released native *N*-glycans, we accurately profiled the secretome *N*-glycosylation of six human epithelial breast cells including normal mammary epithelial cells (HMEC) and breast cancer cells belonging to luminal A subtype (MCF7), HER2-overexpressing subtype (SKBR3), and basal B subtype (MDA-MB157, MDA-MB231, HS578T). On the basis of intact molecular mass, LC retention time, and MS/MS fragmentation, a total of 74 *N*-glycans were confidently identified and quantified. The secretomes comprised significant levels of highly sialylated and fucosylated complex type *N*-glycans, which were elevated in all cancer cells relative to HMEC (57.7−87.2% vs 24.9%, $p < 0.0001$ and 57.1−78.0% vs 38.4%, $p < 0.0001$−0.001, respectively). Similarly, other glycan features were found to be altered in breast cancer secretomes including paucimannose and complex type *N*-glycans containing bisecting β1,4-GlcNAc and LacdiNAc determinants. Subtype-specific glycosylation were observed, including the preferential expression of α2,3-sialylation in the basal B breast cancer cells. Pathway analysis indicated that the regulated *N*-glycans were biosynthetically related. Tight clustering of the breast cancer subtypes based on *N*-glycome signatures supported the involvement of *N*-glycosylation in cancer. In conclusion, we are the first to report on the secretome *N*-glycosylation of a panel of breast epithelial cell lines representing different subtypes. Complementing proteome and lipid profiling, *N*-glycome mapping yields important pieces of structural information to help understand the biomolecular deregulation in breast cancer development and progression, knowledge that may facilitate the discovery of candidate cancer markers and potential drug targets.



KEYWORDS: *subtype classification, luminal breast cancer, triple-negative breast cancer, N-linked glycans, secretome, sialylation, fucosylation, bisecting GlcNAc, LacdiNAc, Lewis antigen*

INTRODUCTION

Breast cancer is the most frequently diagnosed cancer in women worldwide. According to World Health Organization (WHO), the disease accounted for around 14% of all female cancer-related mortalities in 2008, and this figure is estimated to double by 2030.[1] The five-year survival rate for localized breast cancer is almost 99% but falls drastically to 24% following tumor metastasis.[2] This implies that breast cancer is highly curable if diagnosed early, which may be facilitated by identification of specific and sensitive biomarkers for early and accurate detection.

A significant challenge with the identification of suitable biomarkers for early detection lies in the heterogeneous nature of breast cancer pathogenesis. Breast cancer diagnostics are heavily based on histological examination and molecular testing of tumor tissues for staging, grading, and subtyping of the disease. In particular, knowledge of molecular marker status such as estrogen receptor (ER), progesterone receptor (PR), and epidermal growth factor receptor 2 (HER2) has contributed to successful targeted therapy.[3] Gene expression

Special Issue: Proteomics of Human Diseases: Pathogenesis, Diagnosis, Prognosis, and Treatment

Received: April 8, 2014

profiling has revealed the capability of these receptors in segregating the disease broadly into five major subtypes, namely, luminal A (ER positive and/or PR positive, HER2 negative), luminal B (ER positive and/or PR positive, HER2 positive), HER2-enriched (HER2 positive), basal-like (ER negative, PR negative, and HER2 negative), and normal-like.[4] The basal-like subtype is also known as triple-negative breast cancer due to the absence of the three markers. In breast cancer cell lines, gene expression profiling further identified two distinct subgroups in the basal-like subtype: basal A and B.[5] Each subtype is strongly associated with different prognoses of the disease, with better survival outcomes observed in luminal A and HER2-enriched subtypes and significantly poorer prognosis in basal-like subtype. Moreover breast cancer patients with luminal A and HER2-enriched subtypes respond positively to targeted treatment using hormone therapy and monoclonal antibody (Trastuzumab), respectively. In contrast, patients with basal-like tumors lack targetable treatment and have limited therapy options, which include surgery and chemotherapy.[6] Current clinically approved serum biomarkers, such as CA15-3 and carcinoembryonic antigen, lack specificity and sensitivity and are not suitable for screening and early detection of the disease.[7] To improve prognosis outcome for these patients, a better understanding of the underlying molecular mechanisms involved in the highly aggressive and metastatic nature of basal-like subtype is clearly needed in order to identify suitable biomarkers for early diagnosis and as potential therapeutic targets.

Proteomic analyses have identified many secreted and membrane proteins that are involved in tumorigenesis. More than half of these are glycosylated, carrying either *N*- or *O*-linked glycans. The *N*-glycans are known to facilitate essential biological functions of glycoproteins such as cell growth, proliferation, and differentiation; cell−cell or cell−matrix interactions; and immune responses.[8] It is now evident that altered *N*-glycans play key roles in disrupting these functions and contribute to the development and progression of different cancers including those of the colon,[9] pancreas,[10] breast,[11] ovary,[12] prostate,[13] and liver.[14] Sensitive and accurate profiling of protein glycans is now possible due to gradual advances in LC−MS/MS technologies.[15] Comprehensive structural elucidation and quantitative analysis of *N*-glycans have been performed on various types of breast cancer samples including breast tumor tissues, serum of breast cancer patients, and membrane proteins of breast cancer cell lines.[11,16−18] These analyses revealed common aberrant features of *N*-glycosylation such as the relative increase of fucosylation, sialylaton, β1,6-GlcNAc branching, high mannose, and Lewis type determinants, which correlate with poor disease prognosis. Such transformations may be accompanied by concomitant changes in expression levels of the processing glycosidases and glycosyltransferases.[19] Therefore, understanding the molecular changes at the glycome level may provide clues on irregularities of protein glycosylation that drive the invasive and metastatic behaviors of breast tumor cells. Since many biomarkers are glycoproteins, this in turn can aid in identifying suitable early diagnostic and prognostic biomarkers and effective drug targets.

To our knowledge, few studies have performed detailed profiling and characterization of *N*-glycosylation on proteins secreted from breast cancer cell lines. Secretions from cancer cell lines represent an excellent source of glycoproteins. Unlike serum, which is highly complex and carries secreted glycoproteins from various cellular tissues such as stroma or liver, the homogeneity of cancer cell secretions preclude contaminations from other cell types and hence allow for detection of cancer-specfic *N*-glycan changes.

The aim of this study is to map and compare the secretome *N*-glycomes of a panel of breast cancer cell lines. On the basis of existing literature, we hypothesize that unique *N*-glycosylation signatures exist in non-cancer and breast cancer cell lines as well as within the subtypes of breast cancer. Such molecular signatures may serve as potential tumor markers and advance our understanding of the involvement of protein *N*-glycosylation in cancer- and subtype-specific malignant transformation. In this study, we used a non-tumorigenic breast epithelial cell line derived from primary human mammary epithelia cells (HMEC) as a normal reference and five breast cancer cell lines representing different breast cancer subtypes, namely, luminal A (MCF7), HER2-enriched (SKBR3), and basal B (MDA-MB-157, MDA-MB-231, HS578T). Using porous graphitized carbon (PGC)-negative ion-LC−CID-MS/MS of reduced but otherwise native *N*-glycans released from their proteins, we profiled the *N*-glycomes of secreted proteins of these cell lines and identified key glycosylation pathways deregulated in breast cancer. Our results revealed significant tumorigenic and breast cancer subtype-specific *N*-glycosylation signatures, including alterations in glycoprotein sialylation, fucosylation, GlcNAc branching, terminal Lewis determinants, bisecting GlcNAc, and *N*,*N*′-diacetyllactosamine (LacdiNAc).

MATERIALS AND METHODS

Breast Cell Origin and Collection of Secretomes

Human mammary epithelial cells (HMEC) were purchased from Lonza (CC-2551, Walkersville, MD). Human breast cancer cell lines MCF7, SKBR3, MDA-MB-157 (MDA157), MDA-MB-231 (MDA231), and HS578T were obtained from American Type Culture Collection (Manassas, VA). HMEC was grown in HuMEC Ready Media (Invitrogen, CA). The five breast cancer cell lines were grown in RPMI (Sigma, MO) supplemented with 5% FBS (Invitrogen, CA), 10 mM glutamine (Invitrogen, CA), and 10 μg/mL insulin. Cells were maintained at 37 °C in 5% CO_2 for all experiments. Each cell line was grown in triplicate to around 80% confluency, washed at least four times with ice-cold PBS to remove traces of FBS, and incubated in serum-free media at 37 °C in 5% CO_2 for 48 h. Conditioned media containing the serum-free secreted proteins were collected, followed by centrifugation at 2,000 x *g*. Supernatant was collected and concentrated, followed by buffer exchange with PBS (1x) using Amicon Ultra centrifugal filter devices with a 10,000 molecular weight cutoff membrane (Millipore, MA). The concentrations of secreted proteins were measured using Bradford reagent (Sigma, MO) to determine the total amount of proteins secreted by each cell line.

Cell Proliferation Assay

Cells were seeded at 1.3 × 10^4 cells/mL/well in six-well plates and incubated overnight at 37 °C in 5% CO_2. Cells were counted every 24 h over a 4-day period using a cell counter (Biorad, CA). The doubling time for each cell line was determined from their exponential growth phase.

Release of *N*-Glycans from Secreted Proteins for LC−MS/MS Analysis

N-Glycans were released from approximately 20 μg of secreted proteins as previously described.[20] Briefly, proteins were precipitated with acetone overnight at −20 °C. Following

solubilization in 8 M urea, proteins were immobilized on methanol-activated PVDF membrane (Millipore, MA) and allowed to dry overnight. Membrane-bound proteins were incubated with 2.5 U of PNGase F (*Flavobacterium meningospeticum*) for 16 h at 37 °C to ensure complete release of *N*-glycans. Released *N*-glycans were incubated with 100 mM ammonium acetate (pH 5) for 1 h at RT and subsequently dried by vacuum centrifugation. Reduction of *N*-glycans was performed with 20 μL of 1 M sodium borohydride (Sigma, MO) in 50 mM potassium hydroxide (Sigma, MO) for 3 h at 50 °C. Reduced samples were quenched with 2 μL of glacial acetic acid and desalted as described below.

Desalting of Reduced Native *N*-Glycans

Strong cation exchange columns were packed on top of ZipTip C18 columns (Millipore, MA), using 30 μL of AG 50W X8 cation exchange resin (Biorad, CA). Columns were washed three times sequentially with 50 μL of each of the following: 1 M HCl, methanol, and water. *N*-Glycan mixtures were added to the prepared columns, and the flow-through fractions were retained. Columns were washed twice with 50 μL of water, and the flow-through fractions were pooled with the initial fractions and dried by vacuum centrifugation. Residual borate was removed by adding 100 μL of methanol, and samples were allowed to evaporate in the vacuum centrifuge. This step was repeated 4−5 times until the white borate residue disappeared. The desalted samples were kept at −80 °C if not desalted immediately with carbon.

Carbon resin obtained from carbon SPE cartridges (Grace, IL) was suspended in 50% methanol. Small carbon columns were prepared by adding 5 μL of carbon slurry onto an empty TopTip (Glygen, MD). Carbon columns were washed sequentially with 30 μL of 90% acetonitrile containing 0.1% (v/v) TFA, 40% acetonitrile containing 0.1% (v/v) TFA, and water. Samples were dissolved in 15 μL of water, applied to the columns, and washed twice with water. All flow-through fractions were discarded. Desalted glycans were eluted with 20 μL of 40% acetonitrile containing 0.1% (v/v) TFA and dried by vacuum centrifugation. Samples were stored at −80 °C if not analyzed immediately.

Analysis of *N*-Glycans by Mass Spectrometry

N-Glycan alditols were separated using porous graphitized carbon (PGC) LC columns (5 μm Hypercarb KAPPA, 100 mm × 0.2 mm, ThermoFisher, MA) using a Dionex HPLC system (Ultimate 3000) connected directly to an ESI-MS/MS Bruker HCT Ultra ion trap mass spectrometer. Separation was performed using a binary gradient solvent system made up of solvent A (10 mM NH_4HCO_3) and solvent B (90% ACN/10 mM NH_4HCO_3). The flow rate was 2 μL/min, and a total gradient of 100 min was programmed as follows: 0−2.5% solvent B for 0−13 min; 2.5−17.5% solvent B for 14−48 min; 17.5−50% solvent B for 48−65 min; 50−100% solvent B for 65−75 min; 100% solvent B for 75−80 min; back to 0% solvent B for 80−85 min and 100% solvent A equilibration for 15 min. Settings for the MS/MS were as follow: drying gas flow, 6 L/min; drying gas temperature, 300 °C; nebulizer gas, 12 psi; skimmer, −40.0 V; trap drive, −99.1 V; and capillary exit, −166 V. Smart fragmentation was used with start and end amplitude of 30% and 200%, respectively. Ions were detected in ion charge control set at 100,000 ions/scan and with maximum accumulation time of 200 ms. MS spectra were obtained in negative ion mode with two scan events: a full scan (m/z 400−2,200) at scan speed of 8,100 m/z per second and data-dependent MS/MS scans after CID fragmentation of the top two most intense precursor ions with threshold 30,000 and relative threshold of 5% relative to the base peak. Dynamic exclusion was inactivated to ensure MS/MS generation of closely eluting *N*-glycan isomers. Precursors were observed mainly in charge states −1 and/or −2 and rarely in charge state −3. Mass accuracy calibration of the instrument was performed using tuning mix (Agilent, CA) prior to acquisition, and *N*-glycans released from bovine fetuin served as positive controls for the sample preparation and the LC−MS/MS performance before each data acquisition. Differences between observed and theoretical precursor and fragment masses were generally less than 0.2 Da. Three LC−MS/MS technical replicates were performed for each cell line.

Assessing Transcriptome Differences of Selected Glycosyltranferases of Breast Cancer Cells

The ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) was queried for data sets with transcriptomes of human breast cancer cell lines. The data set (E-GEOD-48213) selected for further analysis contained transcriptional profiling of 56 cultured breast cell lines prepared from the TruSeq RNA Illumina platform and analyzed on an Agilent Bioanalyzer High Sensitivity chip.[21] The panel included three breast cancer cell lines (MCF7, SKBR3, and MDA231) that were used for subtype comparisons. Processed data were downloaded, and transcriptomes associated with glycosylation enzymes (glyco-transcriptomes) were selected for further analysis.

Data and Statistical Analysis

The resulting raw data were viewed and analyzed using ESI-Compass v1.4 (Bruker Daltonics). Monoisotopic masses were manually obtained and searched against Glycomod (http://web.expasy.org/glycomod/) to obtain possible glycan monosaccharide compositions. These were subsequently verified manually by *de novo* sequencing of their corresponding MS/MS spectra and their matches to recently published *N*-glycan data sets.[9,22] The relative abundance of each glycan in a sample was determined using the ratio of the extracted ion chromatogram (EIC) peak area of each *N*-glycan over the sum of EIC peak areas of all *N*-glycans in the sample. Glycans were quantified in all of their observed charged states. Three technical replicates were performed for each cell line.

All relative abundances of *N*-glycans were presented as a percentage out of 100%, as mean ± SD. Statistical analyses were conducted using GraphPad Prism (v6) and SPSS for Windows (v21.0). One-way ANOVA, followed by post-hoc analysis (Dunnett or Tukey tests), which was used for comparison between each of the different cancer cell lines to the reference HMEC cell line and for comparison between the three selected cancer cell lines (MCF7, SKBR3, MDA231). All p values were adjusted taking into account the multiple comparisons made and reported as multiplicity adjusted p values. Values that were less than 0.05 were regarded as statistically significant. Due to large orders of magnitude, glycan profiling data were log transformed to remove skewness before performing hierarchical clustering analysis using an in-house program written in R.

Table 1. Characteristics of Investigated Cultured Human Breast Epithelial Cells

cell line	HMEC	MCF7	SKBR3	MDA157	MDA231	HS578T
gene cluster[a]	NA	luminal	HER2-enriched	basal B	basal B	basal B
origin[b] and tumorigenicity[b,c]	human mammary epithelial cells, primary tissue, non-tumorigenic	adenocarcinoma; pleural effusion; tumorigenic; differentiated	adenocarcinoma; pleural effusion; tumorigenic; poorly differentiated	medullary carcinoma; pleural effusion; tumorigenic; poorly differentiated	metastatic adenocarcinoma; pleural effusion; tumorigenic; differentiated	carcinoma sarcoma; primary tumor; non-tumorigenic
gene expression[a]	ER (NA); PR (NA); HER2 (NA)	ER (+); PR (+); HER2 (−)	ER (−); PR (−); HER2 (+)	ER (−); PR (−); HER2 (−)	ER (−); PR (−); HER2 (−)	ER (−); PR (−); HER2 (−)
protein expression[d,e]	E-cad (+) Vim (+)	E-cad (+), NC Vim (−)	E-cad (−) Vim (−)	E-cad (−) Vim (↑)	E-cad (−) Vim (↑)	E-cad (−) Vim (↑)
doubling time[e] (h)	34	28	38	58	21	40
protein secretion rate[e,f] (μg/30 mL serum-free media/48 h)	166	180	194	165	174	100
cell viability[e]	95%	93%	97%	92%	98%	98%

[a]Assignment of subtype and gene expression are from Neve et al.; NA, not applicable; +, detectable; −, undetectable. [b]Origin of cell lines based on information from America Type Culture Collection (ATCC). [c]Tumorigenicity in nude mice. [d]Protein expression in serum-free culture media was determined (data not shown); NC, no change in expression relative to HMEC; ↑, increased expression ≥3-fold relative to HMEC. [e]Experimental data obtained in this study. [f]Obtained after 48 h of serum-free incubation.

dx.doi.org/10.1021/pr500331m | J. Proteome Res. XXXX, XXX, XXX−XXX

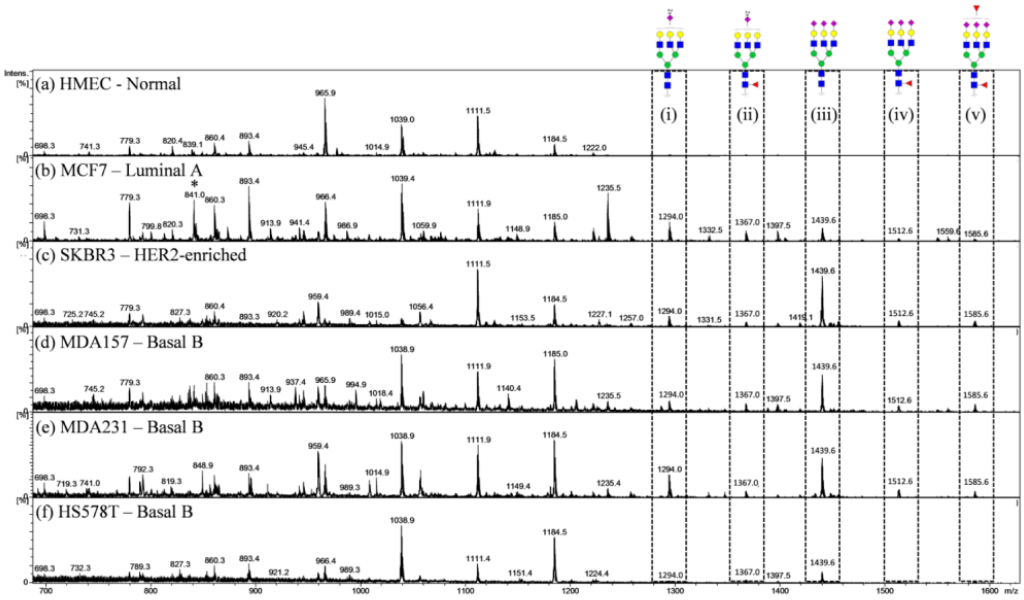


Figure 1. PGC-LC–ESI-IT-MS/MS-based mass profiles (MS1) of secreted *N*-glycans from serum-free conditioned media of cultured breast cells (a–f). *N*-Glycans with higher m/z values (highlighted in boxes, e.g., m/z 1294.0^{2-} (i), 1367.0^{2-} (ii), 1439.6^{2-} (iii), 1512.6^{2-} (iv), and 1585.6^{2-} (v)) corresponding to larger, highly sialylated and fucosylated structures were evidently missing in the non-tumorigenic HMEC cells. (* symbolizes contaminating peak of unknown origin).

RESULTS

Molecular and Cellular Characteristics of Investigated Human Breast Cells

In this study we map the *N*-glycosylation of secreted proteins (hereafter called the secretome) from a panel of human epithelial breast cell lines representing the luminal subtype (MCF7), HER2-enriched subtype (SKBR3), and the basal B subtype (MDA157, MDA231, HS578T) and compare these to a non-tumorigenic human mammary epithelial cell line (HMEC). HS578T was the only cancer cell line derived from primary breast tumors as opposed to the other five cancer cell lines that were derived from metastatic cells obtained by pleural effusion (see Table 1 for molecular and cellular characteristics of the investigated cell lines). The cellular doubling times varied dramatically, with MDA231 having the highest (21 h) while MDA157 displayed the lowest (58 h) proliferation rates. Large variations were also observed in the protein secretion rates ranging from 100 μg/30 mL serum-free media/48 h incubation time for the lowest secretor (MDA157) to 194 μg/30 mL/48 h for SKBR3. The protein secretion rate did not correlate with the cellular doubling times (R^2 = 0.048, data not shown). Loss of E-cadherin and increased vimentin expression levels have been linked to enhanced migratory and aggressive behavior in tumor cells.[23] In a separate LC–MS/MS-based proteomics analysis, E-cadherin and vimentin were both detected in HMEC, whereas only E-cadherin was detected in MCF7 with both proteins absent in SKBR3. Interestingly, the basal B cell lines were devoid of E-cadherin but exhibited more than 3-fold increased expression of vimentin compared to HMEC (Table 1), indicative of the aggressive behavior associated with the basal-like breast cancer subtype. The morphologies of investigated cells are shown in Supplemental Figure S2.

For *N*-glycan analysis, cells were cultured to grow in media supplemented with less than 5% serum for a few passages, followed by incubation in serum-free condition at subconfluency (~80%) for an additional 48 h. This was to minimize the contribution from exogenous fetal calf serum proteins in the *N*-glycome profiles, which was later confirmed by the absence of NeuGc terminating *N*-glycans in the *N*-glycome. No significant changes in the secretome *N*-glycosylation were observed 24 and 48 h after addition of serum-free media (Supplementary Figure S3). The 48 h incubation time was therefore chosen for further analysis due to the higher protein concentration in the culture media. Cell viability assays indicated minimal cell death during the stated growth conditions (92–98% cell viability, Table 1). Thus, we anticipate negligible or no contributions from intracellular *N*-glycosylated proteins released by processes such as apoptosis or cell lysis in the reported *N*-glycomes.

PGC-LC–MS/MS Based Characterization of Secretome *N*-Glycans of Human Breast Cells

N-Glycans released from the secreted glycoproteins were profiled using PGC-LC–negative ion-ESI-IT-MS/MS. The unique stereoselectivity of PGC allows excellent separation of isomeric and isobaric native *N*-glycans.[20] Structural characterization of *N*-glycans was performed partly based on intact molecular mass and partly based on diagnostic and fragment ions arising from glycosidic (B/Y- and C/Z-ions, nomenclature ion by Domon and Costello[24]) and A/X-ion cross-ring cleavages in CID MS/MS. In addition, the characterization relied heavily on the well-described relationship between *N*-glycans and their PGC-LC retention times.[25−27] For instance, fragmentation of core fucosylated *N*-glycan generated diagnostic ions corresponding to the composition of the α1,6-linked fucose (Fuc) attached to the reducing-terminal *N*-

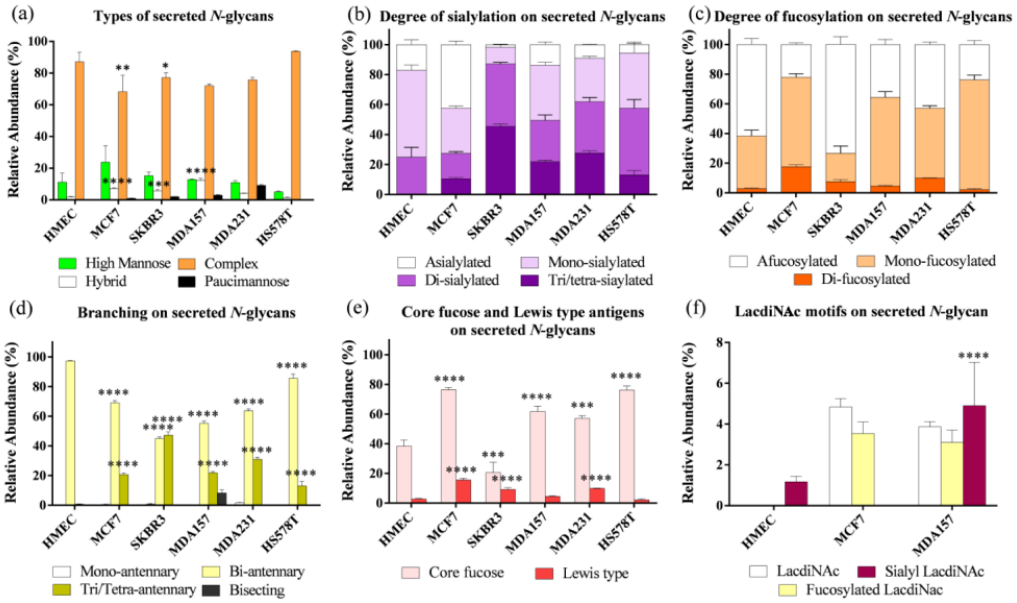


Figure 2. Comparative analyses of secreted *N*-glycans from secretomes of investigated breast cell lines. Statistical analysis, where performed, was between breast cancer cell lines and HMEC. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The relative amounts based on their relative MS intensities (EIC peak area) are presented as mean ± SD ($n = 3$). (a) Bars represent the four *N*-glycan types: high mannose, hybrid, complex, and paucimannose. For panels b–f, the relative intensities were normalized to include only hybrid and complex type *N*-glycans. (b) Degree of sialylation of secreted *N*-glycans as determined by the number of sialic acid residues present on each *N*-glycan. (c) Degree of fucosylation of secreted *N*-glycans as determined by the number of fucose residues present on each *N*-glycan. (d) Distribution of branched *N*-glycans categorized into those with mono-antennary, bi-antennary, tri/tetra-antennary, and β1,4-bisecting GlcNAc-containing *N*-glycans. (e) Distribution of fucosylated *N*-glycans represented by those carrying core fucose or terminal Lewis type antigens. (f) Distribution of LacdiNAc, sialyl LacdiNAc, and fucosyl LacdiNAc motifs in HMEC, MCF7, and MDA157.

acetylglucosamine (GlcNAc) residue (Z_1/Y_1 ion, m/z 350.1^{1-}/368.1^{1-} [Fucα1,6GlcNAc] and/or Z_2/Y_2 ion, m/z 553.3^{1-}/571.2^{1-} [GlcNAcβ1,4(Fucα1,6GlcNAc)]). These fragments were absent in MS/MS spectra from *N*-glycans lacking core fucosylation (Supplemental Figure S1, glycans 14a and 14b). Another prominent diagnostic ion denoted by Harvey[28] as D- and/or [D-18 Da] ion was used to define the antenna topology of the *N*-glycan by yielding the monosaccharide composition of the *N*-glycan 6′-arm. For example, the isomers of bi-antennary structure carrying a sialic acid residue (NeuAc) with the composition $Man_3GlcNAc_2$ (Core) + $HexNAc_2Hex_2NeuAc_1$ (m/z 965.9^{2-}) were differentiated on the basis of the D-ion (m/z 979.4^{1-}) corresponding to the fragment [NeuAcα2,3/6Galβ1,3/4GlcNAcβ1,2Manα1,6Man], indicating that the sialic acid residue is attached to the 6′-arm. In contrast, sialic acid missing on the same arm was represented by a D-ion of m/z 688.4^{1-} [Galβ1,3/4GlcNAcβ1,2Manα1,6Man] (Supplementary Figure S1, glycans 21a and 21b). Terminal galactose (Gal) or fucose residues located on the 3′- or 6′-arm were distinguished using the same approach.

Several *N*-glycans contained key terminal structures such as Lewis$^{x/a}$, LacNAc, and LacdiNAc. It is not possible with this technique to differentiate between Lewis^x [Galβ1,4(Fucα1,3)-GlcNAc] and Lewis^a determinants [Galβ1,3(Fucα1,4)GlcNAc] since their structural differences lie in the terminal fucose/galactose linkages to the antenna GlcNAc residue. Both LacNAc [Galβ1,3/4GlcNAc] and LacdiNAc [GalNAcβ1,3/4GlcNAc] are disaccharide determinants; the former has a Gal residue and latter a *N*-acetylgalactosamines (GalNAc) residue attached to the antenna GlcNAc residue. The Lewis$^{x/a}$, LacNAc and LacdiNAc determinants were discriminated by the presence of distinctive fragment ions m/z 510.3^{1-} [Galβ1,3/4(Fucα1,3/4GlcNAc)], m/z 364.1^{1-} [Galβ1,3/4GlcNAc], and m/z 405.2^{1-} [GalNAcβ1,4GlcNAc], respectively. A number of *N*-glycans with β1,4-bisecting GlcNAc were detected based on the observation of D-221 fragment ions (Supplemental Figure S1, glycans 16a, 24, 35a, 35b, 39). For bisecting GlcNAc containing *N*-glycans, the D-ion typically loses the mass corresponding to the bisecting β1,4-linked GlcNAc (221.0 Da), attached to the chitobiose core. For sialylated *N*-glycans, sialic acid residues are attached to the penultimate galactose residues via α2,3- or α2,6-linkages. These linkages were identified based on differential PGC–LC retention times, i.e., α2,6-linked sialylated structures elute significantly earlier than α2,3-linked sialoglycans[29] (Supplemental Figure S4). Using this set of fragmentation and retention time rules/knowledge, the MS/MS spectra corresponding to all reported *N*-glycans were manually annotated (Supplemental Figure S1).

N-Glycome Profiling of Breast Cell Secretomes

Mass profiles show the global *N*-glycan distribution of secreted proteins derived from HMEC and the five breast cancer cells (Figure 1). Evidently, the secreted *N*-glycomes in normal and breast cancer cell lines are different, including the unique presence of highly branched tri/tetra-antennary *N*-glycans rich in terminal sialic acid and fucose residues (glycan 40a/b, 41, 42a/b, 43a/b, and 44a/b) in all the cancer cell lines. The characterized *N*-glycans and their relative abundances are summarized in Supplemental Table S1.

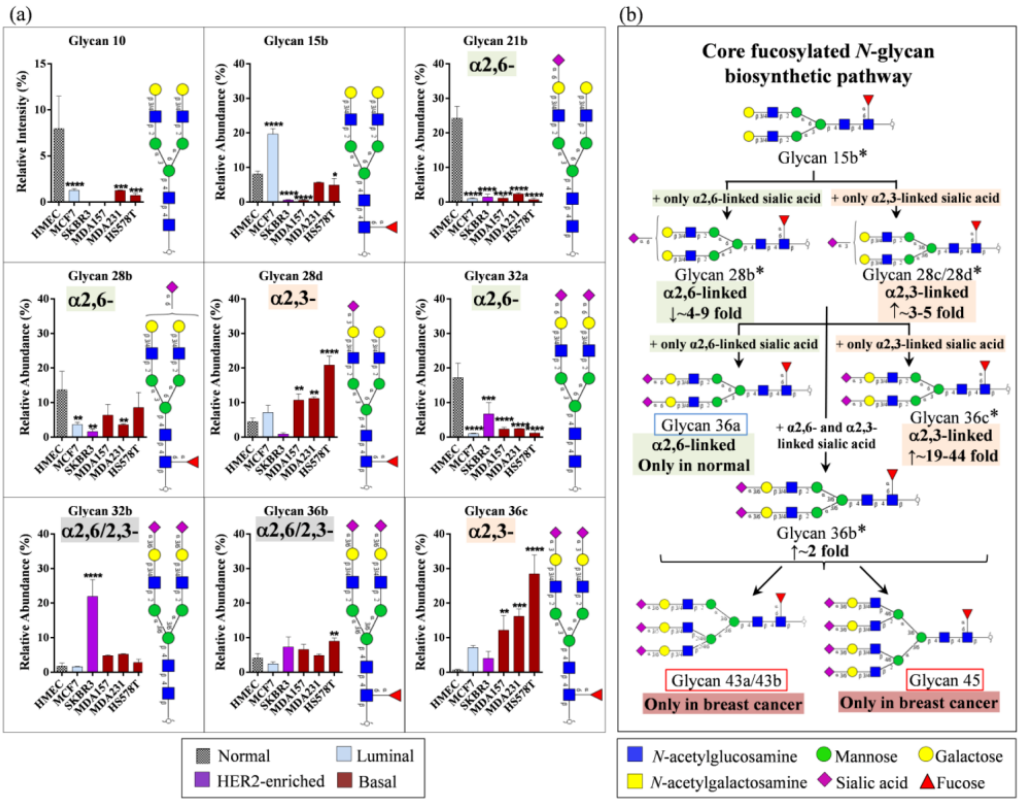


Figure 3. Differentially expressed secreted *N*-glycans in normal breast epithelial cell line (HMEC) and breast cancer cell lines. Statistical analysis was between breast cancer cell lines and HMEC. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Data are presented as mean ± SD ($n = 3$). (a) Seventeen secreted *N*-glycans were observed in both normal and breast cancer cell lines. ANOVA analysis identified nine of these to be significantly up- or down-regulated in the cancer cells as compared to the normal breast epithelial cells. (b) Of the nine regulated *N*-glycans, five were core fucosylated (*) and are mapped to their *N*-glycosylation biosynthetic pathway. Other *N*-glycans (unmarked), either not significantly regulated or found exclusively in normal (blue box) or cancer samples (red box), were also depicted to complete the pathways. The basal B breast cancer cell lines preferentially expressed α2,3-linked sialylated *N*-glycans indicated by reduced expression of α2,6-sialylation (glycans 21b, 28b, 32a) and increased expression of α2,3-sialylation (glycans 28d, 36c). Notably, the tri- and tetra-antennary structures were found only in the cancer cells, but their sialic acid linkages, however, were left undetermined. The relationship between the other four non-core fucosylated *N*-glycans are shown in Supplemental Figure S6.

In total, 45 *N*-glycan monosaccharide compositions were identified in the secretomes from all cell lines. Of these, 22 monosaccharide compositions contained two or more isomeric structures, resulting in a total of 74 *N*-glycan isomers. These structures were classified into the four *N*-glycan types according to their monosaccharide composition, i.e., high mannose, hybrid, complex, and paucimannose. The secreted *N*-glycome comprised predominantly complex-type *N*-glycans (70−90%) regardless of cell line origin, although more heterogeneous complex structures were found in the breast cancer cells (Figure 2a and Supplemental Table S2). High-mannose-type *N*-glycans were less abundant (5−26%) in the secretomes, mainly distributed over the five known high-mannose-type *N*-glycans ($Man_{5-9}GlcNAc_2$) (glycan 4, 5, 6a, 6b, 12, and 18). In the *N*-glycan biosynthetic pathway, mannose residues are sequentially trimmed from the oligomannose precursor ($Glc_3Man_9GlcNAc_2$) to $Man_5GlcNAc_2$, where a GlcNAc residue can be added to the chitoboise core structure at the 3′-arm to form the intermediate, $GlcNAcMan_5GlcNAc_2$. If the mannose residues at the 6′-arm antenna are not removed, hybrid structures are generated. However, if they are removed and replaced by a GlcNAc residue, further processing in the trans-Golgi lead to formation of complex *N*-glycans.[30] The relative unprocessed high mannose structures (e.g., $Man_{6-9}GlcNAc_2$) can be secreted if the glycosylation sites are inaccessible on the protein surface.[31,32] Unexpectedly *N*-glycan precursor with the monosaccharide composition $GlcNAc_2Hex_{10}$ (glycan 27) corresponding to the immature $Glc_1Man_9GlcNAc_2$ were detected in the MCF7 secretome. The glucose residue is normally removed during the protein folding quality control as the initial *N*-glycan processing step in the endoplasmic reticulum. Although only little cell death was observed during incubation, immature *N*-glycans could arise from intracellular sources or from glycoprotein bypassing the quality control and the downstream glycan processing. The relatively low abundance (0.3%) of this precursor argues against its biological significance. Another observation was the presence of paucimannose (0.9−9.1%) in four of the breast cancer cell lines, with the highest abundance in MDA231 (Figure 2a). This class of *N*-glycans is not commonly observed in vertebrates as it

Table 2. Overview of Alterations in Secretome *N*-Glycomes of Investigated Breast Cancer Cells Relative to Non-tumorigenic Breast Cells

	subtype		
	luminal[a]	HER2-enriched[a]	basal B[a]
	cell line MCF7	cell line SKBR3	cell lines MDA157, MDA231, HS578T
N-glycan type			
high mannose	↑	no change	no change
hybrid	↑	↑	↑
paucimannose	↑	↑	↑
branching			
bi-antennary	↓	↓	↓
tri/tetra-antennary	↑	↑	↑
bisecting GlcNAc	ND	ND	+[c]
fucosylation			
degree of fucosylation	↑	↑	↑
core fucosylation	↑	↓	↑
sialylation			
degree of sialylation	↑	↑	↑
α2,6-linked monosialylation[b]	↓	↓	↓
α2,3-linked monosialylation[b]	no change	↑↓	↑
α2,3/α2,6-linked disialylation[b]	no change	↑	↑
terminal determinants			
Lewis$^{x/a}$ determinants	↑	↑	↑
(±fucosyl)(±sialyl)(LacdiNAc)	(+)(ND)(+)	all ND	(+)(↑[c])(+)

[a]*N*-Glycan changes relative to HMEC in this study; ↑, increased expression ($p < 0.05$); ↓, decreased expression ($p < 0.05$); +, present; ND, not detected. [b]With reference to the nine significantly regulated *N*-glycans (see text). [c]Changes observed only in MDA157.

is thought that human cells lack the enzymatic capability for their synthesis.[33]

Comparative Analyses of *N*-glycan Substructures in Secretomes of Tumorigenic and Non-tumorigenic Breast Cells

High Degree of Sialylation and Fucosylation in Breast Cancer Cells. Numerous studies have suggested that the relative increase of global sialylation and fucosylation are salient protein glycosylation features associated with tumorigenesis.[8] To investigate the secretome *N*-glycans for sialylation and fucosylation, the glycan data were normalized to include only the hybrid and complex *N*-glycans, which are the two classes that receive these two terminal modifications. We observed that complex *N*-glycans were more frequently sialylated (85.2 ± 14.6%) than fucosylated (53.6 ± 25.2%) for cancer and normal cell secretomes (Figure 2b,c). Among the cancer cells, secreted *N*-glycans from MCF7 exhibited the highest levels of fucosylation (78.0 ± 1.2%), while SKBR3 had the most sialylated *N*-glycans (98.4 ± 0.3%). Interestingly, secreted *N*-glycans of MCF7 showed the least sialylation and those of SKBR3 the least fucosylation, indicating that the two modifications complement rather than stimulate each other. In-depth analysis on sialylation and fucosylation showed striking differences between normal and breast cancer cells. *N*-Glycans carrying three or more terminal sialic acid residues were prevalent in all five breast cancer cell lines (10.3−44.4%) but absent in HMEC cells (Figure 2b). Similarly, *N*-glycans with two fucose residues were over-represented in the breast cancer cell lines compared to HMEC cells (Figure 2c).

Highly Branched *N*-Glycans Were Preferentially Expressed in Breast Cancer Cells. Significantly higher levels of tri/tetra-antennary *N*-glycans were found in all the cancer cell lines compared to HMEC (Figure 2d). In the cancer cell lines, this increase was generally accompanied by a decrease in the bi-antennary structures. Of the 15 identified tri/tetra-antennary *N*-glycans, three belonged to HMEC, at a total relative abundance of less than 1.0%. The expression of the remaining 12 highly branched and mainly sialylated and fucosylated structures, unique to the breast cancer cells, ranged between 20.4% (MCF7) to 47.3% (SKBR3).

Bisecting GlcNAc-Containing *N*-Glycans Were Exclusively Expressed in MDA157. A total of five bisecting GlcNAc containing complex type *N*-glycans (glycans 16a, 24, 35a, 35b, 39) were detected exclusively in the MDA157 secretome. In the *N*-glycan biosynthetic pathway, bisecting GlcNAcylation is produced by the addition of GlcNAc residues in a β1,4-linked configuration to the β-mannose of the chitobiose core of hybrid and complex *N*-glycans. Two of the five structures were isomers (glycans 35a and 35b) discriminated by their terminal sialic acid linkages localized on the same arm. Expression of the α2,3-linked sialylated glycans was 2-fold higher compared to α2,6-linked sialylation. Bisecting GlcNAc containing *N*-glycans contributed to around 8% of the total *N*-glycome (Figure 3d). Interestingly, all of the bisecting GlcNAcylated structures were core fucosylated and structurally and biosynthetically related (Supplemental Figure S5).

Increased Core and Terminal Fucosylation in Cancer Relative to Normal Breast Cells. The majority of the monofucosylated *N*-glycans were α1,6-fucosylated at the chitobiose core. These were significantly increased in four out of the five breast cancer cell lines relative to HMEC (Figure 2e). Only SKBR3 carried significantly less core fucosylation. The multifucosylated *N*-glycans additionally have fucose linked to the antenna GlcNAc, generating the biologically important Lewis$^{x/a}$ and Lewis$^{y/b}$ determinants. The former was found in both non-tumorigenic and tumorigenic cells in this study, albeit at different amounts; only trace levels of Lewis$^{x/a}$ antigens (2−3%) were expressed in HMEC, while these were significantly

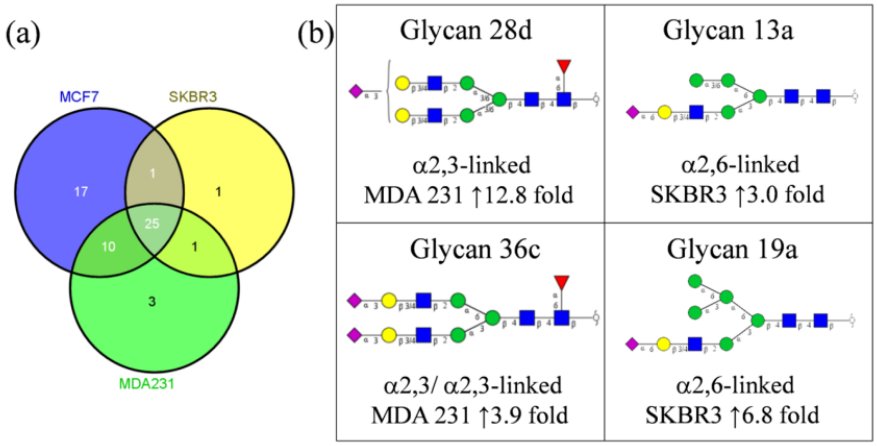


Figure 4. Breast cancer subtype-specific comparative analyses between MCF7 (luminal A), SKBR3 (HER2-enriched), and MDA231 (basal B) cell lines. (a) Venn diagram showing the common and uniquely expressed secreted *N*-glycans between the three cell lines. (b) Differentially expressed α2,3- and α2,6-linked sialylated structures in MDA231 and SKBR3. Glycans 28d and 36c, both with α2,3-linked sialylation, were increased 4- to 13-fold in MDA231 relative to SKBR3, whereas glycans 13a and 19a, both with α2,6-linked sialylation, were increased 3- to-7 fold in SKBR3 relative to MDA231.

increased (4−6-fold) in MCF7, SKBR3, and MDA231, relative to HMEC.

LacdiNAc Determinants and Derivatives Were Detected in a Subset of Breast Epithelial Cells. Several *N*-glycans displayed terminal HexNAc residues attached to the antenna GlcNAc residue. The HexNAc residues were likely to be a β1,4-linked GalNAc creating the LacdiNAc disaccharide determinant. Similar to the more common LacNAc, LacdiNAc can be further modified by either fucosylation or sialylation.[34] LacdiNAc, fucosylated LacdiNAc, and sialylated LacdiNAc determinants were detected only in a subset of the cell lines. Sialyl LacdiNAc was expressed in HMEC and MDA157, albeit at a 2-fold higher level in the latter, while the LacdiNAc and fucosylated LacdiNAc structures were observed only in MCF7 and MDA157 (Figure 2f).

Comparative Analyses of Secretome *N*-Glycome Sub-structures and Determinants in Normal and Tumorigenic Breast Cells

In order to understand how *N*-glycosylation is altered during breast cancer development, an in-depth mining of the glycomes was performed in the context of their known biosynthetic pathway.[35] Since the initial analysis indicated *N*-glycosylation alterations of the hybrid and complex *N*-glycans, we further examined these more processed *N*-glycan classes. In total, the processed *N*-glycans comprised 17 structures (14 complex and 3 hybrid types) in HMEC and at least one breast cancer cell. The complex *N*-glycans were predominantly bi-antennary structures. One-way ANOVA identified nine *N*-glycans of these 17 "common" *N*-glycans to be significantly regulated when comparing the normal and cancerous cells. These consisted of five core and four non-core fucosylated complex bi-antennary *N*-glycans (Figure 3a). The differentially expressed *N*-glycans were mapped according to their respective *N*-glycan biosynthesis (Figure 3b and Supplementary Figure S6). The breast cancer cells preferentially secreted glycoproteins containing bi-antennary *N*-glycans displaying α2,3-sialylation relative to α2,6-sialylation. The most dramatic alterations involved core fucosylated and α2,3-linked sialylated structures (i.e., glycan 28d and glycan 36c), where the expression levels increased 3−44-fold in breast cancer cells compared to HMEC. Notably, all three basal B breast cancer cells, MDA157, MDA231, and HS578T, showed significant up-regulation of both of these α2,3-sialylated *N*-glycans ($p < 0.0001$) with the overexpression of α2,3-linked sialylation most pronounced in HS578T (Figure 3a). Further evidence to support the preferred α2,3-linked sialylation in breast cancer was the simultaneous 3−33-fold reduction of α2,6-sialylated glycans (glycans 21b, 28b, and 32a) across the breast cancer cell secretomes (Figure 3b and Supplemental Figure S6). These data clearly demonstrate alteration of the *N*-glycan sialylation process in breast malignancy and in particular the observation that the basal B breast cancer cell lines heavily favor α2,3-linked sialylation over α2,6-linked sialylation. The overall changes observed between HMEC and breast cancer cell lines are summarized in Table 2.

Subtype-Specific Comparison of Secretome *N*-Glycan Substructures in MCF7, SKBR3, and MDA231 Breast Cancer Cells

Molecular subtypes of breast cancer are known to generate distinct molecular expression patterns. We conducted further analyses by comparing the secreted *N*-glycomes of MCF7, SKBR3, and MDA231, representative of luminal, HER2-enriched, and basal-like breast cancer subtypes, respectively. In total, 58 *N*-glycans were present in all three cell lines, Figure 4a. MCF7 contained the highest number of unique *N*-glycans (17). Thirty-seven *N*-glycans were shared in at least two cell lines. Statistical analyses revealed 14 structures, predominantly sialylated, were more than 3-fold regulated (Supplemental Table 3). There were three notable observations. First, sialylated structures were significantly up-regulated in SKBR3 and MDA231 compared to MCF7. However, when comparisons were made between SKBR3 and MDA231, the increased expression of sialylated glycans in SKBR3 involved the α2,6-linked sialylation (glycans 13a, 9.9-fold and 19a, 8.6-fold). In contrast, α2,3-sialylated glycans (glycan 28d, 12.0-fold and 36c, 3.9-fold) were increased in MDA231 (Figure 4b). These data supported earlier observation that the basal B cells preferentially express α2,3-linked sialylated structures. Second, SKBR3 was characterized by an increase in tri-antennary structures

(glycan 42a, 42b and 44b; 5.7- to 7.1-fold) relative to MCF7, while MDA231 exhibited more than a 4 fold increase in paucimannose (glycan 1 and 3) relative to SKBR3 and MCF7. Third, a core fucosylayted structure (glycan 15b) was elevated 3.3- to 32.2-fold in MCF 7 relative to SKBR3 and MDA231. Of the 17 uniquely expressed *N*-glycans in MCF7 not found in SKBR3 and MDA231, 16 were core fucosylated with half of these also carrying a terminal fucose (Supplemental Table 3b).

Taken together, these data suggest subtype-specific glycosylation for the secretomes: MCF7 (luminal subtype) displayed a high degree of fucosylation; SKBR3 (HER2-enriched subtype) was characterized by α2,6-linked sialylated and triantennary structures; and MDA231 (basal-like subtype) was characterized by α2,3-linked sialylated and paucimannose structures.

Correlation of Transcriptional Levels of Selected Glycosyltransferases and the Differential Expression of Secretome *N*-Glycosylation in MCF7, SKBR3, and MDA231

We sought to investigate the relationship between the altered *N*-glycosylation and changes in the associated glycosylation enzyme levels using an *in silico* approach to mine available transcriptome data for the three breast cancer cell lines. The latter was performed on the subset of glycosyltransferases associated with the α2,3-/α2,6-linked sialylation and core fucosylation. Both SKBR3 and MDA231 showed increased sialyltransferase expression relative to MCF7 involving either the α2,3- or α2,6-linked sialylation, but not both together, Table 3. In SKBR3, an increase in both ST6Gal1 and ST3Gal2/4 was observed; however, ST6Gal1 showed a much higher (30-fold) expression compared to the 3- to 4-fold increase in ST3Gal2/4. In contrast, ST3Gal2/3 expression was increased 8- to 33-fold but displayed no change in ST6Gal 1 expression in MDA231 relative to MCF7. When compared to SKBR3, MDA231 had a 54-fold reduction in ST6Gal1 expression. These data validated the observation that the secreted *N*-glycome of SKBR3 was characterized by enhanced expression of α2,6-linked sialylation, whereas MDA231 typically comprised higher levels of α2,3-linked sialylation. In addition, the high degree of fucosylation observed on secretome *N*-glycans of MCF7 correlated well with the 3- to 7-fold increase in fucosyltransferase 8 (Fut8) expression when compared to those of SKBR3 and MDA231.

Table 3. Comparative Analyses of Transcriptional Levels (mRNA)[21] of Selected Glycosyltransferases in MCF7, SKBR3, and MDA231[a]

gene name	SKBR3 vs MCF7 (HER2-enriched vs luminal A)	MDA231 vs MCF7 (basal B vs luminal A)	MDA231 vs SKBR3 (basal B vs HER2-enriched)
ST3Gal2	↑4.3	↑33.4	↑7.6
ST3Gal3	N/A	↑8.1	N/A
ST3Gal4	↑2.8	no change	↓5.1
ST3Gal6	N/A	N/A	no change
ST6Gal1	↑30.0	no change	↓54.8
Fut8	↓7.4	↓3.4	↑2.2

[a]Transcription fold change indicated.

Secretome *N*-Glycome-Based Clustering of Tumorigenic and Cancer Subtypes

Breast tumors are normally classified into various subtypes according to their gene expression profiles. To assess whether the known differences in genotype, phenotype, and growth characteristics of the investigated breast epithelial cells correlated with their *N*-glycosylation, a cluster analysis was performed using log-transformed relative abundances of the observed *N*-glycans for each secretome and by applying hierarchical clustering with Pearson correlation (Figure 5a). Two major clusters were observed, one representing the non-tumorigenic breast cell (HMEC) and another representing the five tumorigenic breast cells. Within the tumorigenic samples, the division between luminal and basal subtypes was easily discerned with the basal B breast cancer cell lines (MDA157, MDA231, and HS578T) clustering separately from the luminal A breast cells (MCF7) and SKBR3 (HER2-enriched).

The sample inter-relationship was also examined using principal component analysis (PCA). In PCA, new sets of variables are created as linear combinations of the original sets of variables (*N*-glycans) to reduce data dimensionality while capturing the directions of most variability. The coefficients of these linear combinations are the principal component *loadings*, and the values of these new combinations are the component *scores*. The scores of the first two or three principal components can be visualized on 2D or 3D plots.

PCA in a 3D plot revealed unique molecular features in the *N*-glycomes that not only discriminated cancerous from non-cancerous cells but also separated luminal A and HER2-enriched subtypes from those of basal B subtypes (Figure 5b). Within the three basal B cell lines, HS578T differed from the two more similar MDA157 and MDA231 cell lines. A possible explanation for this division may be that HS578T was established from a primary breast tumor, while MDA157 and MDA231 were derived from pleural effusions of breast cancer patients, therefore representing the metastatic state of the disease.

To investigate which secreted *N*-glycans contributed with the largest variance in the first principal component, data with the highest loadings (>0.1 and <0.1) were extracted. Twelve secreted *N*-glycans were identified to be the dominant features, creating distinctive patterns between the six breast cell secretomes (Supplemental Figure S7). These included the nine differentially expressed *N*-glycans between the non-tumorigenic and tumorigenic cell lines (Figure 4a) and the eight glycans that were differentially expressed between SKBR3, MDA231, and MCF7 (Supplemental Table S3).

These distinctive features suggest that the secretome *N*-glycosylation may serve as a potential diagnostic feature in clinical settings to differentiate between healthy and breast cancer tissues; between luminal A, HER2-enriched, and basal-like subtypes; and possibly at the early stages of disease progression.

■ DISCUSSION

Although altered *N*-glycans have been implicated in breast malignancy, to our knowledge, the secretome *N*-glycosylation of breast cancer cell lines has not been extensively investigated. Breast cancer cell lines remain a valuable *in vitro* model for exploring the biology of cancer through their ability to recapitulate both the distinctive normal/cancer and the luminal/basal subtype division observed in breast tumor tissues.[36] We demonstrate that the secretome *N*-glycosylation is capable of distinguishing normal and tumorigenic cells as well as the tumorigenic subtypes belonging to luminal, HER2-enriched, and basal types. Nevertheless, there are limitations associated with cell lines, for example, the risk of contamination from exogenous sources or intracellular proteins arising from

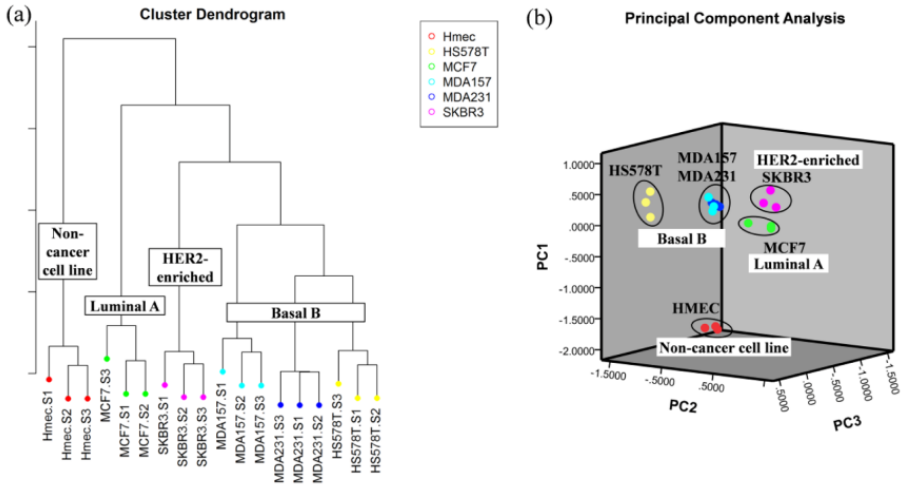


Figure 5. Dendrogram cluster analysis using hierarchical clustering (a) and 3D plot of PCA (b) of secretome *N*-glycosylation profiles of the six epithelial breast cells investigated in this study. PC1, principal component 1; PC2, principal component 2; PC3, principal component 3.

cell death. However, such challenges can be overcome with proper experimental design and careful interpretation of results. Taking these into account, our data provide validity for using secretions of established breast epithelial cell lines to investigate the underlying molecular mechanisms associated with subtype-specific glycosylation changes.

On the basis of the hypothesis that breast tumor cells secrete proteins carrying altered *N*-glycosylation into the circulation, secretions from breast cancer cell lines can provide an attractive source for detection of potential glycan-based biomarkers, in addition to using more complex serum samples. Although the use of serum is clinically feasible as it can be obtained noninvasively, it is a highly complex biological fluid with protein concentration that spans at least 10 orders of magnitude. Proteomic analysis has shown that high abundance glycoproteins, many of which are acute phase proteins produced from the liver in response to inflammation, are manifested in the serum profile of cancer patients.[37] Additionally, the presence of low abundance glycoproteins secreted from tumor cells may be masked by high abundance proteins, thus precluding the detection of *N*-glycans released from these proteins. Our data showed significant breast cancer-specific alterations of protein *N*-glycosylation, such as a high degree of sialylation, fucosylation, and branching, which correlate well with changes observed in the serum of breast cancer patients.[16,38,39] However, we also identified some uncommon (sub)structures in the secretome *N*-glycomes of the investigated breast cancer cell lines, structures that only rarely have been described in clinical samples and/or *in vitro* studies, including LacdiNAc and related fucosylated and sialylated derivatives. Recent studies have suggested the involvement of LacdiNAc determinants in ovarian cancers.[40] However, the biological roles and significance of these structures in cancerous breast cells remain unclear and require further investigation.

The unique glycan structures observed in our breast cancer cell line models, which have not been reported in previous serum studies, provide further suggestion for the apparent usefulness of cell lines as *in vitro* models for biomarker discovery studies. Additionally, proteomics analysis of the secretome generated from the *in vitro* system has mapped a significant proportion of these proteins to those found in the *in vivo* environment.[41,42] In a separate analysis of the secreted proteome of the breast epithelial cells, we found over 80% overlap with those in the Plasma Proteome Database (http://www.plasmaproteome database.org/) (data not shown), supporting the biological relevance of secreted glycoproteins from cultured cells. However, we were restricted to a panel of just six breast cell lines in this study; thus, it may be necessary to validate these observations with a larger panel of breast cell types and to further assess their potential clinical utility. Nevertheless, our data provide a detailed map of the cancer-specific *N*-glycan alterations in the secretome across the different breast cancer subtypes, which may help substantiate the findings from serum-based discovery studies.

It has been suggested that identifying suitable biomarkers for breast cancer may be subtype-dependent as the individual subtypes generate distinct molecular profiles.[43] Our findings on the expression of specific *N*-glycan substructures within a subset of cell lines and the characteristic *N*-glycan changes between the three subtypes underscore the importance of subtype-specific studies. LacdiNAc and its fucosylated form were exclusively expressed in MCF7 and MDA157, whereas bisecting β1,4-GlcNAc structures were detected only in MDA157. Notably, the following subtype-specific expression patterns were evident: increased α2,6-linked sialylation associated with HER2-enriched cells; increased α2,3-linked sialylation associated with basal B cells; and increased α1,6-fucosylation associated with luminal A cells. These results were corroborated by strong correlation to the transcription of the relevant glycosylation enzymes. Similar to our observation, high expression of α2,3-linked sialylation in MDA231 relative to MCF7 has been previously reported.[17] Interestingly, elevated α2,6-linked sialylation was observed in the glycome of sialylated glycoproteins derived from serum of breast cancer patients.[44] However, this study, as with many other serum-based breast cancer glycome investigations, was based on stages of breast cancer progression and not subtype-specificity. Given that sialylated structures are involved in tumor invasion and metastasis, the differential expression of serum *N*-glycan structures bearing these two linkages in the different subtypes

of breast cancer warrant further investigation. As expression of these structures may be driven by higher activity of associated sialyltransferases, as indicated by our data, it is also useful to explore the potential of these enzymes as biomarkers or drug targets.

CONCLUSION

This is the first study to profile and compare the *N*-glycosylation of secretomes from a panel of normal and cancerous human breast epithelial cells. Significant alterations and unique *N*-glycosylation signatures, as a consequence of breast malignancy, were observed. Overall, an increase in sialylation, fucosylation, and highly branched structures were observed in the cultured breast cancer cells. Expression of bisecting GlcNAc and LacdiNAc and its derivative structures was restricted to MDA157 and MCF7. In addition, we observed unique subtype-specific *N*-glycosylation. These global and subtype-specific *N*-glycan changes could help delineate differences between non-tumorigenic and tumorigenic cellular mechanisms and distinguish breast cancer cells of luminal, HER2-enriched, and basal B subtypes. In turn, this knowledge may lead to better understanding of the molecular mechanisms involved in breast cancer and further strengthen the validity of potential biomarkers mined from serum-based glycome studies.

ASSOCIATED CONTENT

Supporting Information

Supplementary Figures S1–S7 and Tables S2 and S3 as described in the main text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: 61-2-98508260. Fax: 61-2-98508313. E-mail: susan.fanayan@mq.edu.au.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research project was facilitated by access to the Australian Proteomics Analysis Facility (APAF) established under the Australian Government's NCRIS program. We thank Dana Pascovici for valuable assistance in performing the clustering analysis. This project was supported by Macquarie University Research Excellence Scheme postgraduate scholarship and the Northern Translational Cancer Research Unit (Kolling Institute of the University of Sydney) through a Cancer Institute NSW competitive grant. M.T.-A. was funded by an Early Career Fellowship Grant from the Cancer Institute NSW. The views expressed herein are those of the authors and are not necessarily those of the Cancer Institute NSW.

ABBREVIATIONS

LC, liquid chromatography; CID, collision-induced fragmentation; EIC, extracted ion chromatogram; MS/MS, tandem mass spectrometry; ESI, electrospray ionization; RT, room temperature; PNGase F, peptide-*N*-glycosidase F; HexNAc, *N*-acetylhexoseamine; GlcNAc, *N*-acetyl-glucosamine; LacdiNAc, *N*,*N*′-diacetyllactosamine; NeuAc, *N*-acetylneuraminic acid; NeuGc, *N*-glycolylneuraminic acid

REFERENCES

(1) *The global burden of disease: 2004 update*; World Health Organization: Geneva, 2013.

(2) Siegel, R.; Ma, J.; Zou, Z.; Jemal, A. Cancer statistics, 2014. *Ca–Cancer J. Clin.* **2014**, *64* (1), 9–29.

(3) Higgins, M. J.; Baselga, J.; xE. Targeted therapies for breast cancer. *J. Clin. Invest.* **2011**, *121* (10), 3797–3803.

(4) Perou, C.; Sorlie, T.; Eisen, M.; Rijn, M.; Jeffrey, S.; Rees, C.; Pollack, J.; Ross, D.; Johnsen, H.; Akslen, L.; Fluge, O.; Pergamenschikov, A.; Williams, C.; Zhu, S.; Lonning, P.; Borresen-Dale, A.; Brown, P.; Botstein, D. Molecular portraits of human breast tumours. *Nature* **2000**, *406* (6797), 747–752.

(5) Neve, R. M.; Chin, K.; Fridlyand, J.; Yeh, J.; Baehner, F. L.; Fevr, T.; Clark, L.; Bayani, N.; Coppe, J.-P.; Tong, F.; Speed, T.; Spellman, P. T.; DeVries, S.; Lapuk, A.; Wang, N. J.; Kuo, W.-L.; Stilwell, J. L.; Pinkel, D.; Albertson, D. G.; Waldman, F. M.; McCormick, F.; Dickson, R. B.; Johnson, M. D.; Lippman, M.; Ethier, S.; Gazdar, A.; Gray, J. W. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **2006**, *10* (6), 515–527.

(6) Badve, S.; Dabbs, D. J.; Schnitt, S. J.; Baehner, F. L.; Decker, T.; Eusebi, V.; Fox, S. B.; Ichihara, S.; Jacquemier, J.; Lakhani, S. R.; Palacios, J.; Rakha, E. A.; Richardson, A. L.; Schmitt, F. C.; Tan, P.-H.; Tse, G. M.; Weigelt, B.; Ellis, I. O.; Reis-Filho, J. S. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod. Pathol.* **2011**, *24* (2), 157–167.

(7) Duffy, M. J. Serum tumor markers in breast cancer: Are they of clinical value? *Clin. Chem.* **2006**, *52* (3), 345–351.

(8) Christiansen, M. N.; Chik, J.; Lee, L.; Anugraham, M.; Abrahams, J. L.; Packer, N. H. Cell surface protein glycosylation in cancer. *Proteomics* **2013**, *14*, 525–46.

(9) Sethi, M. K.; Thaysen-Andersen, M.; Smith, J. T.; Baker, M. S.; Packer, N. H.; Hancock, W. S.; Fanayan, S. Comparative N-glycan profiling of colorectal cancer cell lines reveals unique bisecting GlcNAc and α-2,3-linked sialic acid determinants are associated with membrane proteins of the more metastatic/aggressive cell lines. *J. Proteome Res.* **2013**, *13* (1), 277–288.

(10) Nakano, M.; Nakagawa, T.; Ito, T.; Kitada, T.; Hijioka, T.; Kasahara, A.; Tajiri, M.; Wada, Y.; Taniguchi, N.; Miyoshi, E. Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: A novel approach for the development of tumor markers. *Int. J. Cancer* **2008**, *122* (10), 2301–2309.

(11) Liu, X.; Nie, H.; Zhang, Y.; Yao, Y.; Maitikabili, A.; Qu, Y.; Shi, S.; Chen, C.; Li, Y. Cell surface-specific N-glycan profiling in breast cancer. *PLoS One* **2013**, *8* (8), e72704.

(12) Anugraham, M.; Jacob, F.; Nixdorf, S.; Everest-Dass, A. V.; Heinzelmann-Schwarz, V.; Packer, N. H. Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: glycan structures reflect gene expression and DNA methylation status. *Mol. Cell. Proteomics* **2014**, DOI: 10.1074/mcp.M113.037085.

(13) Saldova, R.; Fan, Y.; Fitzpatrick, J. M.; Watson, R. W. G.; Rudd, P. M. Core fucosylation and α2–3 sialylation in serum N-glycome is significantly increased in prostate cancer comparing to benign prostate hyperplasia. *Glycobiology* **2011**, *21* (2), 195–205.

(14) Comunale, M. A.; Wang, M.; Hafner, J.; Krakover, J.; Rodemich, L.; Kopenhaver, B.; Long, R. E.; Junaidi, O.; Bisceglie, A. M. D.; Block, T. M.; Mehta, A. S. Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. *J. Proteome Res.* **2008**, *8* (2), 595–602.

(15) Thaysen-Andersen, M.; Packer, N. H. Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim. Biophys. Acta, Proteins Proteomics* **2014**, *1844* (9), 1437–1452.

(16) Saldova, R.; Asadi Shehni, A.; Haakensen, V. D.; Steinfeld, I.; Hilliard, M.; Kifer, I.; Helland, Å.; Yakhini, Z.; Børresen-Dale, A.-L.; Rudd, P. M. Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC. *J. Proteome Res.* **2014**, *13* (5), 2314–2327.

(17) Cui, H.; Lin, Y.; Yue, L.; Zhao, X.; Liu, J. Differential expression of the alpha2,3-sialic acid residues in breast cancer is associated with metastatic potential. *Oncol. Rep.* **2011**, *25* (5), 1365−71.
(18) de Leoz, M. L. A.; Young, L. J. T.; 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* (1), No. M110.002717.
(19) Wang, P. H. Altered glycosylation in cancer: Sialic acids and sialyltransferases. *J. Cancer Molecules* **2005**, *1* (2), 73−81.
(20) Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protoc.* **2012**, *7* (7), 1299−1310.
(21) Heiser, L. M.; Sadanandam, A.; Kuo, W. L.; Benz, S. C.; Goldstein, T. C.; Ng, S.; Gibb, W. J.; Wang, N. J.; Ziyad, S.; Tong, F.; Bayani, N.; Hu, Z.; Billig, J. I.; Dueregger, A.; Lewis, S.; Jakkula, L.; Korkola, J. E.; Durinck, S.; Pepin, F.; Guan, Y.; Purdom, E.; Neuvial, P.; Bengtsson, H.; Wood, K. W.; Smith, P. G.; Vassilev, L. T.; Hennessy, B. T.; Greshock, J.; Bachman, K. E.; Hardwicke, M. A.; Park, J. W.; Marton, L. J.; Wolf, D. M.; Collisson, E. A.; Neve, R. M.; Mills, G. B.; Speed, T. P.; Feiler, H. S.; Wooster, R. F.; Haussler, D.; Stuart, J. M.; Gray, J. W.; Spellman, P. T. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (8), 2724−9.
(22) Chik, J. H. L.; Zhou, J.; Moh, E. S. X.; Christopherson, R.; Clarke, S. J.; Molloy, M. P.; Packer, N. H. Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies. *J. Proteomics* **2014**, *108* (0), 146−162.
(23) Kokkinos, M. I.; Wafai, R.; Wong, M. K.; Newgreen, D. F.; Thompson, E. W.; Waltham, M. Vimentin and epithelial-mesenchymal transition in human breast cancer − Observations in vitro and in vivo. *Cells Tissues Organs* **2007**, *185* (1−3), 191−203.
(24) Domon, B.; Costello, C. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate J.* **1988**, *5* (4), 397−409.
(25) Harvey, D. J. Fragmentation of negative ions from carbohydrates: Part 3. Fragmentation of hybrid and complex N-linked glycans. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (5), 647−659.
(26) Harvey, D. J.; Jaeken, J.; Butler, M.; Armitage, A. J.; Rudd, P. M.; Dwek, R. A. Fragmentation of negative ions from N-linked carbohydrates, Part 4. Fragmentation of complex glycans lacking substitution on the 6-antenna. *J. Mass Spectrom.* **2010**, *45* (5), 528−535.
(27) Harvey, D. J.; Rudd, P. M. Fragmentation of negative ions from N-linked carbohydrates. Part 5: Anionic N-linked glycans. *Int. J. Mass Spectrom.* **2011**, *305* (2−3), 120−130.
(28) Harvey, D. J. Fragmentation of negative ions from carbohydrates: Part 2. Fragmentation of high-mannose N-linked glycans. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (5), 631−646.
(29) Nakano, M.; Saldanha, R.; Gobel, A.; Kavallaris, M.; Packer, N. H. Identification of glycan structure alterations on cell membrane proteins in desoxyepothilone B resistant leukemia cells. *Mol. Cell Proteomics* **2011**, *10* (11), M111 009001.
(30) Aebi, M. N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta, Mol. Cell Res.* **2013**, *1833* (11), 2430−2437.
(31) Thaysen-Andersen, M.; Packer, N. H. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* **2012**, *22* (11), 1440−1452.
(32) Lee, L. Y.; Lin, C.-H.; Fanayan, S.; Packer, N. H.; Thaysen-Andersen, M. Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. *Frontiers in Immunology* **2014**, DOI: 10.3389/fimmu.2014.00404.
(33) Schachter, H. Paucimannose N-glycans in Caenorhabditis elegans and Drosophila melanogaster. *Carbohydr. Res.* **2009**, *344* (12), 1391−1396.
(34) Kawar, Z. S.; Haslam, S. M.; Morris, H. R.; Dell, A.; Cummings, R. D. Novel poly-GalNAcβ1−4GlcNAc (LacdiNAc) and fucosylated poly-LacdiNAc N-glycans from mammalian cells expressing β1,4-N-acetylgalactosaminyltransferase and α1,3-Fucosyltransferase. *J. Biol. Chem.* **2005**, *280* (13), 12810−12819.
(35) Potapenko, I. O.; Haakensen, V. D.; Luders, T.; Helland, A.; Bukholm, I.; Sorlie, T.; Kristensen, V. N.; Lingjaerde, O. C.; Borresen-Dale, A. L. Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Mol. Oncol.* **2010**, *4* (2), 98−118.
(36) Prat, A.; Karginova, O.; Parker, J. S.; Fan, C.; He, X.; Bixby, L.; Harrell, J. C.; Roman, E.; Adamo, B.; Troester, M.; Perou, C. M. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res. Treat* **2013**, *142* (2), 237−55.
(37) Diamandis, E. P.; van der Merwe, D.-E. Plasma protein profiling by mass spectrometry for cancer diagnosis: Opportunities and limitations. *Clin. Cancer Res.* **2005**, *11* (3), 963−965.
(38) Saldova, R.; Reuben, J. M.; Abd Hamid, U. M.; Rudd, P. M.; Cristofanilli, M. Levels of specific serum N-glycans identify breast cancer patients with higher circulating tumor cell counts. *Ann. Oncol* **2011**, *22* (5), 1113−9.
(39) Pierce, A.; Saldova, R.; Abd Hamid, U. M.; Abrahams, J. L.; McDermott, E. W.; Evoy, D.; Duffy, M. J.; Rudd, P. M. Levels of specific glycans significantly distinguish lymph node-positive from lymph node-negative breast cancer patients. *Glycobiology* **2010**, *20* (10), 1283−1288.
(40) Machado, E.; Kandzia, S.; Carilho, R.; Altevogt, P.; Conradt, H. S.; Costa, J. N-Glycosylation of total cellular glycoproteins from the human ovarian carcinoma SKOV3 cell line and of recombinantly expressed human erythropoietin. *Glycobiology* **2011**, *21* (3), 376−86.
(41) Kulasingam, V.; Diamandis, E. P. Proteomics analysis of conditioned media from three breast cancer cell lines. *Mol. Cell. Proteomics* **2007**, *6* (11), 1997−2011.
(42) Pavlou, M. P.; Dimitromanolakis, A.; Diamandis, E. P. Coupling proteomics and transcriptomics in the quest of subtype-specific proteins in breast cancer. *Proteomics* **2013**, *13* (7), 1083−95.
(43) Gonzalez, R. M.; Daly, D. S.; Tan, R.; Marks, J. R.; Zangar, R. C. Plasma biomarker profiles differ depending on breast cancer subtype but RANTES is consistently increased. *Cancer Epidemiol., Biomarkers Prev.* **2011**, *20* (7), 1543−1551.
(44) Alley, W. R., Jr.; Novotny, M. V. Glycomic analysis of sialic acid linkages in glycans derived from blood serum glycoproteins. *J. Proteome Res.* **2010**, *9* (6), 3062−72.

CHAPTER 3

Part 2

N-glycome analysis of membrane proteins from a panel of breast epithelial cell lines

3.1 Introduction

Membrane proteins are frequently *N*-glycosylated as they traffic through the ER-Golgi secretory pathway to their final destination on the cell surface or intracellular organelles. They perform diverse biological functions essential to cellular growth and development; many of these processes are mediated or modulated by the conjugated *N*-glycans. It is thus not too surprising that aberrant *N*-glycosylation of membrane proteins have been linked to various pathological conditions including the promotion of tumor invasion and metastatic behaviours in many human cancers [341]. Delineating the aberrant *N*-glycosylation may therefore shed light on the underlying mechanisms in tumorigenesis. Additionally, it may unravel *N*-glycan expression patterns predictive of breast cancer specific stage of the disease or identify certain *N*-glycan changes amenable to drug treatments.

Part 1 of this chapter investigated the *N*-glycosylation expression patterns on secreted *N*-glycorproteins from a panel of cultured breast epithelial cells, including the non-tumorigenic breast epithelial cell line (HMEC) and five tumorigenic cell lines representing three common breast cancer subtypes, namely MCF7 for luminal A subtype, SKBR3 for HER2-enriched subtype and MDA231, MDA157 and HS578T for basal B subtype. The results strongly underpinned the role of altered *N*-glycosylation in breast tumorigenesis and identified breast cancer subtype-specific *N*-glycan patterns. In Part 2 of this chapter, the *N*-glycosylation of membrane glycoproteins from the same set of cell lines was investigated for tumor-defining features. In addition, the *N*-glycan expression patterns between the two subcellular fractions are compared.

3.2 Materials and Methods

Following the collection of conditioned media from the breast epithelial cells cultured under serum-free conditions, membrane proteins were extracted from the cells using

ultracentrifugation, followed by Triton X-114 phase partitioning (described in the Chapter 2, Page 90).

Approximately 50 μg of membrane proteins were used for each *N*-glycome preparation. The methodologies used for the membrane *N*-glycan analysis including the enzymatic release from enriched membrane proteins, reduction and desalting of released *N*-glycans, their separation and analysis on PGC-LC-MS/MS and finally the data and statistical analysis were identical to those applied to the secreted *N*-glycans (Part 1, Chapter 3).

3.3 Results

3.3.1 Comparative analyses of *N*-glycan sub-structures on membrane proteins from tumorigenic and non-tumorigenic breast cells

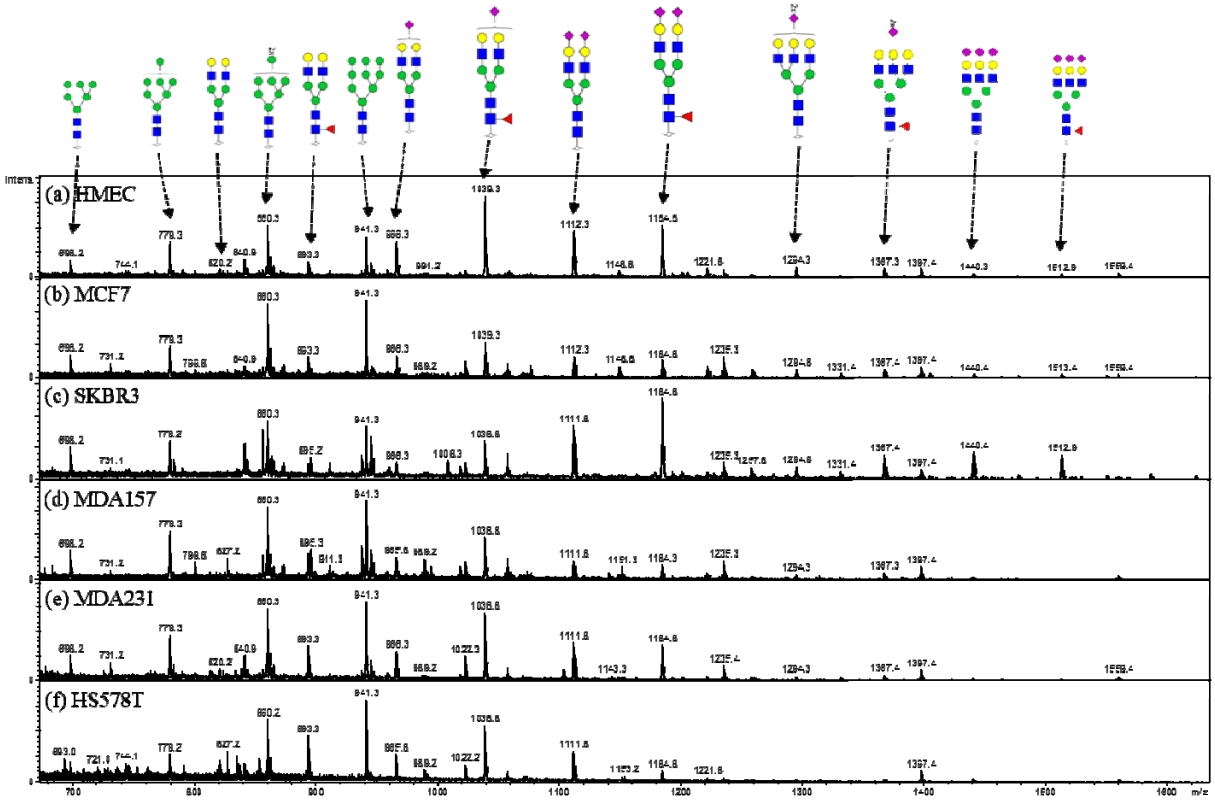


Figure 3.1 PCG-LC-MS/MS-based MS1 mass profiles of *N*-glycans released from membrane proteins derived from cultured breast cells (a-f). Although no apparent qualitative differences between HMEC (a) and the breast cancer cell lines (b-f) were observed, in-depth quantitative analysis revealed notable tumor phenotypic specific alterations (see text).

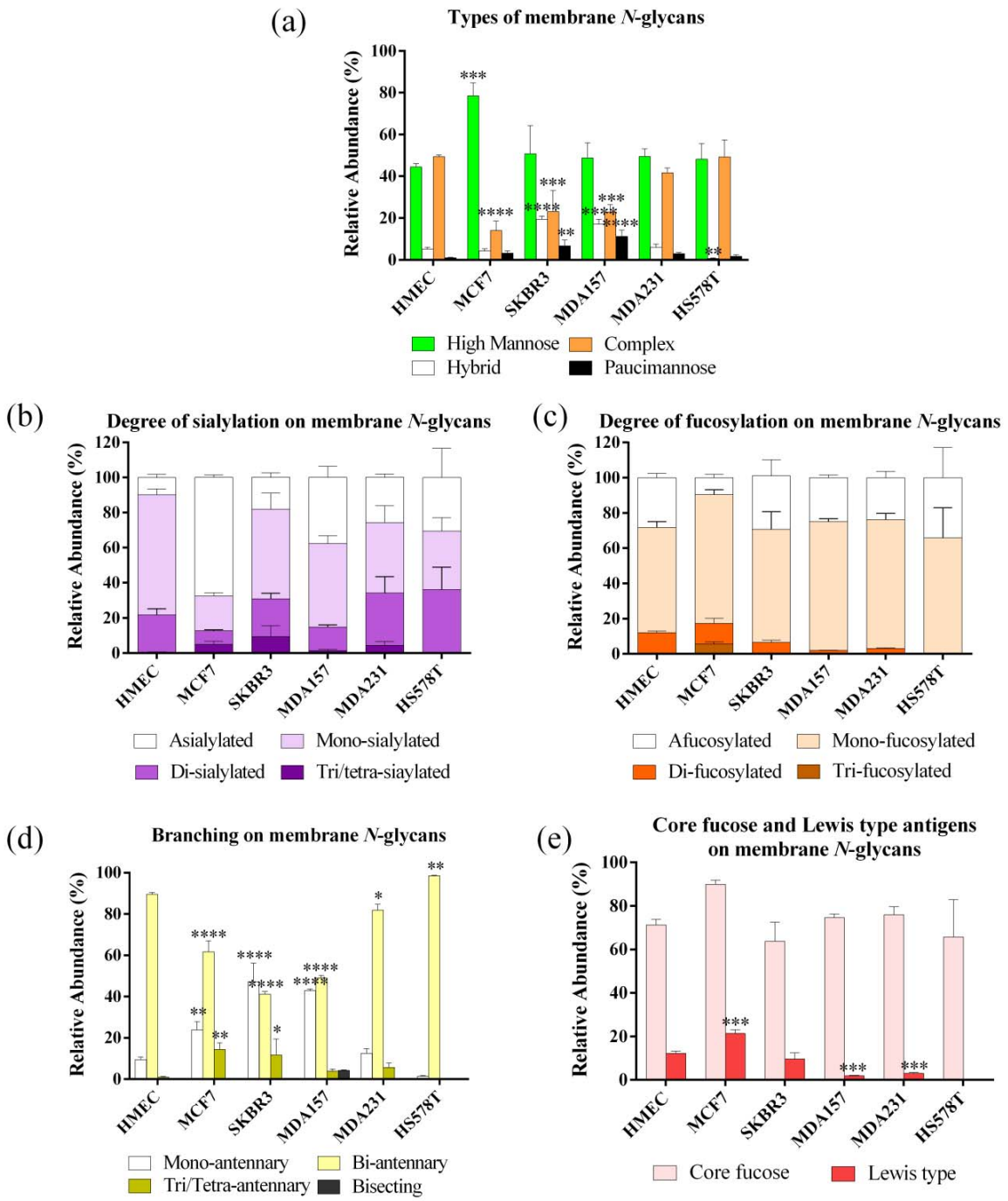


Figure 3.2 Comparative analyses of membrane *N*-glycan of investigated breast cell lines. Statistical analysis was between breast cancer cell lines and HMEC. **** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05. The relative amounts based on their relative MS intensities (EIC peak area) are presented as mean ± SD (n = 3). (a) Bars represent the four *N*-glycan types: high mannose, hybrid, complex, and paucimannose. For the panels b−e, the relative intensities were normalized to include only hybrid and complex type *N*-glycans. (b) The degree of sialylation of membrane *N*-glycans as determined by the number of sialic acid residues present on each *N*-glycan. (c) The degree of fucosylation of membrane *N*-glycans as determined by the number of fucose residues present on each *N*-glycan. (d) The distribution of branched *N*-glycans categorized into those with mono-antennary, bi-antennary, tri/tetra-antennary, and β1,4-bisecting GlcNAc-containing *N*-glycans. (e) Distribution of fucosylated *N*-glycans represented by those carrying core fucose or terminal Lewis type antigens.

Unlike the *N*-glycomes derived from the secreted proteins of epithelial breast cells, no major discriminatory features were directly visible in the MS1 mass profiles of membrane *N*-glycans between the non-tumorigenic breast epithelial cell line (HMEC) and the five breast cancer cell lines (Figure 3.1). A total of 34 *N*-glycan monosaccharide compositions released from the membrane glycoprotein samples. These *N*-glycans were categorized into the commonly-observed mammalian *N*-glycan types, i.e. high mannose, complex, hybrid, and paucimannose (Figure 3.2a). Using the same approach as for the secreted *N*-glycans (See publication II), the relative abundances of the membrane *N*-glycan structures were determined and the MS/MS spectra for each structure was manually assigned, yielded a total of 53 *N*-glycan structures, including 15 observed isomers (Appendix 4 and 4b). The abundant *N*-glycans observed in all cell lines were of the high mannose types constituting 44.4-78.5% (mol/mol) of the entire *N*-glycome followed by the less abundant complex types (14.1-49.4%). Amongst the six cell lines, MCF7 membrane proteins exhibited the highest expression of high mannose type *N*-glycan. When compared to HMEC normal reference, membrane proteins of MCF7, SKBR3, and MDA157 displayed significantly reduced levels of complex type *N*-glycans. In comparison to high mannose and complex types, the hybrid and paucimannose type membrane *N*-glycans were less abundant in all cell lines (hybrid, 0.5-20.3%; paucimannose, 1.0-11.2%). However, both SKBR3 and MDA157 had significantly higher expression of these two *N*-glycan types relative to HMEC.

N-glycans that have do not undergo complete processing during the *N*-glycan biosynthesis to form hybrid and complex structures will remain as the high mannose type containing five to nine mannose residues. By normalizing the membrane *N*-glycan data to total abundance of high mannose type structures, the internal distribution of the high mannose *N*-glycan species were mapped to four groups, i.e. *N*-glycans with five and six mannose residues (Man_5/Man_6); seven mannose residues (Man_7); eight mannose residues (Man_8); and nine mannose residues with or without the terminal glucose residue ($Man_9 \pm Glc_1$) (Figure 3.3). The latter is considered as an

immature *N*-glycan form thereby yielding an indication of the amount of intracellular *N*-glycosylation. At least two thirds of high mannose *N*-glycans was distributed into Man_8 and $Man_9 \pm Glc_1$ categories indicating little *N*-glycan processing of the *N*-glycosylation presented on the membrane proteome. The doubling time and the secretion rate for each cell line were investigated (Part 1, Chapter 3, page 133). No correlation was found between high mannose *N*-glycan expression and rate of cell growth whereas a weak correlation was observed between the secretion rate of the secretome and the *N*-glycosylation processing.

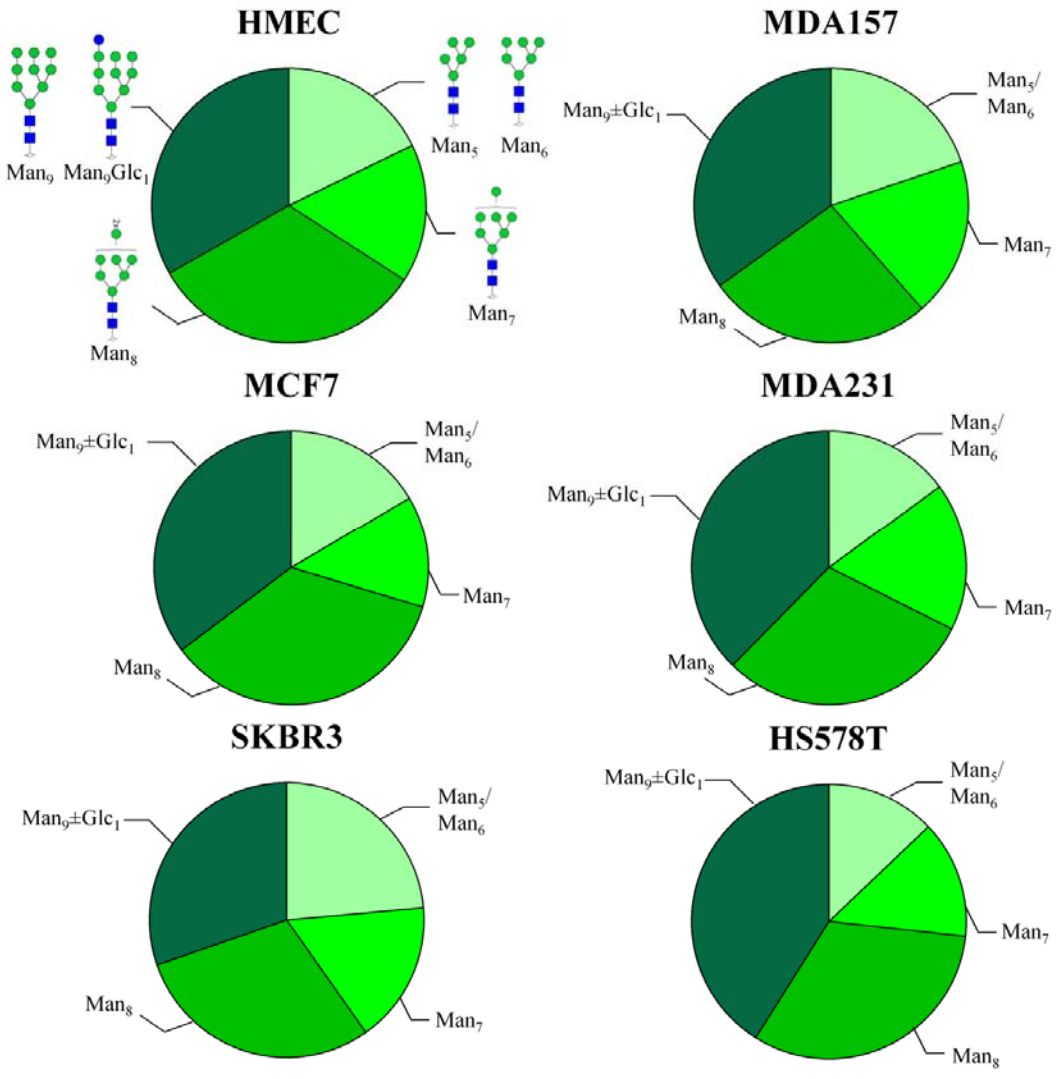


Figure 3.3 The distribution of high mannose *N*-glycans in the membrane protein fractions of the six investigated breast epithelial cells grouped into Man_5 and Man_6, Man_7, Man_8 and $Man_9 \pm Glc_1$, with the latter representing immature *N*-glycans normally only associated with intracellular ER *N*-glycosylation.

Despite the absence of visible differences in the MS1 mass profiles of the membrane *N*-glycomes of the cancer and non-cancer cell lines, in-depth quantitative analysis revealed notable tumor

phenotype-specific alterations. Similar to the secreted *N*-glycan analysis, the membrane glycan data were normalized to include only highly-processed *N*-glycan structures, i.e. hybrid and complex *N*-glycans for subsequent analyses. The membrane *N*-glycans in most of the breast cancer cell lines exhibited a higher degree of sialylation, fucosylation and branching as well as enhanced expression of Lewis types epitopes compared to those in HMEC. Notably, the expression of bisecting GlcNAc structures was restricted to MDA157 (Figure 3.2b-e).

Of the 43 hybrid and complex type membrane *N*-glycan structures, 23 structures were detected in both HMEC and at least one of the breast cancer cell lines. Comparative analysis of the sub-structures and determinants of this "common pool" of *N*-glycans in the normal and tumorigenic breast cells was performed. In total, 9 membrane *N*-glycans, five core fucosylated and four non-core fucosylated complex types *N*-glycans, were found to be significantly altered between cancer and non-cancer samples (Figure 3.4). The membrane proteins of the breast cancer cells displayed a significantly higher expression of α2,3-linked sialylated *N*-glycans (glycans 24c, 27c and 30c) while the α2,6-linked sialylated *N*-glycans (glycans 19a, 24b and 3a) were significantly under-represented. The nine regulated and biosynthetically related *N*-glycans were mapped to the *N*-glycosylation biosynthetic pathways to obtain a better overview of the *N*-glycan regulation (Figure 3.5).

Cluster analysis performed using the log transformed relative abundance data of the membrane *N*-glycomes did not result in a clear division between the non-tumorigenic and tumorigenic cell lines (Figure 3.6a). Similarly, no statistical distinction between HMEC and the breast cancer cell lines was observed by principal component analysis (Figure 3.6b). The non-tumorigenic cell line, HMEC was not visibly distinguished from the other tumorigenic cell lines but rather exhibited a relative close relationship to the three breast cancer cell lines of basal B subtype (MDA157, MDA231 and HS578T), which were observed to aggregate in the PCA analysis. Since no clear

tumor-specific clustering were observed between HMEC and the breast cancer cell lines, no further subtype-specific analysis was performed for membrane *N*-glycome profiles.

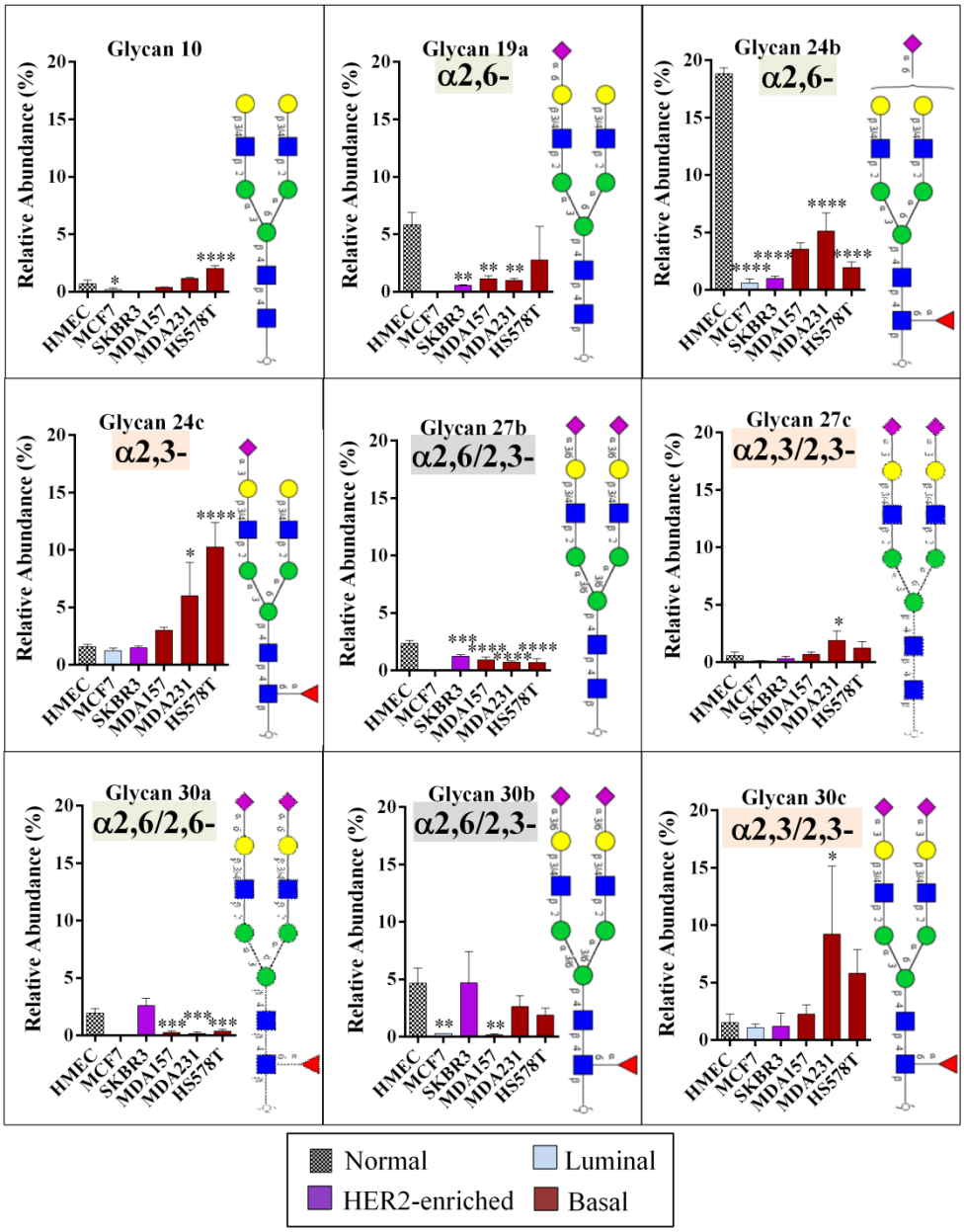


Figure 3.4 Nine differentially expressed *N*-glycans of membrane proteins derived from the normal breast epithelial cell line (HMEC) and the breast cancer cell lines. Statistical analyses were performed between breast cancer cell lines and HMEC. **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. Data are presented as mean ± SD (n = 3).

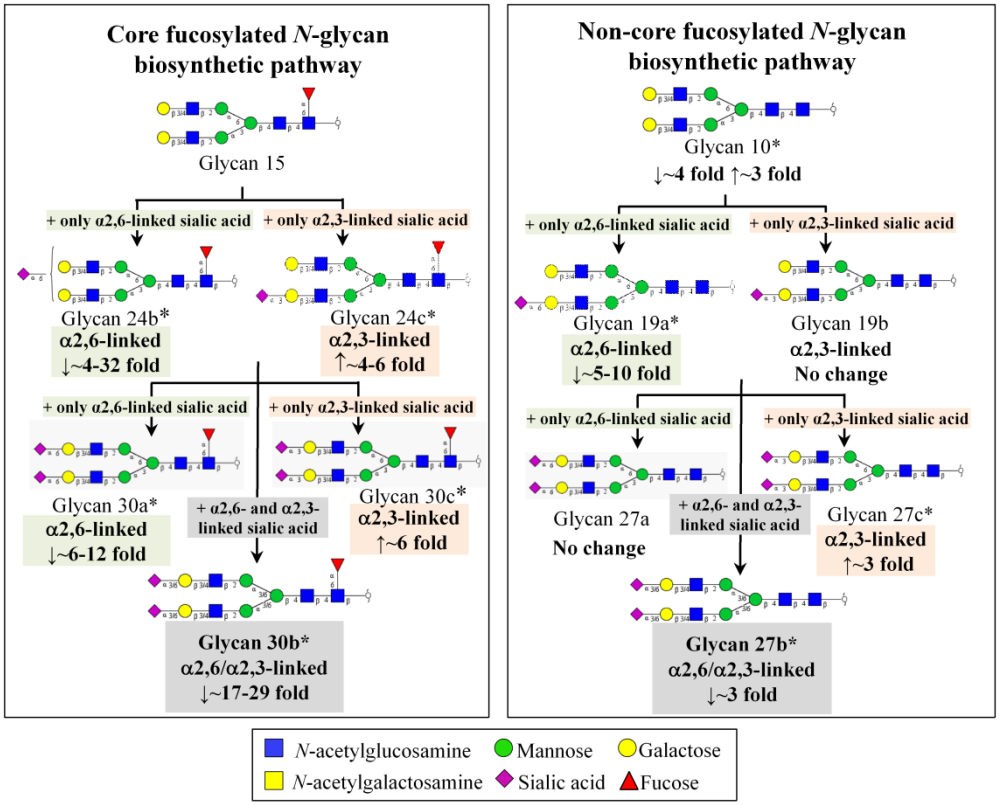


Figure 3.5 The nine regulated membrane *N*-glycans (five core fucosylated and four non-core fucosylayed, marked as *) are mapped according to their *N*-glycosylation biosynthetic pathways. Other *N*-glycans (unmarked) that were not significantly regulated in breast cancer were also depicted to complete this part of the pathway. Breast cancer cell lines, notably of the basal B subtype, preferentially expressed α2,3-linked sialylated *N*-glycans which was indicated by reduced expression of α2,6-sialylation (green box, glycans 19a, 24b, 30a) and increased expression of α2,3-sialylation (pink box, glycans 24c, 27c and 30c).

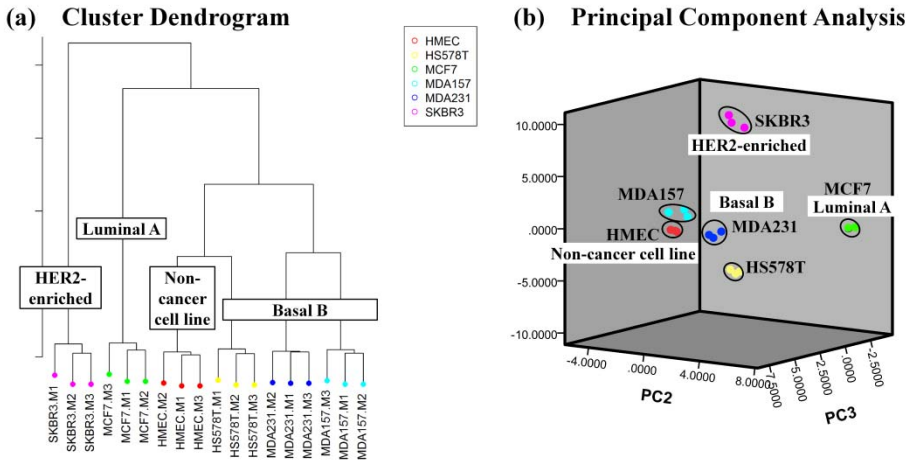


Figure 3.6 Dendrogram cluster analysis using hierarchical clustering (a) and 3D plot of PCA (b) of membrane *N*-glycosylation profiles of the six epithelial breast cells investigated in this study. PC1, principal component 1; PC2, principal component 2; PC3, principal component 3.

3.3.2 Global comparison between secreted and membrane *N*-glycans in the panel of breast epithelial cell lines

The *N*-glycan expression patterns of proteomes derived from the two subcellular fractions, i.e. secreted and membrane-bound proteins were remarkably distinct when evaluated on the individual *N*-glycan structural level yet shared some similar features. By categorizing the more processed *N*-glycans into the hybrid, complex and paucimannose types and the less processed *N*-glycans into the high mannose types, a striking difference in the distribution of *N*-glycan types were identified between the secreted and membrane *N*-glycome (Figure 3.7). *N*-glycans on the secreted glycoproteins were significantly more processed (76.2-95 % of total *N*-glycome) compared to those on the membrane-bound glycoproteins, which displayed significant amounts of high mannose type *N*-glycans. The differential expressions of *N*-glycosylation between the two subcellular fractions were further explored and the results are presented in the Chapter 5.

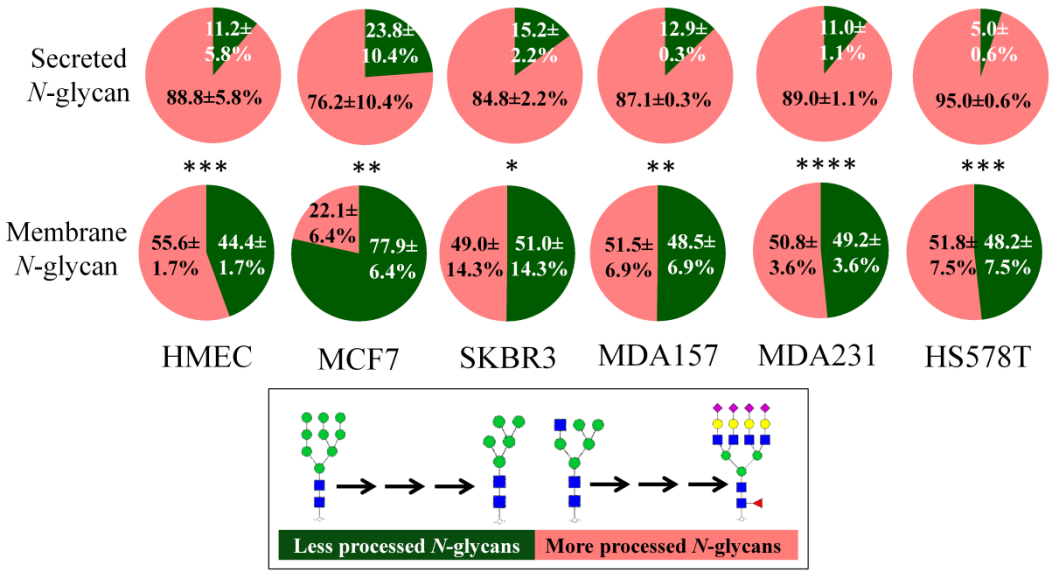


Figure 3.7 Secreted *N*-glycomes displayed more processed *N*-glycans compared to membrane N-glycomes. The relative abundances (mean ± SD) of the more processed *N*-glycans comprising the complex, hybid and paucimannose types are presented in light red and the less processed *N*-glcyans (i.e high mannose) are in green. The immature high mannose *N*-glycan is progressively trimmed and subsequently processed to produce hybrid, complex and paucimannose types *N*-glycans (inset).

Several *N*-glycans displayed LacdiNAc determinants were observed in the secretome of a subset of breast cell lines (HMEC, MCF7, MDA157) but were absent in the membrane fractions. One of these structures was identified as a hybrid-type carrying high mannose residues on arm-6 and LacdiNAc epitopes on arm-3 (Glycan no. 9, page 254). Although unusual, this structure was characterized based on existence of diagnostic fragments in the MS/MS spectra and need to be further validated.

The *N*-glycan type distribution of the *N*-glycan structures identified in the subcellular proteome of the individual cell lines are presented in Table 3.1. Structural characterization of the combined secreted and membrane *N*-glycan data yielded a total of 80 non-redundant *N*-glycans across all six cell lines, with a higher number of *N*-glycans detected uniquely in the secreted fractions (27 structures) (Figure 3.8). However, the largest fraction accounting for around 60% of the *N*-glycans (47 structures) were detected in both fractions.

Table 3.1 The *N*-glycan type distribution of *N*-glycan structures identified from the secreted and membrane proteome fractions from the six investigated breast cell lines.

Cell line	**Number of secreted *N*-Glycan structures according to glycan types**				
	High mannose	**Hybrid**	**Complex**	**Paucimannose**	**Total**
HMEC	6	4	18	0	28
MCF7	7	9	36	1	53
SKBR3	6	5	15	2	28
MDA157	6	11	26	3	46
MDA231	6	7	23	3	39
HS578T	5	2	16	0	23
Cell line	**Number of membrane *N*-Glycan structures according to glycan types**				
	High mannose	**Hybrid**	**Complex**	**Paucimannose**	**Total**
HMEC	7	7	17	2	33
MCF7	7	8	20	2	37
SKBR3	7	11	15	3	36
MDA157	7	11	23	3	44
MDA231	7	7	18	3	35
HS578T	7	1	12	2	22

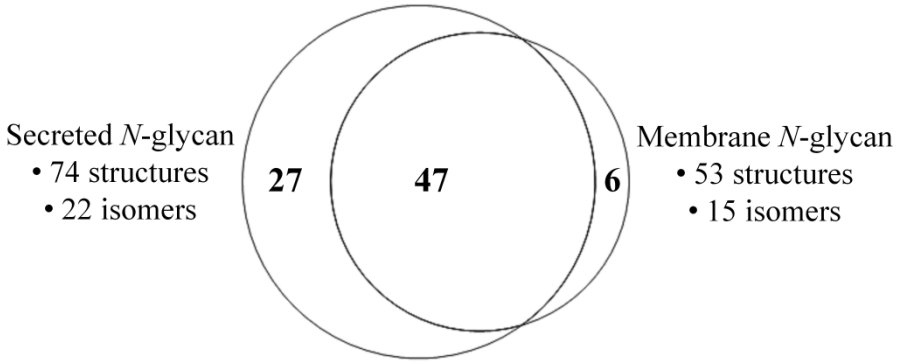


Figure 3.8 The common and unique *N*-glycans observed in the secreted and membrane proteomes of the six investigated breast cell lines

Despite the expression of subcellular-specific *N*-glycan type profiles, both of the subcellular fractions displayed similar tumor-associated trends, i.e. a high degree of sialylated, fucosylated and branched structures. In addition, common expression patterns in the subcellular *N*-glycomes within each cell line were identified. Amongst the tumorigenic cell lines, the secreted and membrane *N*-glycomes of MCF7 contained the highest levels of fucosylation (including core and Lewis type), and the lowest level of sialylation. Both subcellular *N*-glycomes of SKBR3 displayed the highest level of *N*-glycan sialylation while the bisecting GlcNAc structures were detected exclusively on the secreted and membrane *N*-glycans of MDA157.

3.4 Discussion

Aberrant expression of cell surface *N*-glycans on breast cancer cells is well-documented [201]. Several studies have demonstrated that certain tumor-associated *N*-glycans are associated with metastasis and poor prognosis of breast cancer patients [241, 246, 258]. In this study, *N*-glycome profiling and characterization were performed from isolated membrane proteins extracted from a panel of cultured breast epithelial cells including five from tumorigenic and one from non-tumorigenic origins. The results presented here are in general agreement with the *N*-glycome changes frequently reported to be associated with breast cancer including an increase in the sialylation and fucosylation as well as the β1,6-branching of *N*-glycans [240, 254, 257]. These molecular trends were observed in both the secreted and membrane *N*-glycomes across all

investigated cancer cell lines. Specifically, bisecting *N*-glycan structures were found uniquely in both secreted and membrane fractions of MDA157. In both cases, their occurrence was consistent with reduced expression of tri- and tetra-antennary structures as increased GnT3 activity (bisecting GlcNAcylation) has been reported to inhibit GnT5 extension (tri- and tetra-antenna) [342]. Lectin cytochemistry of this panel of breast cells would represent a complementary technique to validate the MDA157-specific bisecting GlcNAcylation using lectins recognizing such glyco-determinants i.e. PHA-E. In breast cancer, bisecting structures were demonstrated to inhibit growth factor signalling, slow tumor progression, reduce cell adhesion and migration [342, 343]. It is interesting to note that although MDA157 belongs to the triple negative breast cancer subtype, it is derived from a rare subtype known as medullary carcinoma with low-grade aggressiveness. Patients diagnosed with this subtype have better prognosis compared with other invasive ductal tumors [344]. More research is thus needed to understand the role of bisecting *N*-glycans in breast cancer metastasis.

Another notable observation was the presence of a significant amount of high mannose *N*-glycans in the membrane proteome. The under processed *N*-glycan structures carrying eight or nine mannose residues were dominant features of the membrane *N*-glycomes across all breast epithelial cell types suggesting that this phenomenon may not be associated with breast tumorigenesis. Nevertheless, a previous *N*-glycome study investigating the cell membrane component of a number of cancer cell lines reported a significantly higher expression of high mannose type *N*-glycans although they did not report on their internal distribution [291]. Elevated levels of Man_9 structures have been observed in both human and mouse breast cancer sera and correlated well with the cancer progression [293]. The significance of high mannose in breast cancer remains unclear and needs to be further explored.

In Chapter 2, proteomic analysis of the membrane fraction revealed that in addition to the plasma membrane proteins, a considerable proportion of the membrane proteins were endoplasmic reticulum- and Golgi-residing membrane proteins suggesting that majority of the membrane *N*-glycoproteins may be derived from an intracellular organelle origin. In contrast, the *N*-glycan profiling of the secreted proteome captured a relatively accurate extracellular *N*-glycome landscape. One of the most interesting observations in these datasets were the unique differential expression patterns of the membrane and secreted *N*-glycomes, which was consistently observed in the investigated cell line. This observation is intriguing given the fact that both subcellular proteome fractions share a common *N*-glycosylation biosynthetic pathway. In Chapter 5, this observation was further explored to understand the mechanisms involved in the generation of subcellular-specific *N*-glycosylation.

3.5 Conclusion

This chapter has profiled the membrane *N*-glycosylation of a panel of breast cancer cell lines representing various breast cancer subtypes. Together, with the secreted *N*-glycome (Part 1), this data provide an in-depth investigation of the altered protein *N*-glycosylation associated with breast malignancy. Significant tumor- and subtype-specific changes were identified in the membrane *N*-glycomes corroborating similar trends observed in the secreted *N*-glycomes. Thus, this consolidated knowledge brings a more complete picture of the molecular alterations relating to the disease, but may also facilitate the discovery of promising glycan biomarkers and potential drug targets to improve the therapeutic treatment breast cancer.

CHAPTER 4

Publication III - Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants

In Chapter 3, N-glycome profiling of the secreted and membrane protein fractions extracted from a panel of cultured breast cancer cells revealed consistent differential distribution of the N-glycan types between the two fractions. This chapter performs a systematic investigation to understand the mechanism involved in the generation of this subcellular-specific N-glycosylation feature, using a combination of structural knowledge, computational and analytical tools.

frontiers in IMMUNOLOGY

ORIGINAL RESEARCH ARTICLE
published: 25 August 2014
doi: 10.3389/fimmu.2014.00404

Differential site accessibility mechanistically explains subcellular-specific *N*-glycosylation determinants

Ling Yen Lee, Chi-Hung Lin, Susan Fanayan, Nicolle H. Packer and Morten Thaysen-Andersen*

Department of Chemistry and Biomolecular Sciences, Biomolecular Frontiers Research Centre, Macquarie University, Sydney, NSW, Australia

Edited by:
Elizabeth Yuriev, Monash University, Australia

Reviewed by:
Jamie Heimburg-Molinaro, Emory University, USA
Tony Velkov, Monash University, Australia

Correspondence:
Morten Thaysen-Andersen, Department of Chemistry and Biomolecular Sciences, Biomolecular Frontiers Research Centre, Macquarie University, Sydney, NSW 2109, Australia
e-mail: morten.andersen@mq.edu.au

Glycoproteins perform extra- and intracellular functions in innate and adaptive immunity by lectin-based interactions to exposed glyco-determinants. Herein, we document and mechanistically explain the formation of subcellular-specific *N*-glycosylation determinants on glycoproteins trafficking through the shared biosynthetic machinery of human cells. LC-MS/MS-based quantitative glycomics showed that the secreted glycoproteins of eight human breast epithelial cells displaying diverse geno- and phenotypes consistently displayed more processed, primarily complex type, *N*-glycans than the high-mannose-rich microsomal glycoproteins. Detailed subcellular glycome profiling of proteins derived from three breast cell lines (MCF7/MDA468/MCF10A) demonstrated that secreted glycoproteins displayed significantly more α-sialylation and α1,6-fucosylation, but less α-mannosylation, than both the intermediately glycan-processed cell-surface glycoproteomes and the under-processed microsomal glycoproteomes. Subcellular proteomics and gene ontology revealed substantial presence of endoplasmic reticulum resident glycoproteins in the microsomes and confirmed significant enrichment of secreted and cell-surface glycoproteins in the respective subcellular fractions. The solvent accessibility of the glycosylation sites on maturely folded proteins of the 100 most abundant putative *N*-glycoproteins observed uniquely in the three subcellular glycoproteomes correlated with the glycan type processing thereby mechanistically explaining the formation of subcellular-specific *N*-glycosylation. In conclusion, human cells have developed mechanisms to simultaneously and reproducibly generate subcellular-specific *N*-glycosylation using a shared biosynthetic machinery. This aspect of protein-specific glycosylation is important for structural and functional glycobiology and discussed here in the context of the spatio-temporal interaction of glyco-determinants with lectins central to infection and immunity.

Keywords: *N*-glycosylation, solvent accessibility, *N*-glycome, subcellular location, glycoproteome, glycosylation site, *N*-glycan, glycoprotein

INTRODUCTION

Significant parts of the human genome and cellular energy are dedicated to produce and regulate protein glycosylation (1). Hence, it is no surprise that this abundant post-translational modification is important in a wide spectrum of biological processes to maintain cellular homeostasis (2). Dysregulation of protein glycosylation is a cause and/or effect of numerous pathological conditions including, but not limited to, congenital disorder of glycosylation (3), cystic fibrosis (4), inflammation (5), auto-immunity (6), and cancer (7). The extracellular location of secreted and cell-surface-tethered proteins carrying *N*-linked glycosylation is ideal for facilitating molecular interactions with the surrounding environment (8). Intracellular functions of *N*-glycoproteins are also known (9, 10). The terminal determinants of host *N*-glycans (so-called "self" and "altered self" in disease) are recognized by endogenous and exogenous glycan-binding proteins commonly called lectins. Interactions between lectins and *N*-glycans are central in innate and adaptive immunity (11). Important examples include the C-type lectins, which may be crudely divided into lectins having affinity for α-mannose/α-fucose-terminated *N*-glycans including dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN), macrophage mannose receptors and Langerin (12), and lectins having affinity for galactose/GalNAc terminating glycans such as macrophage galactose lectin and DC-asialoglycoprotein receptor (13, 14). In addition, siglecs (I-type lectins) and galectins (S-type lectins) are important for facilitating a functional immune response (15).

The human *N*-glycosylation biosynthetic machinery is relatively well understood (16, 17). In brief, the synthesis is initiated by the transfer of common immature glycan precursors i.e., $Glc_3Man_9GlcNAc_2$ to conserved sequons (NxT/S, x ≠ P) on translocating polypeptide chains. The glycan precursor is then remodeled through sequential trimming and elongation by specific glycosidases and glycosyltransferases located in the endoplasmic reticulum (ER) and the *cis*-, medial, and *trans*-Golgi, respectively. This series of enzymatic processes first results in the trafficking *N*-glycoproteins being comprised of attached high-mannose-type *N*-glycans, which progresses to the hybrid- and complex-type stage if sufficient interactions with the processing enzymes occur (17). The Golgi-based *N*-glycan processing, including the formation of glycan types and the addition of terminal determinants such as α-fucosylation and α-sialylation, occurs on maturely folded

glycoproteins (18, 19). An extensive and reproducible repertoire of *N*-glycans is usually present on a given glycosylation site (20). This *N*-glycan microheterogeneity on proteins results from incomplete processing by the multiple competing enzymatic reactions that can be influenced by cellular factors including the availability of nucleotide sugars, glycosylation enzyme activity, and glycoprotein trafficking time through the biosynthetic machinery. Such cellular factors contribute to cell- and tissue-specific *N*-glycosylation (21). Importantly, the structures of the individual glycoproteins trafficking through the glycosylation machinery dramatically influence the degree of *N*-glycan processing creating protein- and site-specific *N*-glycosylation (22). By thorough literature-based curation of published site-specific glycoprofiling data of mammalian *N*-glycoproteins, we recently confirmed that several structural features including glycan type formation, α1,6-(core) fucosylation, and β1,4/6-GlcNAc branching of *N*-glycans are strongly correlated with the solvent accessibility of the glycosylation sites of maturely folded glycoproteins (19). As such, extensive *N*-glycan processing was observed for proteins displaying solvent accessible glycosylation sites relative to spatially hidden sites. Thus, differential site accessibility can explain how glycoproteins produced simultaneously in the same cell, and even sequons on the same glycoproteins, can present widely different *N*-glycan structural repertoires.

Considering the importance of protein- and site-specific *N*-glycosylation in many aspects of glycobiology including glycoimmunology, we here seek to further explore this feature in the context of the multiple subcellular glycoproteomes that traffic through the shared glycosylation machinery in the secretory pathway of human cells, yet end up at different cellular locations. Due to the functional implications of both intra- and extracellular *N*-glycoproteins, we focus on the secreted, cell-surface, and intracellular glycoproteomes, the latter fraction largely represented by microsomal proteins (23). Understanding, how the subcellular glycoproteomes are generated and regulated under normal and altered physiological conditions of the cell is valuable to the understanding of their involvement in immune biology. Recent analytical developments in glycomics (24–27) and glycoproteomics (28–31) have, together with more conventional proteomics, enabled sensitive, and detailed system-wide investigations of the regulation of protein *N*-glycosylation in immunity (32).

Using LC-MS/MS-based glycomics and proteomics on multiple subcellular fractions from a panel of human cell lines displaying diverse cellular characteristics, we here document that human cells have developed a general mechanism to reproducibly generate vastly different *N*-glycan determinants on their differently located subcellular glycoproteomes that trafficked simultaneously through a shared biosynthetic machinery. We provide evidence that the subcellular-specific protein *N*-glycosylation arises from differential solvent accessibilities of the glycosylation sites of maturely folded glycoproteins that localize to different subcellular compartments following the glycan processing. This aspect of protein-specific glycosylation is discussed here in the context of immunity and infection due to the crucial role of endogenous and exogeneous lectins recognizing exposed self, and altered self, glyco-determinants to facilitate the functional immune response.

MATERIALS AND METHODS

CELLULAR ORIGIN, CULTURE CONDITIONS, AND DOUBLING TIME

Multiple human cells showing diverse geno- and phenotypical characteristics were used to demonstrate the general nature of the cellular mechanisms observed in this study. Human mammary epithelial cells (HMEC) were purchased (product # CC-2551, Lonza). Human breast epithelial cell lines MCF10A, MCF7, SKBR3, MDA-MB-157 (MDA157), MDA-MB-231 (MDA231), and HS578T as well as a human colon cancer epithelial cell line SW480 were obtained from American Type Culture Collection (Manassas, VA, USA). HMEC was grown in HuMEC Ready Media (Invitrogen). MCF10A was cultured in DMEM/F12 with the addition of 5% horse serum (Invitrogen), 20 ng/mL epidermal growth factor (EGF) (Invitrogen), 0.5 μg/mL hydrocortisone (Sigma), 100 ng/mL cholera toxin (Sigma), and 8 μg/mL insulin (Invitrogen). Other cell lines were grown in RPMI (Sigma) supplemented in 5% fetal bovine serum (FBS) (Invitrogen), 10 mM glutamine (Invitrogen), and 10 μg/mL insulin. Cells were maintained at 37°C in 5% CO_2 for all experiments. The breast cell lines were grown in triplicates to ~80% confluence and washed at least four times with ice-cold phosphate buffered saline (PBS) to remove traces of FBS and incubated in serum-free media at 37°C in 5% CO_2 for 48 h prior to subcellular fractionation.

To measure the cellular doubling times of the breast cell lines, cells were seeded at 1.3×10^4 cells/mL/well in six-well plates and incubated overnight at 37°C in 5% CO_2. Cells were counted every 24 h over a four-day period using a cell counter (Bio-Rad). The doubling time for each cell line was determined from their exponential growth phase. For overview of the investigated cells and associated data, see Table S1 in Supplementary Material.

COLLECTION AND PREPARATION OF SUBCELLULAR GLYCOPROTEOMES FROM BREAST CELL LINES

The *secreted* subcellular glycoproteomes were collected by sampling 30 mL of serum-free culture media followed by centrifugation at 2,000 × *g* to pellet any floating cells. The supernatants were concentrated and buffer exchanged into PBS (1×) using 10,000 MWCO Amicon Ultra membranes (Millipore). Proteins were then precipitated with nine volumes of acetone overnight at −20°C. The pellets were stored at −80°C until further analysis.

The *cell-surface* subcellular glycoproteomes were isolated from MCF7, MDA468, and MCF10A breast epithelial cell lines using a commercial biotinylation kit (product # 89881, Pierce) to specifically biotinylate the cell-surface glycoproteins. The protocol supplied by the manufacturer was followed. Briefly, monolayers of cultured cells grown in 75 cm^2 culture flasks were washed three-times with PBS (1×) before incubation in EZ-Link sulfo-NHS-SS-biotin in ice-cold PBS (1×) for 30 min at 4°C on a rocking platform. The labeling reactions were terminated and the biotinylated cells were washed and collected by scraping in Tris-buffered saline (TBS) (1×), followed by centrifugation at 500 × *g* for 3 min. The supernatants were discarded and the cell pellets were disrupted in manufacturer-provided lysis buffer by ultra-sonication using five 1-s bursts with a Sonifier 450 (Branson Sonifier, Wilmington, NC, USA). The cell lysates were centrifuged at 10,000 × *g* for 2 min at 4°C. Solubilized biotinylated cell-surface proteins in the clarified supernatants were isolated using NeutrAvidin Agarose.

Cell-surface-bound proteins were eluted using 50 mM DTT and precipitated with acetone overnight at −20°C. The pellets were stored at −80°C until analysis.

The *microsome* (total membrane) subcellular glycoproteomes were obtained by first removing serum-free media, thoroughly washing cells with PBS (1×), and harvesting cells in 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA containing a protease inhibitor cocktail (Roche Diagnostics). The cells were ultrasonicated on ice for three rounds of 10-s bursts using a Sonifier 450 and centrifuged at 2,000 × *g* for 20 min at 4°C to remove intact cells and nuclei. The supernatants were ultra-centrifuged at 120,000 × *g* for 80 min after which the supernatants were discarded. The microsomal membrane pellets were washed twice with ice-cold 0.1 M sodium carbonate and resuspended in 25 mM Tris-HCl pH 7.4, 150 mM NaCl, and 1% (v/v) Triton X-114. Samples were phase partitioned by incubation at 37°C for 20 min, followed by 1,000 × *g* centrifugation for 10 min. The upper aqueous layer was carefully removed and nine volumes of ice-cold acetone were added to the lower detergent phase and incubated overnight at −20°C to precipitate the proteins. The pellets were stored at −80°C until further analysis.

The protein concentrations of the subcellular fractions were measured using Bradford reagents (Sigma). Equal protein amounts were precipitated in the three subcellular fractions and the resulting pellets were solubilized in 8 M urea for spotting on PVDF membranes for *N*-glycome profiling or in NuPAGE LDS sample buffer for gel electrophoresis prior to proteome profiling.

SUBCELLULAR FRACTIONATION OF HUMAN COLON CANCER CELL LINES

SW480 cells (5×10^7) were washed twice with homogenization buffer (20 mM HEPES, pH 7.5, and 0.25 M sucrose). Cell pellets were resuspended to a final volume of 2 mL in homogenization buffer and lysed using an Ultra-Turrax disperser (Ika). After a low speed centrifugation at 1,000 × *g* for 10 min, the supernatant was collected as the post-nuclear fraction (PNF). The PNF was subjected to ultracentrifugation at 30,000 rpm for 1 h in a SW41Ti rotor (Beckman Coulter) to pellet the microsome. ER and Golgi-enriched membranes were prepared as described (33). Briefly, 1 mL of PNF (usually 2.5–3 mg protein) was adjusted to 1.4 M sucrose by adding 2 mL of 2 M sucrose. A discontinuous sucrose gradient was made by sequentially loading 1.5 mL of 1.6 M sucrose, 3 mL PNF in 1.4 M sucrose, 3 mL of 1.2 M sucrose, and 3 mL of 0.8 M sucrose. All sucrose solutions contained 20 mM HEPES pH 7.5. Ultracentrifugation was conducted at 28,500 rpm for 2 h in a SW41Ti rotor. Enriched-Golgi membranes were harvested at the 0.8 M/1.2 M interface. Enriched ER membranes were harvested from the 1.4 M layer. The collected ER and Golgi membranes were diluted by homogenization buffer to reduce concentration of sucrose and subsequently pelleted by ultracentrifugation at 30,000 rpm for 1 h in a SW41Ti rotor. Pelleted ER- and Golgi-enriched membranes were resuspended in 8 M urea and protein concentrations were determined by BCA assays (Pierce).

RELEASE AND PREPARATION OF *N*-GLYCANS FROM SUBCELLULAR GLYCOPROTEOMES

N-glycans were released from ~20 μg secreted proteins, 50 μg cell-surface proteins, and 50 μg microsome membrane proteins as previously described (27). Briefly, protein mixtures were immobilized on methanol-activated PVDF membranes (Millipore) and allowed to dry overnight. Membrane-bound proteins were incubated with 2.5 U PNGase F (*Flavobacterium meningosepticum*, Roche) for 16 h at 37°C to ensure complete release of *N*-glycans. Released *N*-glycans were incubated with 100 mM ammonium acetate (pH 5) for 1 h at RT and subsequently dried by vacuum centrifugation. Reduction of *N*-glycans was performed with 20 μL 1 M sodium borohydride (Sigma) in 50 mM potassium hydroxide (Sigma) for 3 h at 50°C. Reactions were quenched with 2 μL glacial acetic acid. Dual desalting was performed in micro-SPE formats using strong cation exchange/C_{18} and carbon columns (27). Desalted *N*-glycans were eluted from the carbon columns with 20 μL 40% acetonitrile (ACN) containing 0.1% (v/v) trifluoroacetic acid and dried by vacuum centrifugation (34). Samples were stored at −80°C if not analyzed immediately.

DIGESTION AND PREPARATION OF PEPTIDE MIXTURES FROM SUBCELLULAR GLYCOPROTEOMES

The subcellular glycoproteomes of the breast cells (~50 μg protein/fraction) i.e., secreted, cell surface, and microsomes and of colon cells (~10 μg protein/fraction) i.e., microsome and ER- and Golgi-enriched membrane fractions were reduced and alkylated and subsequently in-gel (breast cells) or in-solution (colon cells) digested. Prior to in-gel digestion, samples were loaded in 10 μL NuPAGE LDS buffer and separated on 4–12% Bis-Tris PAGE gels (Invitrogen). Electrophoresis was performed at 200 V for 50 min. After separation of proteins, gels were fixed in 40% (v/v) ethanol and 10% (v/v) acetic acid for at least 2 h, stained overnight with Coomassie Blue G250 (Bio-Rad) and destained in ultra-pure water (Millipore). In-gel trypsin digestion of all samples was performed from eight equal sized gel fractions. Each fraction was sliced into 1 mm pieces and placed in a 96-well plate. The gel pieces were destained with 50% (v/v) ACN in 50 mM ammonium bicarbonate until clear, dehydrated in 100% (v/v) ACN, and dried. Sequence-grade porcine trypsin (Promega) (1:30 enzyme/substrate, w/w) was used to digest the proteins overnight at 37°C. Tryptic peptide mixtures were then collected and two rounds of gel extractions of peptides were performed with 2% (v/v) formic acid in 50% (v/v) ACN and 50 mM ammonium bicarbonate. The extracts were combined, peptide mixtures dried by vacuum centrifugation, redissolved in 10 μL 0.1% (v/v) formic acid, and desalted as described below. For in-solution digestion, samples were diluted to <1 M urea (final concentration) and trypsinized (sequence-grade porcine trypsin, 1:40 enzyme/substrate, w/w) overnight at 37°C. Following proteolysis, the peptide mixtures were acidified by adding formic acid to a final concentration of 0.1% (v/v). Desalted of peptide mixtures were performed using self-packed C_{18} SPE tips. Briefly, C_{18} tips were washed three-time with 20 μL 100% ACN, three-times with 20 μL 50% (v/v) ACN in 0.1% formic acid, and equilibrated with 50 μL 0.1% (v/v) formic acid. After sample loading, tips were washed three-times with 20 μL 0.1% formic acid. Peptides were eluted with 20 μL 60% (v/v) ACN in 0.1% formic acid and 20 μL 90% (v/v) ACN in 0.1% formic acid and dried. The desalted fractions were dried and stored at −80°C until LC-MS/MS.

LC-MS/MS-BASED *N*-GLYCOMICS

N-glycans alditols were separated using a porous graphitized carbon (PGC) LC column [5 μm (particle size) Hypercarb KAPPA, 100 mm (length) × 200 μm (ID), 250 Å (pore size), Thermo Scientific] using an Ultimate 3000 HPLC system (Dionex) connected directly to an ESI-MS/MS HCT Ultra ion trap (Bruker Daltonics). Separation was performed using a binary gradient solvent system made up of solvent A (aqueous 10 mM NH_4HCO_3) and solvent B (90% ACN/10 mM ammonium bicarbonate). The flow rate was 2 μL/min and a total gradient of 100 min was programed as follows: 0–2.5% solvent B for 0–13 min; 2.5–17.5% solvent B for 14–48 min; 17.5–50% solvent B for 48–65 min; 50–100% solvent B for 65–75 min; 100% solvent B for 75–80 min; back to 0% solvent B for 80–85 min, and 100% solvent A equilibration for 15 min. Settings for the MS/MS were as follows: drying gas flow: 6 L/min; drying gas temperature: 300°C; nebulizer gas: 12 p.s.i.; skimmer: −40.0 V; trap drive: −99.1 V; and capillary exit: −166 V. Smart fragmentation was used with start- and end-amplitude of 30 and 200%, respectively. Ions were detected in ion charge control set at 100,000 ions/scan and with maximum accumulation time of 200 ms. MS spectra were obtained in negative ion mode with three scan events: a full scan (*m/z* 400–2,200) at a scan speed of 8,100 *m/z*/s and data-dependent MS/MS scans after CID fragmentation of the top two most intense precursor ions with an absolute intensity threshold of 30,000 and a relative intensity threshold of 5% relative to the base peak. Dynamic inclusion was inactivated to ensure MS/MS generation of closely eluting *N*-glycan isomers. Precursors were observed mainly in charge states $Z = -1$ and/or −2. Mass accuracy calibration of the mass spectrometer was performed using a well-defined tune mix (Agilent) prior to acquisition. *N*-glycans released from bovine fetuin served as positive controls for the sample preparation and the LC-MS/MS performance. Differences between observed and theoretical precursor and fragment masses were generally <0.2 Da. Three LC-MS/MS technical replicates were performed for the subcellular fractions.

LC-MS/MS-BASED PROTEOMICS

Three LC-MS/MS technical replicates of the subcellular proteomes of the breast cells were analyzed using a Q-Exactive (Thermo Scientific). Peptide mixtures in 0.1% (v/v) formic acid were loaded onto a C_{18} reversed phase column packed in-house [2.7 μm (particle size) HaloLink Resins, Promega, column dimensions: 100 mm (length) × 75 μm (ID)]. Separation of peptides was performed over a 60 min gradient with the first 50 min of the linear gradient increasing from 0 to 50% in solvent B [0.1% (v/v) aqueous formic acid/100% (v/v) ACN] and then to 85% solvent B for the next 2 min and maintained at 85% for 8 min. The flow rate was constant at 300 nL/min. The Easy-nLC (Thermo Scientific) was connected directly to the nano-ESI source of the Q-Exactive. MS full scans were acquired with resolution of 35,000 in the positive ion mode over *m/z* 350–2,000 range and an automatic gain control (AGC) target value of 1×10^6. The top 10 most intense precursor ions were then isolated for MS/MS using higher energy collisional dissociation fragmentation at 17,500 resolution with the following settings: collision energy: 30%; AGC target: 2×10^5; isolation window: *m/z* 3.0; and dynamic exclusion enabled. Precursors with unassigned or $Z = +1$ charge states were ignored for MS/MS selection.

The subcellular proteomes of the colon cells were LC-MS/MS analyzed using a Triple TOF 5600 (ABSciex). Peptides were separated by a nanoLC system (Eksigent) on a C_{18} reversed phase column [ProteCol 100 mm (length) × 150 μm, (ID): 3 μm (particle size), 300 Å (pore size); SGE Analytical Science] with a 90 min gradient from 5 to 40% solvent B [90% (v/v) ACN with 0.1% formic acid] at a constant flow rate of 600 nL/min. The top 10 most intense precursor ions with $Z = +2$, +3, and +4 were selected for MS/MS using CID fragmentation.

ANALYSIS OF *N*-GLYCOME LC-MS/MS DATA

N-glycome raw data for all subcellular glycoproteomes were viewed and manually analyzed using DataAnalysis v4.0 (Bruker Daltonics). Monoisotopic masses were obtained and searched against GlycoMod[1] to obtain possible monosaccharide compositions, which were subsequently verified manually by *de novo* sequencing of corresponding MS/MS spectra and by taking account of PGC chromatographic retention time. The glycan type and the terminating monosaccharide determinants could unambiguously be identified using this method (27). The relative abundances of the observed *N*-glycans were determined using the ratio of the extracted ion chromatogram (EIC) peak area of each *N*-glycan species over the sum of EIC peak areas of all observed *N*-glycans in the sample. This has been shown to be a reasonably accurate method for relative *N*-glycan quantitation (35). The extent of *N*-glycan processing was measured by evaluating the relative molar proportion of the relative unprocessed species (i.e., immature mono-glucosylated glycans and high-mannose type *N*-glycans) and the processed species (i.e., hybrid, complex, and paucimannose type *N*-glycans) of the total *N*-glycome. In addition, the degree of monosaccharide determinants including α1,2/3/6-mannose, β1,3/4-galactose, α1,3/4/6-fucose, and α2,3/6-sialic acid terminating *N*-glycans were calculated as a relative molar abundance of both the entire *N*-glycome and of the potentially modified *N*-glycan substrates (e.g., complex/hybrid-types). Since multiple determinants may be displayed by a given *N*-glycan, the total summed to more than 100%.

ANALYSIS OF LC-MS/MS-BASED PROTEOMIC DATA AND GENE ONTOLOGY

For breast cell proteomes, raw spectra were converted to .mgf files using Proteome Discoverer Daemon v1.3 (Thermo Scientific) and searched against SwissProt protein database (*Homo sapiens*, 20,279 reviewed entries, November 2013 release) using the Global Proteome Machine (Cyclone). The following search criteria were used: carbamidomethylation was a fixed modification and oxidation and deamidation were variable modifications for methionine and asparagine/glutamine residues, respectively. Mass tolerances of 10 ppm and 0.02 Da were selected for precursor and product ions, respectively, with a maximum of two missed tryptic cleavages.

For colon cell proteomes, MS/MS spectra were extracted by ProteinPilot v4.2 (ABSciex) and searched using Mascot v2.4.0

[1] http://web.expasy.org/glycomod/

(Matrix Science) against SwissProt protein database (*Homo sapiens*, 20,253 entries, April 2013 release) using trypsin as the digestion enzyme. Precursor and product ion tolerances were 20 ppm and 0.50 Da, respectively. Oxidation of methionine residues and carbamidomethylation of cysteine residues were used as variable modifications.

Scaffold v4.2.1 (Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptides were accepted if they were confidently identified at ≥95.0% probability as evaluated by the local false discovery rate (FDR) algorithm. Proteins were included if they were confidently identified at ≥99.0% probability as assigned by the Protein Prophet algorithm incorporated in the software. Proteins containing shared or similar peptides, and which could not be differentiated based on MS/MS analysis alone, were grouped to satisfy the principles of parsimony. Proteins, which confidently shared identified peptides were grouped into clusters. Proteins were annotated using gene ontology (GO) terms from NCBI. The protein identifications were stringently filtered based on the presence of a minimum of two peptides in all replicates. The relative abundances of proteins were determined by conventional spectral counting and adjusted by taking the polypeptide length into account. Putative *N*-glycoproteins in the proteome of the subcellular fractions were predicted *in silico* based on the presence of one or more sequons (NxT/S, x ≠ P) and a signal peptides (for secreted proteins) and/or transmembrane regions (for cell-surface and microsome proteins) using prediction tools including SignalIP (v4.1) (36), Transmembrane Hidden Markov Model (TMHMM v2.0) (37), PrediSi (38), and Phobius (39). Mitochondrial and nuclear membrane proteins were excluded as these are unlikely to enter the ER–Golgi glycosylation pathway. Ambiguous assignments were manually checked (validated or discarded) with information from Uniprot. Potential sequons were obtained using NetNGlyc (40). These *in silico* prediction tools generated lists of experimentally validated and putative glycoproteins. The 100 most abundant glycoproteins in each subcellular fraction were used to assess glycosylation site accessibility. The contribution of these glycoproteins to the total glycoproteome in each sample was estimated by multiplying the normalized spectral count of the individual glycoproteins with their potential glycosylation sites, a measure termed "sequon-weighted normalized spectral count."

SELECTION OF PDB 3D STRUCTURE FOR GLYCOSYLATION SITE ACCESSIBILITY DETERMINATION

Three-dimensional protein structures were obtained from the protein data bank (PDB) database[2]. If multiple structures were available for a glycoprotein, the best match to the naturally occurring variant was chosen by considering the following parameters in a prioritized order: (1) high protein sequence coverage and resolution of the 3D structure, (2) source of protein (purified from organism/tissue over artificial expression system), (3) known site-specific mutations, (4) presence of artificial/natural ligands, and (5) oligomerization of the solved 3D structure. The experimentally obtained PDB structures used in this study were all based on X-ray crystallography, Table S2 in Supplementary Material. Where no experimentally determined structures were available (43%), structure homologs were obtained from ProteinModelPortal[3], Swiss-model repository[4], or ModBase[5]. High sequence homology was used as a selection criterion when choosing homology model. The average sequence homology for all structures was 67%, which is considered very reliable for homology modeling (41), Table S1 in Supplementary Material. 3D protein structures were viewed with RasMol v2.7.5 (RasWin Molecular Graphics) for visual inspection.

GLYCOSYLATION SITE ACCESSIBILITY DETERMINATION FROM MATURELY FOLDED GLYCOPROTEINS

The glycosylation site solvent accessibility was determined by measuring the accessibility to the individual asparagine residues forming the glycosylation sites using NACCESS[6] (42), an accurate and frequently used solvent accessibility determination program (19, 43–45). NACCESS calculates the atomic accessible area by predicting van der Waal's interactions when a probe is rolled around on the protein surface (46, 47). The maximum probe size offered by the program (5 Å radius) was used as a default in this study to simulate as closely as possible the accessibility of the glycosylation enzymes to the glycosylation sites. NACCESS produces unit-less and absolute accessibility values as the output format (denoted "arb. units"), which are comparable between glycosylation sites of different glycoproteins (19). Prior to the measurements of site accessibility, any water molecules, sugars, ligands, and other hetero-atoms/molecules, not part of the core polypeptide chain, were removed from the protein surface. Negligible accessibility differences were observed for the "native" and the monomeric form of glycoproteins with quaternary structures (data not shown). Hence, in the case of multimers, glycosylation site solvent accessibilities derived from the monomeric structures were not considered in the analysis.

STATISTICAL ANALYSIS

All relative abundances of *N*-glycans were presented as a percentage out of 100% as mean ± SD. Glycosylation site accessibilities were presented as mean ± SEM to illustrate the potential spread of mean instead of the individual data points, which can be hugely influenced by the (local) accuracy and quality of the PDB structures. To overcome this potential issue of PDB "noise," relative large numbers of data points (n) were needed. Data were analyzed using Prism v6 (GraphPad). One-way ANOVA analysis was performed for statistical comparison between the three subcellular fractions followed by *post hoc* Tukey's tests. All p values were adjusted taking into account the multiple comparisons made and reported as multiplicity adjusted p values. $p < 0.05$ was regarded as statistically significant and indicated with "*." Stronger statistical significance was indicated as follows: $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$. Simple linear regression and corresponding correlation coefficients (R^2) were obtained to evaluate the relationship between

[2] http://www.rcsb.org/pdb

[3] http://www.proteinmodelportal.org

[4] http://swissmodel.expasy.org

[5] http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi

[6] http://wolf.bms.umist.ac.uk/naccess/

the degree of *N*-glycan processing in terms of glycan type and expression of terminal glycan determinants and the glycosylation site solvent accessibility.

RESULTS

SUBCELLULAR-SPECIFIC *N*-GLYCOSYLATION OF HUMAN BREAST EPITHELIAL CELLS

Label-free quantitative *N*-glycome mapping of the secreted and microsome (total membrane) subcellular glycoproteomes of a panel of eight cultured human breast cells (i.e., MCF7, SKBR3, MDA157, MDA231, MDA468, HS578T, HMEC, and MCF10A) displaying diverse cellular features showed differential *N*-glycan processing of the two fractions, **Figure 1A**. The glycoproteins secreted into the cultured media consistently displayed a significantly higher proportion of processed *N*-glycan types (i.e., hybrid, complex, and paucimannose) (74.2–95.0% mol/mol of total *N*-glycome) than the high-mannose-rich microsomal subcellular glycoproteomes (22.1–55.6%, $p < 0.0001$–0.05). Little, if any, correlation between the *N*-glycan processing stage and the cellular doubling time ($R^2 = 0.13$) or the protein secretion rate ($R^2 = 0.35$), respectively, was detected of the secreted glycoproteomes across the cell line panel, Figure S1 in Supplementary Material. No correlation was detected between the *N*-glycan processing stage of the microsomal glycoproteins and the cellular doubling time ($R^2 = 0.04$) or the protein secretion rate ($R^2 = 0.01$).

In-depth, *N*-glycan profiling of the secreted, microsomal, and cell-surface enriched glycoproteomes was carried out for MCF7, MDA468, and MCF10A cells as representative cells for the breast cell line panel. Differential *N*-glycan processing was evident as exampled by the clear differences seen in the *N*-glycome *m/z* profiles of the three subcellular fractions of MCF7 cells, **Figure 1B**. The cell-surface glycoproteins derived from MCF7 and MDA468 (but not MCF10A) cells were subjected to more *N*-glycan processing than microsomal proteins ($p < 0.01$–0.05) and all the three cell lines showed significantly increased abundance of the more processed *N*-glycans on the secreted proteins ($p < 0.0001$–0.01), **Figure 1C**.

SUBCELLULAR-SPECIFIC DISTRIBUTION OF *N*-GLYCAN DETERMINANTS

To further evaluate the subcellular-specific distribution of common *N*-glycosylation determinants, which may be recognized by different immuno-lectins, terminal α-mannose, α-fucose, and α-sialic acid residues were mapped based on the obtained *N*-glycome profiles, **Figure 1D**. As expected from the glycan type distribution, terminating α-mannosylation was found to be significantly reduced on the secreted and cell-surface proteins relative to the microsomal proteins. The α-fucosylation, primarily of the α1,6-(core) type, and α2,3/6-sialylation were concomitantly significantly higher in the secreted fractions than in the cell-surface-enriched fraction (with the exception of fucosylation of MCF7) and in the microsomal fraction of all three cell lines. Taking the incomplete subcellular fractionation into account (see "Proteomics- and GO-Based Assessment of Subcellular Fractionation"), we estimate that very little terminal α-mannosylation is present on protein *N*-glycans in contact with the extracellular environment in the investigated cells and that little α-sialylation and α-fucosylation are carried by intracellular (microsomal) *N*-glycoproteins.

PROTEOMICS- AND GO-BASED ASSESSMENT OF SUBCELLULAR FRACTIONATION

In total, 2,297, 2,636, and 2,042 human proteins were identified across the three subcellular fractions in MCF7, MDA468, and MCF10A, respectively. Putative *N*-glycoproteins fulfilling our strict prediction criteria i.e., presence of the following: one or more sequons (NxT/S, $x \neq P$); and signal peptides (for secreted proteins); and/or transmembrane regions (for membrane-tethered proteins) comprised significant proportions of the subcellular proteomes (15.7–31.0%), Table S3A in Supplementary Material. The GO terms "ER", "Golgi/endosome/plasma membrane", and "extracellular" were used to evaluate the localization/origin of the glycoproteins identified in the subcellular fractions. In agreement with a previous study (23), the GO annotation of the identified proteins showed that the microsomes in general contained a high proportion of ER-residing proteins, **Figures 2A–C**. Although the proteins are only broadly, and possibly somewhat inaccurately, classified on the basis of GO terms, the trends clearly indicated significant enrichment, although not complete isolation, of the desired proteins in the respective subcellular fractions. The ER-based contribution to the microsome was supported by the fact that a significant proportion of the high-mannose *N*-glycans identified in this fraction were of the immature type i.e., $Man_9 \pm Glc_1$ (MCF7: 35.3 ± 0.9%, MDA468: 40.2 ± 2.0%, and MCF10A: 31.8 ± 0.4%, mol/mol of the total high-mannose *N*-glycans), **Figure 2D** (MCF7 data) and Figure S2 in Supplementary Material (MDA468 and MCF10A data).

To further investigate the intracellular *N*-glycosylation and confirm the presence of ER-rich microsomes, the *N*-glycome and proteome of ER- and Golgi-enriched fractions of human colon epithelial cancer cells (SW480) as prepared by the method of sucrose density gradient centrifugation, were mapped and compared to the microsome profiles derived from the same cells, Figure S3A in Supplementary Material. Quantitative analysis of four reliable and representative markers of the ER (i.e., 78 kDa glucose-regulated protein, protein disulfide bond isomerase, calreticulin, and protein transport protein Sec61 alpha isoform 1) and Golgi (i.e., polypeptide *N*-acetylgalactosaminyltransferase 2, β-1,4-galactosyltransferase 1, Golgi apparatus protein 1, and Golgi membrane protein 1) compartments revealed a high abundance of ER-specific proteins in the ER-enriched fraction, Figure S3B in Supplementary Material. However, there was still a significant presence of ER proteins in the Golgi-enriched and microsome fractions. In contrast, the ER-enriched and microsome fractions were essentially free of Golgi proteins, Figure S3C in Supplementary Material. In line with our breast epithelial cell data, the proteins in the ER-enriched fraction contained a significantly higher degree of high-mannose ($Glc_{0\text{-}1}Man_{5\text{-}9}GlcNAc_2$) (92%) *N*-glycans than the proteins in the microsome (75%) and the Golgi-enriched fraction (51%). Taken together, the data confirm that the microsomes of human breast and colon epithelial cells predominantly contain ER proteins and that such intracellular proteins mostly carry high-mannose type *N*-glycosylation. Since the Golgi fraction contains few, if any, ER proteins, it becomes clear that the majority of post-ER *N*-glycans are of the complex type.

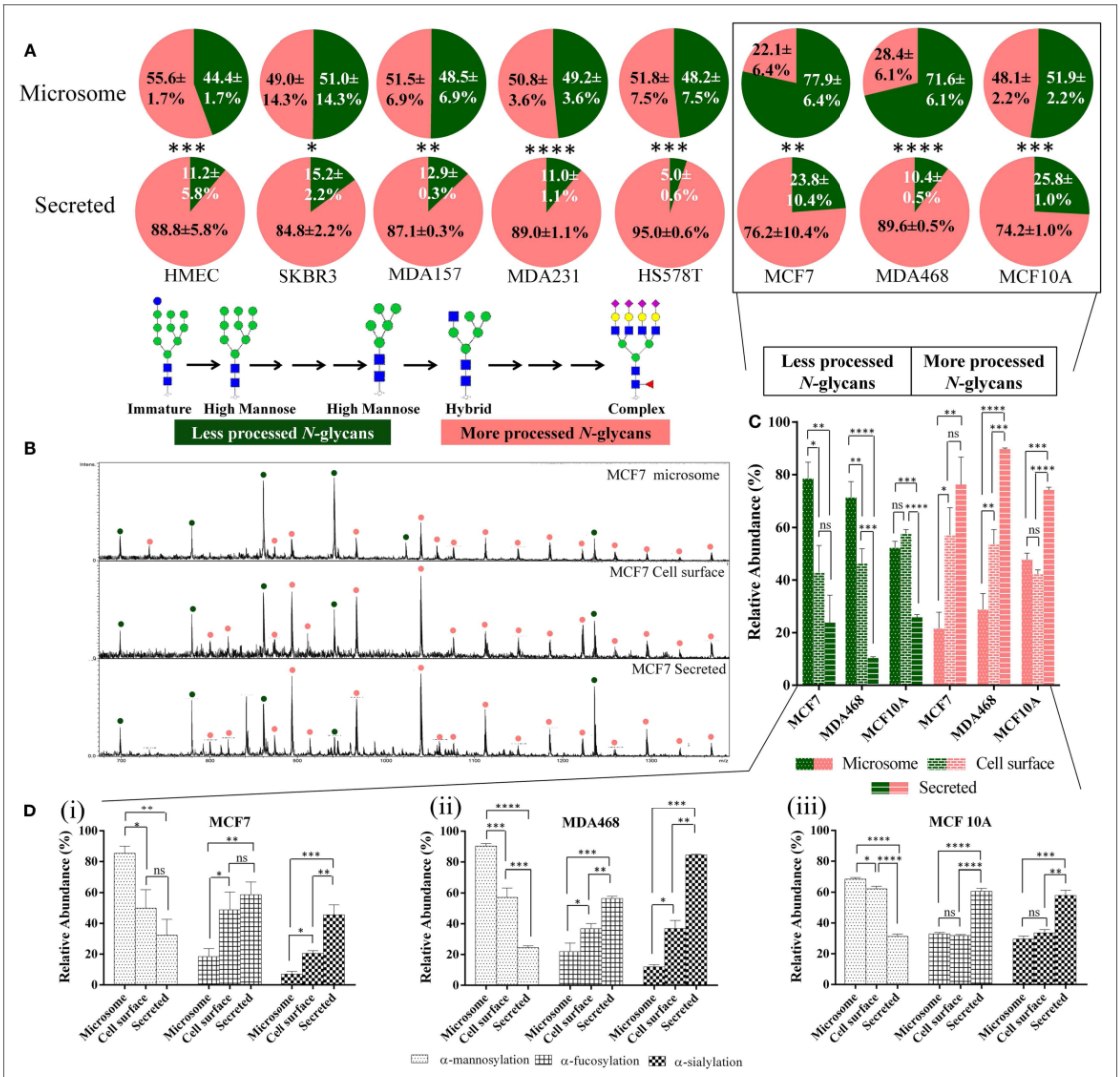


FIGURE 1 | Secreted glycoproteins display more *N*-glycan type processing than microsomal glycoproteins. **(A)** The *N*-glycomes of the microsomal (top) and secreted (bottom) proteins of a panel of eight geno- and phenotypically different cultured human breast epithelial cells (i.e., MCF7, SKBR3, MDA157, MDA231, MDA468, HS578T, HMEC, and MCF10A) were profiled, see Table S1 in Supplementary Material for information of investigated cells. The relative molar abundances (mean ± SD) of more processed *N*-glycans comprising the complex, hybrid, and paucimannose type are presented in light red and the less processed *N*-glycans of the immature and high-mannose type in green (inset). Subcellular-specific *N*-glycosylation of boxed cell lines was investigated further in greater detail. **(B)** Summed *m/z* profiles of the *N*-glycomes derived from microsomal (top), cell-surface (middle), and secreted (bottom) proteins of MCF7 cells. Signals corresponding to *N*-glycans have been assigned as less processed (green) or more processed (light red) *N*-glycan types following the same classification as in **(A)**. **(C)** Relative molar distribution (mean ± SD) of more (right, hybrid/complex/paucimannose, light red bars) and less (left, high mannose, green bars) processed *N*-glycan types of the microsomal (dotted bars), cell-surface (brick), and secreted (banded) proteins of MCF7, MDA468, and MCF10A. **(D)** Subcellular-specific distribution of the *N*-glycan determinants. The proportion of terminal α-mannosylation, α-fucosylation, and α-sialylation (non-reducing end) *N*-glycans of the total *N*-glycome (mol/mol %) on the microsome, cell-surface, and secreted glycoproteomes across MCF7 (i), MDA468 (ii), and MCF10A (iii) breast cell lines were determined from the *N*-glycome profiles. *N*-glycans may terminate with multiple monosaccharide determinants making the values sum to more than 100%. For all panels: ns, not significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $****p < 0.0001$.

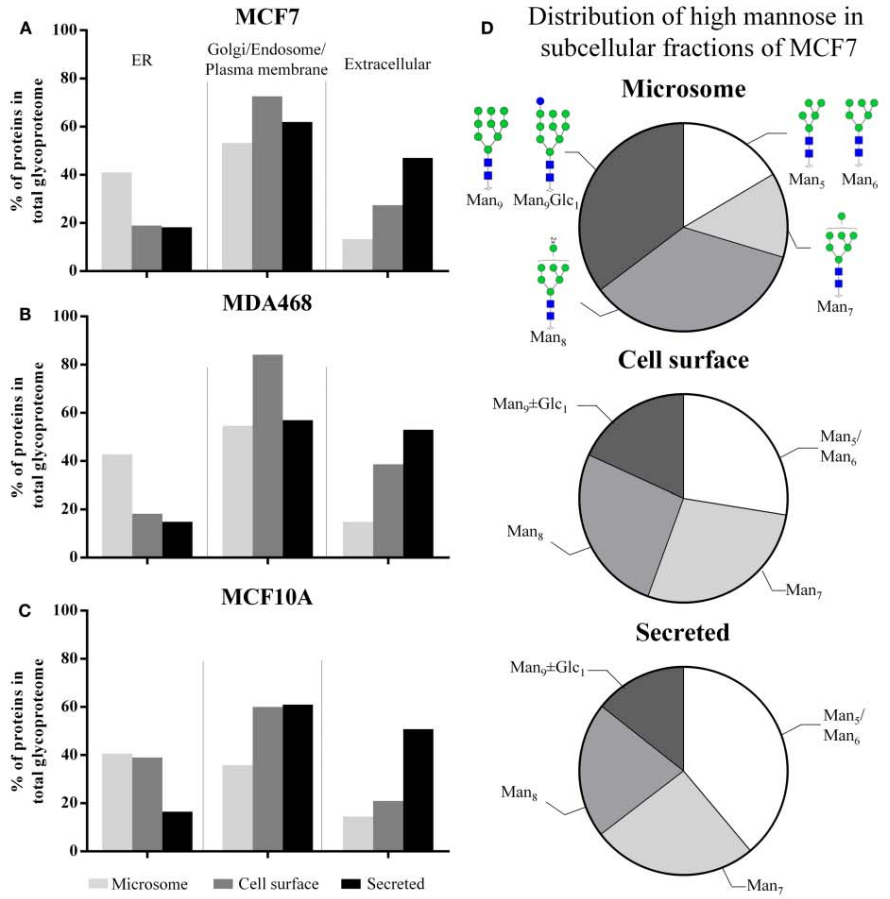


FIGURE 2 | (A–C) The subcellular proteomes of MCF7, MDA468, and MCF10A breast epithelial cell lines were mapped according to GO terms into ER, Golgi/endosome/plasma membrane, and extracellular region classifiers. This confirmed enrichment, but not isolation, of cell-surface and secreted proteins in the respective subcellular fractions. In addition, the classification confirmed that the microsomes contained a significant proportion of ER-residing proteins. **(D)** The subcellular distribution of the high-mannose glycan type series on proteins derived from MCF7 into Man_5, Man_6, Man_7, Man_8, $Man_9 \pm Glc_1$, the latter representing immature *N*-glycans normally only associated with intracellular ER *N*-glycosylation. See Figure S2 in Supplementary Material for the subcellular distribution of the high-mannose glycan type series of MDA468 and MCF10A.

DIFFERENTIAL Asn SITE ACCESSIBILITIES EXPLAIN SUBCELLULAR-SPECIFIC *N*-GLYCOSYLATION

To investigate a possible link between the observed subcellular-specific *N*-glycosylation and protein *N*-glycosylation site accessibility, *in silico* assessment of site accessibility was performed of the identified proteins predicted to be *N*-glycosylated. Due to the laborious and time-consuming approach of determining glycoprotein site accessibility (19), only the most abundant subset of the putative *N*-glycoproteins observed in the subcellular fractions were included in the accessibility assessment. The relative abundances of the individual putative glycoproteins were calculated by a conventional normalized spectral counting strategy; however, the number of sequons of the individual proteins was factored into the calculation to ensure a fair representation of heavily and lightly *N*-glycosylated proteins. We call this term "sequon-weighted normalized spectral counts." Based on sequon-weighted normalized spectral counts, the 100 most abundant glycoproteins uniquely present in the three subcellular fractions, which, by weight, comprised 70–100% of the individual subcellular glycoproteomes, were used to assess glycosylation site accessibility, Table S3B in Supplementary Material. The solvent site accessibilities were determined using an established approach based on van der Waal interactions of the asparagine residue of the glycosylation sites to solvent (19). 3D-glycoprotein structures (experimental or homology modeled) were available for approximately one-third of the 189, 89, and 183 putative *N*-glycoproteins identified

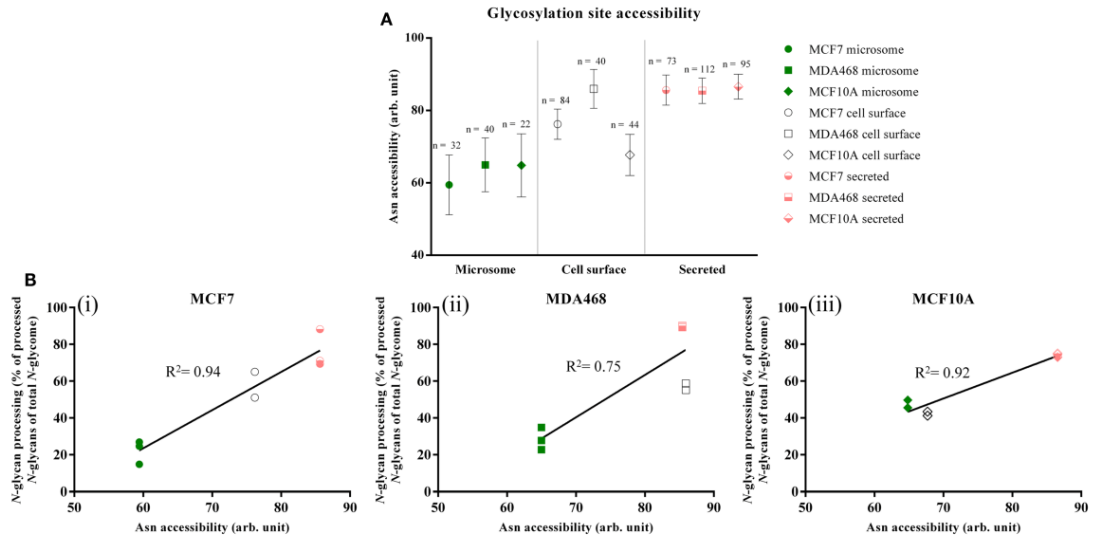


FIGURE 3 | (A) Glycosylation site accessibilities (unit-less, arbitrary values, mean ± SEM) of the microsomal (green), cell-surface (white), and secreted (light red) proteins derived from MCF7, MDA468, and MCF10A breast epithelial cell lines. **(B)** Correlation between the site accessibilities and the *N*-glycosylation processing as measured by the more processed *N*-glycan types (hybrid, complex, and paucimannose) as a molar proportion of the total *N*-glycome for the three subcellular fractions. High correlation coefficients (R^2) indicate strong correlation.

uniquely in the microsome, cell-surface, and secreted fraction, respectively, Figure S4 in Supplementary Material. This yielded site-accessibility datasets covering in total 161 (microsome), 189 (cell-surface), and 236 (secreted) *N*-glycosylation sites from the three cell types.

Differential site accessibilities were observed for the three subcellular glycoproteomes for all three investigated breast cell lines, **Figure 3A** (see also Figures S5A–C in Supplementary Material for an alternative representation showing 95% confidence intervals). Glycosylation sites of secreted glycoproteins were significantly more accessible [MCF7: 85.63 ± 35.47, $n = 73$; MD468: 85.44 ± 36.85, $n = 112$; MCF10A: 86.56 ± 33.54 (all unit-less arbitrary values), $n = 95$] than sites on microsomal proteins (MCF7: 59.44 ± 46.58, $n = 32$; MD468: 64.98 ± 46.99, $n = 40$; MCF10A: 64.84 ± 40.97, $n = 22$, $p < 0.01$). In agreement with the *N*-glycomes that carried a mixture of less processed high-mannose and more processed *N*-glycan types, the sites of cell-surface proteins were intermediately accessible: cell-surface sites were either statistically similar in accessibility to the microsomal protein sites (MCF10A: 67.70 ± 37.66, $n = 44$) or similar to the secreted protein sites (MCF7: 76.20 ± 38.13, $n = 84$; MD468: 85.95 ± 34.08, $n = 40$). For all three breast cell lines, the glycosylation site accessibilities were strongly correlated with the *N*-glycan processing as measured by their glycan type (MCF7: $R^2 = 0.94$; MD468: $R^2 = 0.75$; MCF10A: $R^2 = 0.92$), **Figure 3B**. Higher average glycosylation site accessibility of the secreted and partly also the cell-surface glycoproteins resulted, as such, in more *N*-glycan processing in terms of glycan type formation. Other subcellular-specific *N*-glycosylation signatures including core fucosylation, β-galactosylation, and α-sialylation were found to correlate only weakly or not at all with glycosylation site accessibility upon search for consistent trends across the three different cell lines, Table S4 in Supplementary Material.

DISCUSSION

SUBCELLULAR-SPECIFIC PROTEIN *N*-GLYCOSYLATION OF HUMAN CELLS

All *N*-linked glycoproteins synthesized by a given cell are processed by a common glycosylation machinery. Despite this shared biosynthetic machinery, we observed that a panel of human breast epithelial cells of different geno- and phenotypes, reproducibly produced subcellular glycoproteomes with distinct *N*-glycosylation signatures. The *N*-glycans attached to proteins enriched from the cell-surface, and in particular the secreted glycoproteins, were significantly more processed with respect to their glycan type (i.e., hybrid/complex/paucimannose) than the predominantly high-mannose type microsomal proteins for all investigated cells. As such, subcellular-specific *N*-glycosylation can be predicted to be a general cellular feature not restricted to the investigated breast epithelial cells. Deeper dissection of the intracellular organelle-specificity of colon cell *N*-glycosylation supported this concept. The capacity of human cells to generate multiple subcellular glycoproteomes displaying specific *N*-glycosylation profiles has, to the best of our knowledge, not been systematically investigated.

The importance of cell-surface *N*-glycosylation for cell–cell and cell–protein interactions has prompted several investigations

of the cell-surface (alternatively termed plasma membrane) *N*-glycosylation. High-mannose type *N*-glycans, in particular Man_{8-9} structures, were previously reported to be the dominating features of the plasma membrane of human embryonic stem cells (48) and of cancer cells (49, 50). However, cell lysates and total membrane fractions similar to our microsome preparations were used in these studies suggesting significant contributions from intracellular high-mannose-rich ER-residing *N*-glycoproteins (23). Hence, the actual cell-surface *N*-glycomes in the previous work may not have been accurately captured. Specific cell-surface enrichment methods such as biotinylation labeling strategies used in this study or adhesion-based isolation methods (23) indicate that human cell-surfaces instead are generally decorated with more processed *N*-glycan types.

Of the six cancerous breast cells investigated in this study, only MCF7 and MDA468 displayed predominantly (>70%) high-mannose *N*-glycans of the microsomal proteins. Approximately equal distribution of high-mannose and the more processed *N*-glycan types of microsomal proteins were detected in the remaining four cancerous (SKBR3, MDA157, MDA231, and HS578T) and the two non-cancerous cells (HMEC and MCF10A). In addition, no consistent over-representation of high-mannose *N*-glycans were detected for the secreted proteins derived from the cancerous cell lines relative to the non-cancerous cell lines. Together this indicates that high-mannose *N*-glycosylation is not linked directly to tumorigenesis. Others have associated serum-derived high-mannose *N*-glycoproteins to pathogenesis including cancer and inflammation (5, 51); however, whether these under-processed species are a result of leakage of intracellular glycoproteins as a consequence of cell death or active cellular secretion from intact cells remains to be described. Based on in-depth comparative analysis of the *N*-glycomes derived from secreted proteins of breast and colon epithelial cells of non-cancerous and cancerous nature, we have recently identified several tumor- and sub-type specific *N*-glycosylation signatures amongst the complex *N*-glycans including alterations of sialylation, α1,6-fucosylation, and bisecting β1,4-GlcNAcylation (submitted) (52).

SITE ACCESSIBILITIES MECHANISTICALLY EXPLAIN SUBCELLULAR-SPECIFIC *N*-GLYCOSYLATION

We have previously shown that solvent accessibility of the glycosylation site of *N*-glycoproteins is an important factor in generating protein- and site-specific *N*-glycosylation (19). We used literature-based glycoprofiling of more than 100 mammalian glycoproteins produced under different cellular and physiological conditions to establish that site accessibility of maturely folded glycoproteins correlates with *N*-glycan processing features including glycan type, α1,6-fucosylation and β1,4/6-GlcNAc-branching. We emphasized in that study that relatively large datasets were required to compensate for the potential inaccuracy of the individual PDB structures and the relative simplistic solvent accessibility assessment simulating the accessibility of the processing glycosylation enzymes to the protein glycosylation sites.

Herein, we used a similar approach using our own *N*-glycosylation data acquired from eight cell lines fractionated into subcellular glycoproteomes to further explore the determining features of site-specific *N*-glycosylation in the context of subcellular localization of proteins. Homogenous cell cultures were an essential tool to ensure that the isolated subcellular glycoproteomes were produced simultaneously under the same physiological conditions of the glycosylation machinery. Although the *N*-glycomes, as expected, varied considerably between the different cell lines, our experimental data not only validated the strong correlation of the *N*-glycan type and the glycosylation site accessibility of maturely folded glycoproteins in agreement with our previously report (19), but also mechanistically explained that subcellular-specific *N*-glycosylation is driven by differences in site accessibilities of the individual glycoproteins ending up at different subcellular destinations, **Figure 4**. Intracellular (microsome) *N*-glycoproteins receive little glycan processing of the high-mannose intermediates as a result of limited site accessibility, whereas the secreted *N*-glycoproteins are modified almost entirely to more processed *N*-glycan types due to high site accessibilities. As such, *N*-glycan processing may be a targeting signal or a requirement for intracellular (ER–Golgi-residing) glycoproteins to translocate to the surface for cell-surface integration/secretion via vesicles. Keeping in mind there may be many exceptions to the molecular trends presented here, it is tempting to view the glycosylation site accessibility, and, thus, the *N*-glycan type, as a crude predictor of subcellular location of human glycoproteins.

We have previously linked core fucosylation to glycosylation site accessibility (19). Interestingly, glycosylation site accessibility alone could not explain the differential core fucosylation of the subcellular fractionated proteins in our data: the secreted proteins did not have a higher degree of core fucosylation of complex/hybrid-type *N*-glycans than the cell-surface proteins although the secreted proteins had significantly higher accessibilities. This surprising observation may be explained by a possible advantage of the membrane-embedded cell-surface glycoproteins to achieve preferential interaction with the membrane-bound fucosyltransferase 8 (FUT8) facilitating the addition of α1,6-fucose residues to the chitobiose cores of *N*-glycans. Soluble (luminal) glycoproteins may be less likely to interact with FUT8. This explanation is congruent with our previous observation describing FUT8 discrimination of soluble *N*-glycoproteins over membrane *N*-glycoproteins (19). Similar processing preference was not observed for the multiple processing enzymes responsible for the formation of the glycan type. As expected, the glycan modification more distal from the protein surface i.e., β-1,3/4-galactosylation and α2,3/6-sialylation were not found to be correlated with glycosylation site accessibility since the glycosyltransferases most likely have unhindered access to the substrates relatively far from the protein surface. By the same token, we cannot rule out that a more refined accessibility determination approach, which not only takes into account the glycosylation site solvent accessibility, but also the conjugated *N*-glycans (53–56), may expose that other subcellular-specific *N*-glycan features correlate with site accessibility. New developments in glycoproteomics may also support and strengthen these observations by giving more accurate insight into the connectivity of glycosylation of the individual protein carriers (31). Finally, it should be emphasized that although the subcellular glycoproteomes share a common biosynthetic machinery, slightly different trafficking rates and/or routes to their final destinations are factors that may

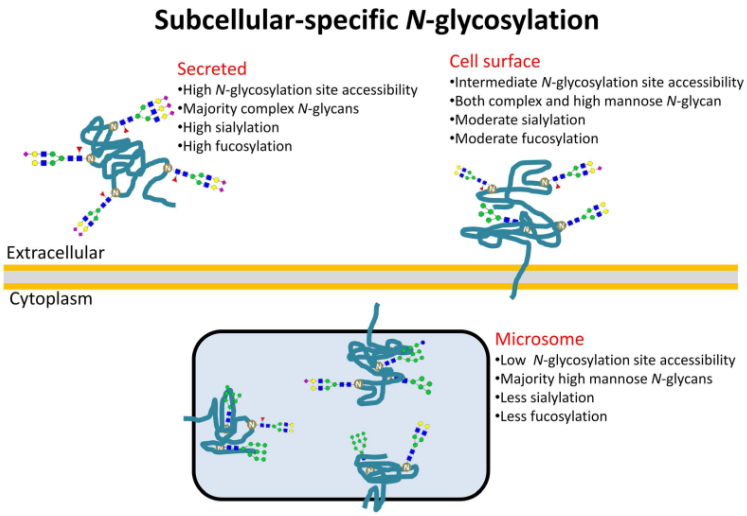


FIGURE 4 | Subcellular-specific *N*-glycosylation is driven by differential solvent accessibility to the *N*-glycosylation sites on maturely folded glycoproteins. Consequently, the *N*-glycans of secreted, cell-surface, and microsomal proteins receive high, intermediate, and low *N*-glycosylation processing, respectively, and as a result, display distinct glyco-determinant signatures.

contribute to yield distinct subcellular *N*-glycosylation. Other cellular factors including the glycosylation enzyme activity or the availability of nucleotide donors may also indirectly contribute to subcellular-specific *N*-glycosylation by having differential effects on the individual subcellular glycoproteomes.

SUBCELLULAR-SPECIFIC GLYCO-DETERMINANTS IN IMMUNITY

The distinct *N*-glycosylation signatures carried by the subcellular glycoproteomes may be functionally important in immunity if we consider the key role of *N*-glycans as mediators for an effective innate and adaptive immune response through their specific interaction with endogenous lectins. In addition, opportunistic pathogens often use exposed *N*-glycan determinants as receptors for adhesion using exogenous lectins (11). The observed subcellular-specific glycosylation is here briefly discussed in the context of glyco-immunity and infection; it is stressed that further empirical evidence is required to validate these proposed relationships.

We found that α-sialylation was a more abundant feature of the secreted *N*-glycoproteins than cell-surface proteins. High sialylation of secreted glycoproteins is essential to mask penultimate galactose residues from being exposed and recognized by asialoglycoprotein receptors, a C-type lectin (12). Thus, the high sialylation of secreted glycoproteins may be a requirement to ensure prolonged circulation half-life. In addition, high sialylation of secreted glycoproteins can act as a strong decoy for the less sialylated cell-surface proteins, to which opportunistic pathogens are known to adhere through sialic acid-recognizing I-type lectins (alternatively termed siglecs) (57, 58). Displaying less-than-complete sialylation of the cell-surface proteins also ensures that a gradient of biological activity toward endogenous siglecs for cellular signaling and endocytosis (59) is maintained through structural diversity, which may confer an immunological advantage to the host cells (60).

The secreted *N*-glycoproteins were over-represented in α1,6-core fucosylation relative to the cell-surface proteins. In line with our previous observations, the higher degree of core fucosylation may serve to either mask hydrophobic patches to regulate stability/solubility of the secreted *N*-glycoproteins (19) or to protect these more exposed proteins from proteolytic degradation in the extracellular environment. It could be speculated that the membrane-embedded nature of cell-surface glycoproteins would make them more stable by not facing solubility issues in their local environment and less vulnerable to proteolytic digestion, thereby having less requirement for steric protection provided by a bulky fucose residue proximal to the protein surface.

We and others have observed that α-mannose is an unusual terminating structural determinant in the extracellular environment (61,62). This may partly be explained by the intracellular functions of mannose (and glucose) terminating *N*-glycans (16, 17). The presence of several mannose recognizing lectins in the extracellular environment including mannan binding protein (MBP), DC-SIGN, and macrophage mannose receptors may be relevant in the context of apoptosis when mannose terminating *N*-glycoproteins are exposed to the extracellular environment. In particular, MBP is a key player and a first line of defense in innate immunity, enabling phagocytosis of apoptotic cells through its binding to exposed immature or under-processed glycans or to pathogens carrying mannosylated glycoproteins (63, 64). Hiding mannose inside cells under physiological conditions could thus be viewed as being

critical to avoiding the unnecessary onset of inflammation and auto-immunity. The presence of extracellular α-mannosylation would, as such, be indicative of pathophysiological conditions. In support of this hypothesis, high-mannose containing glycoforms of intracellular adhesion molecule 1 and EGF receptor on cell-surfaces were shown to contribute to endothelial inflammation (61) and correlated with poor prognosis of various cancers, respectively (61, 62).

It has been noted that the structure and function of the protein *N*-glycome is different within and outside human cells and that these differences may be shaped by evolutionary forces (60). We are the first to systematically investigate and mechanistically explain some aspects of subcellular-specific *N*-glycosylation. We conclude that human cells have developed protein structure-specific mechanisms including differential *N*-glycosylation site accessibilities to generate subcellular glycoproteomes that display distinct *N*-glycosylation phenotypes using a shared biosynthetic machinery. Establishing this relationship is of general significance to glycobi ologists and in particular to molecular immunologists due to the functional relevance of *N*-glycan determinants acting as ligands for the spectrum of endogenous lectins involved in facilitating an efficient immune response.

ACKNOWLEDGMENTS

This work was supported by Macquarie University Research Excellence Scheme postgraduate scholarship and ARC Super Science (FS110200026) and Discovery (DP110104958) Grants. Morten Thaysen-Andersen was funded an Early Career Fellowship by Cancer Institute NSW. The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/Journal/10.3389/fimmu.2014.00404/abstract

REFERENCES

1. Freeze HH. Genetic defects in the human glycome. *Nat Rev Genet* (2006) **7**(7):537–51. doi:10.1038/nrg1894
2. Freeze HH. Understanding human glycosylation disorders: biochemistry leads the charge. *J Biol Chem* (2013) **288**(10):6936–45. doi:10.1074/jbc.R112.429274
3. Cylwik B, Naklicki M, Chrostek L, Gruszewska E. Congenital disorders of glycosylation. Part I. Defects of protein *N*-glycosylation. *Acta Biochim Pol* (2013) **60**(2):151–61.
4. Venkatakrishnan V, Packer NH, Thaysen-Andersen M. Host mucin glycosylation plays a role in bacterial adhesion in lungs of individuals with cystic fibrosis. *Expert Rev Respir Med* (2013) **7**(5):553–76. doi:10.1586/17476348.2013.837752
5. Scott DW, Patel RP. Endothelial heterogeneity and adhesion molecules *N*-glycosylation: implications in leukocyte trafficking in inflammation. *Glycobiology* (2013) **23**(6):622–33. doi:10.1093/glycob/cwt014
6. Stuchlová Horynová M, Raška M, Clausen H, Novak J. Aberrant O-glycosylation and anti-glycan antibodies in an autoimmune disease IgA nephropathy and breast adenocarcinoma. *Cell Mol Life Sci* (2013) **70**(5):829–39. doi:10.1007/s00018-012-1082-6
7. Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. Cell surface protein glycosylation in cancer. *Proteomics* (2014) **14**(4–5):525–46. doi:10.1002/pmic.201300387
8. Boscher C, Dennis JW, Nabi IR. Glycosylation, galectins and cellular signaling. *Curr Opin Cell Biol* (2011) **23**(4):383–92. doi:10.1016/j.ceb.2011.05.001
9. Schwarz F, Aebi M. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* (2011) **21**(5):576–82. doi:10.1016/j.sbi.2011.08.005
10. Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* (2004) **73**:1019–49. doi:10.1146/annurev.biochem.73.011303.073752
11. van Kooyk Y, Rabinovich GA. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol* (2008) **9**(6):593–601. doi:10.1038/ni.f.203
12. Figdor CG, van Kooyk Y, Adema GJ. C-type lectin receptors on dendritic cells and Langerhans cells. *Nat Rev Immunol* (2002) **2**(2):77–84. doi:10.1038/nri723
13. Kawasaki T, Ii M, Kozutsumi Y, Yamashina I. Isolation and characterization of a receptor lectin specific for galactose/N-acetylgalactosamine from macrophages. *Carbohydr Res* (1986) **151**:197–206. doi:10.1016/S0008-6215(00)90340-9
14. van Vliet SJ, Saeland E, van Kooyk Y. Sweet preferences of MGL: carbohydrate specificity and function. *Trends Immunol* (2008) **29**(2):83–90. doi:10.1016/j.it.2007.10.010
15. Rabinovich GA, van Kooyk Y, Cobb BA. Glycobiology of immune responses. *Ann N Y Acad Sci* (2012) **1253**:1–15. doi:10.1111/j.1749-6632.2012.06492.x
16. Aebi M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta* (2013) **1833**(11):2430–7. doi:10.1016/j.bbamcr.2013.04.001
17. Aebi M, Bernasconi R, Clerc S, Molinari M. *N*-glycan structures: recognition and processing in the ER. *Trends Biochem Sci* (2010) **35**(2):74–82. doi:10.1016/j.tibs.2009.10.001
18. Parodi AJ. Protein glucosylation and its role in protein folding. *Annu Rev Biochem* (2000) **69**:69–93. doi:10.1146/annurev.biochem.69.1.69
19. Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of *N*-glycan type, core fucosylation and branching. *Glycobiology* (2012) **22**(11):1440–52. doi:10.1093/glycob/cws110
20. Sumer-Bayraktar Z, Nguyen-Khuong T, Jayo R, Chen DD, Ali S, Packer NH, et al. Micro- and macroheterogeneity of *N*-glycosylation yields size and charge isoforms of human sex hormone binding globulin circulating in serum. *Proteomics* (2012) **12**(22):3315–27. doi:10.1002/pmic.201200354
21. Parekh RB, Dwek RA, Thomas JR, Opdenakker G, Rademacher TW, Wittwer AJ, et al. Cell-type-specific and site-specific *N*-glycosylation of type I and type II human tissue plasminogen activator. *Biochemistry* (1989) **28**(19):7644–62. doi:10.1021/bi00445a021
22. Rudd PM, Dwek RA. Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol* (1997) **32**(1):1–100.
23. Mun JY, Lee KJ, Seo H, Sung MS, Cho YS, Lee SG, et al. Efficient adhesion-based plasma membrane isolation for cell surface *N*-glycan analysis. *Anal Chem* (2013) **85**(15):7462–70. doi:10.1021/ac401431u
24. Bones J, Mittermayr S, O'Donoghue N, Guttman A, Rudd PM. Ultra performance liquid chromatographic profiling of serum *N*-glycans for fast and efficient identification of cancer associated alterations in glycosylation. *Anal Chem* (2010) **82**(24):10208–15. doi:10.1021/ac102860w
25. North SJ, Hitchen PG, Haslam SM, Dell A. Mass spectrometry in the analysis of N-linked and O-linked glycans. *Curr Opin Struct Biol* (2009) **19**(5):498–506. doi:10.1016/j.sbi.2009.05.005
26. Hua S, Lebrilla C, An HJ. Application of nano-LC-based glycomics towards biomarker discovery. *Bioanalysis* (2011) **3**(22):2573–85. doi:10.4155/bio.11.263
27. Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc* (2012) **7**(7):1299–310. doi:10.1038/nprot.2012.063
28. Parker BL, Thaysen-Andersen M, Solis N, Scott NE, Larsen MR, Graham ME, et al. Site-specific glycan-peptide analysis for determination of *N*-glycoproteome heterogeneity. *J Proteome Res* (2013) **12**(12):5791–800. doi:10.1021/pr400783j
29. Nwosu CC, Seipert RR, Strum JS, Hua SS, An HJ, Zivkovic AM, et al. Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *J Proteome Res* (2011) **10**(5):2612–24. doi:10.1021/pr2001429
30. Zauner G, Koeleman CA, Deelder AM, Wuhrer M. Nano-HPLC-MS of glycopeptides obtained after nonspecific proteolysis. *Methods Mol Biol* (2013) **951**:113–27. doi:10.1007/978-1-62703-146-2_9
31. Thaysen-Andersen M, Packer NH. Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteomes. *Biochim Biophys Acta* (2014) **1844**(9):1437–52. doi:10.1016/j.bbapap.2014.05.002
32. Kolarich D, Lepenies B, Seeberger PH. Glycomics, glycoproteomics and the immune system. *Curr Opin Chem Biol* (2012) **16**(1–2):214–20. doi:10.1016/j.cbpa.2011.12.006

33. Balch WE, Dunphy WG, Braell WA, Rothman JE. Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. *Cell* (1984) **39**(2 Pt 1):405–16. doi:10.1016/0092-8674(84)90019-9
34. Packer NH, Lawson MA, Jardine DR, Redmond JW. A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconj J* (1998) **15**(8):737–47. doi:10.1023/A:1006983125913
35. Leymarie N, Griffin PJ, Jonscher K, Kolarich D, Orlando R, McComb M, et al. Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: the ABRF glycoprotein research multi-institutional study 2012. *Mol Cell Proteomics* (2013) **12**(10):2935–51. doi:10.1074/mcp.M113.030643
36. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* (2011) **8**(10):785–6. doi:10.1038/nmeth.1701
37. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* (2001) **305**(3):567–80. doi:10.1006/jmbi.2000.4315
38. Hiller K, Grote A, Scheer M, Münch R, Jahn D. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* (2004) **32**:W375–9. doi:10.1093/nar/gkh378
39. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* (2004) **338**(5):1027–36. doi:10.1016/j.jmb.2004.03.016
40. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* (2002):310–22.
41. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* (1986) **5**(4):823–6.
42. Hubbard SJ, Thornton JM. *NACCESS Computer Program*. Department of Biochemistry and Molecular Biology, University Colloge London (1993). Available from: http://wolf.bms.umist.ac.uk/naccess/
43. Engelen S, Trojan LA, Sacquin-Mora S, Lavery R, Carbone A. Joint evolutionary trees: a large-scale method to predict protein interfaces based on sequence sampling. *PLoS Comput Biol* (2009) **5**(1):e1000267. doi:10.1371/journal.pcbi.1000267
44. Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat Protoc* (2011) **6**(9):1341–54. doi:10.1038/nprot.2011.367
45. Joseph AP, Valadié H, Srinivasan N, de Brevern AG. Local structural differences in homologous proteins: specificities in different SCOP classes. *PLoS One* (2012) **7**(6):e38805. doi:10.1371/journal.pone.0038805
46. Chothia C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* (1976) **105**(1):1–12. doi:10.1016/0022-2836(76)90191-1
47. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* (1971) **55**(3):379–400. doi:10.1016/0022-2836(71)90324-X
48. An HJ, Gip P, Kim J, Wu S, Park KW, McVaugh CT, et al. Extensive determination of glycan heterogeneity reveals an unusual abundance of high mannose glycans in enriched plasma membranes of human embryonic stem cells. *Mol Cell Proteomics* (2012) **11**(4):M111010660. doi:10.1074/mcp.M111.010660
49. Hua S, Saunders M, Dimapasoc LM, Jeong SH, Kim BJ, Kim S, et al. Differentiation of cancer cell origin and molecular subtype by plasma membrane *N*-glycan profiling. *J Proteome Res* (2013) **13**(2):961–8. doi:10.1021/pr400987f
50. Liu X, Nie H, Zhang Y, Yao Y, Maitikabili A, Qu Y, et al. Cell surface-specific *N*-glycan profiling in breast cancer. *PLoS One* (2013) **8**(8):e72704. doi:10.1371/journal.pone.0072704
51. de Leoz ML, Young LJ, An HJ, Kronewitter SR, Kim J, Miyamoto S, et al. High-mannose glycans are elevated during breast cancer progression. *Mol Cell Proteomics* (2011) **10**(1):M110.002717. doi:10.1074/mcp.M110.002717
52. Sethi MK, Thaysen-Andersen M, Smith JT, Baker MS, Packer NH, Hancock WS, et al. Comparative *N*-glycan profiling of colorectal cancer cell lines reveals unique bisecting GlcNAc and alpha-2,3-linked sialic acid determinants are associated with membrane proteins of the more metastatic/aggressive cell lines. *J Proteome Res* (2014) **13**(1):277–88. doi:10.1021/pr400861m
53. Frank M, Schloissnig S. Bioinformatics and molecular modeling in glycobiology. *Cell Mol Life Sci* (2010) **67**(16):2749–72. doi:10.1007/s00018-010-0352-4
54. Lutteke T. Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr D Biol Crystallogr* (2009) **65**(Pt 2):156–68. doi:10.1107/S0907444909001905
55. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* (2004) **14**(2):103–14. doi:10.1093/glycob/cwh008
56. Petrescu AJ, Wormald MR, Dwek RA. Structural aspects of glycomes with a focus on *N*-glycosylation and glycoprotein folding. *Curr Opin Struct Biol* (2006) **16**(5):600–7. doi:10.1016/j.sbi.2006.08.007
57. Sharon N. Carbohydrate – lectin interactions in infectious disease. In: Kahane I, Ofek I, editors. *Toward Anti-Adhesion Therapy for Microbial Diseases*. Springer (1996). p. 1–8. doi:10.1007/978-1-4613-0415-9_1
58. Peterson R, Cheah WY, Grinyer J, Packer N. Glycoconjugates in human milk: protecting infants from disease. *Glycobiology* (2013) **23**(12):1425–38. doi:10.1093/glycob/cwt072
59. Crocker PR, Paulson JC, Varki A. Siglecs and their roles in the immune system. *Nat Rev Immunol* (2007) **7**(4):255–66. doi:10.1038/nri2056
60. Varki A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* (2006) **126**(5):841–5. doi:10.1016/j.cell.2006.08.022
61. Johns TG, Mellman I, Cartwright GA, Ritter G, Old LJ, Burgess AW, et al. The antitumor monoclonal antibody 806 recognizes a high-mannose form of the EGF receptor that reaches the cell surface when cells over-express the receptor. *FASEB J* (2005) **19**(7):780–2. doi:10.1096/fj.04-1766fje
62. Scott DW, Dunn TS, Ballestas ME, Litovsky SH, Patel RP. Identification of a high-mannose ICAM-1 glycoform: effects of ICAM-1 hypoglycosylation on monocyte adhesion and outside in signaling. *Am J Physiol Cell Physiol* (2013) **305**(2):C228–37. doi:10.1152/ajpcell.00116.2013
63. Ogden CA, deCathelineau A, Hoffmann PR, Bratton D, Ghebrehiwet B, Fadok VA, et al. C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. *J Exp Med* (2001) **194**(6):781–95. doi:10.1084/jem.194.6.781
64. Takahashi K, Ezekowitz RA. The role of the mannose-binding lectin in innate immunity. *Clin Infect Dis* (2005) **41**(Suppl 7):S440–4. doi:10.1086/431987

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 July 2014; accepted: 07 August 2014; published online: 25 August 2014.
Citation: Lee LY, Lin C-H, Fanayan S, Packer NH and Thaysen-Andersen M (2014) Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. Front. Immunol. 5:404. doi: 10.3389/fimmu.2014.00404
This article was submitted to Immunotherapies and Vaccines, a section of the journal Frontiers in Immunology.

CHAPTER 5

Publication IV - An optimized approach for enrichment of glycoproteins using native multi-lectin affinity chromatography

Proteomics has already made significant discoveries and advanced our understanding in cancer research. While glycomics is just maturing these years it is moving at a rapid pace and tangible examples in this thesis have been presented showcasing the strength of glycomics to explore the involvement of protein glycosylation in breast cancer. However, one of the remaining challenges is to link the knowledge of these two "omics" disciplines to enable an improved molecular understanding of the disease. As such, the ability to characterize intact glycopeptides provides the means to integrate the information of the carrier proteins, the specific modified site and the structure of the attached glycans. This last chapter focuses on optimizing a multi-lectin affinity chromatography platform to enrich for specific subsets of disease-related glycoproteins from the breast cancer cell line, MCF7, which can then be profiled using LC-MS/MS technologies. It is expected that this targeted glycoprotein enrichment strategy together with more untargeted enrichment strategies will provide important tools needed to mature the important field of glycoproteomics.

Pages 172-180 (Publication 4) of this thesis have been removed as they contain published material under copyright. Removed contents published as:

Lee, L. Y., Hincapie, M., Packer, N., Baker, M. S., Hancock, W .S. and Fanayan, S. (2012), An optimized approach for enrichment of glycoproteins from cell culture lysates using native multi-lectin affinity chromatography. *Journal of Separation Science*, vol. 35, no.18, pp. 2445-2452. https://doi.org/10.1002/jssc.201200049

CHAPTER 6

GENERAL DISCUSSION

6.1 Thesis summary

The primary hypothesis driving the studies in this thesis is that changes in the expression of the cellular proteome and *N*-glycome are bio-indicators of cancer development. This hypothesis was tested by surveying the proteomes and *N*-glycomes profiles of two subcellular fractions (i.e. secreted and membrane fractions) extracted from a panel of tumorigenic human cell lines models of breast cancer representing the most common clinical breast cancer subtypes, i.e. ER+/PR+, HER2-enriched and basal-like (basal A and B) breast cancers. The resulting profiles generated using well-established proteomes and glycomics technologies were related to the corresponding molecular "fingerprints" derived from normal reference cells representing healthy physiology.

Chapter 2 described how shotgun proteomics identified over 3,000 proteins in the secreted and membrane fractions of the cultured breast epithelial cells. The secreted fraction demonstrated to be a promising source for the detection of molecular changes associated with cancer. Nearly twice as many known breast cancer-related proteins were secreted by the tumor cells compared to the non-tumorigenic cells. In both subcellular fractions, the majority of the differentially expressed proteins present in all three breast cancer subtypes were regulated in the same direction, indicating that the core molecular signatures underlying breast malignancy are likely to be conserved across the various subtypes of breast cancer. However, subtype-specific functional analysis of the resulting proteomes revealed that unique proteins in each of the three subtypes may still be involved in the activation of the common pathological mechanism driving the development and progression of cancer, e.g. in the GPCR pathway. Additionally, it is evident that the cellular-matrix integrity in the highly metastatic and invasive MDA231 cells was notably perturbed compared to the more weakly metastatic cells, MCF7 and SKBR3. These results clearly underpin the significance of the in-depth functional proteome analysis and highlighted the importance of integrating bioinformatics tools capable of handling large datasets into the workflow and the data interpretation.

Although most proteomics researchers are satisfied with confidently identifying and accurately quantifying the proteins in biological samples, additional layers of structural complexity exist for the proteome. Proteins often undergo PTMs, which are often ignored in regular proteomics workflow. Protein glycosylation is one of the most abundant types of modification, which gives rise to extensive macro- and micro-heterogeneity. Aberrant *N*-glycosylation has been implicated in many types of human cancers as well as in other diseases. In *Chapter 3*, these molecular changes were probed by performing global *N*-glycome analyses of the secreted and membrane proteins derived from five epithelial breast cancer cells and non-tumorigenic breast epithelial cell line. The tumor cells corresponded to the three common breast cancer subtypes – luminal A (MCF7), HER2-enriched (SKBR3) and basal B (MDA157, MDA231 and HS578T). To the best of our knowledge, this is the first detailed characterization of the *N*-glycosylation of secreted proteins from a panel of breast cell lines.

An important aim of this PhD study is to improve our understanding of the role of altered glycosylation in various subtypes of breast cancer. The *N*-glycomes of both secreted and membrane proteins fractions were in good agreement with the tumor-associated *N*-glycosylation features consistently reported in the literature including a higher degree of sialylation, fucosylation and branching of the *N*-glycans. Importantly, breast cancer subtype-specific *N*-glycan analysis, which is rarely performed in glycomics type studies as judging from the available literature, revealed distinctive trends in the *N*-glycosylation associated with each cancer subtype. This clearly demonstrates the need to perform detailed glycome analysis to unravel subtle molecular changes on tumor-associated proteins and improve our understanding of breast cancer given the heterogeneous nature of the disease.

Concomitantly, *N*-glycan changes have been correlated with altered expression of the relevant enzymes responsible for their occurrence such as sialylatransferases, fucosyltransferases, *N*-

acetylglucosaminyltransferases and galactosyltransferases [247, 263, 345, 346]. Several investigations have also demonstrated a positive correlation of cancer-associated glycan structures with metastasis and prognosis of patients [241, 246, 258]. Serum-based studies are clinically relevant as cancer-associated glycan structures identified have the potential to serve as disease biomarkers. Over-expression of bi-, tri- and tetra-antennary glycans containing Lewis-type epitopes are often observed in the serum of breast cancer patients [259, 347]. Although protein *N*-glycosylation changes in the breast cancer cultured cells investigated correlated well with those found in the serum, some of the structures identified, such as LacdiNAc and bisecting *N*-glycans, have rarely been described in clinical samples and/or *in vitro* studies. A possible explanation that they have not been observed in serum could be that their presence is masked by the Lewis-type epitopes, which have been shown to be released from the highly abundant glycoproteins produced in the liver in response to inflammation [348].

The *N*-glycome data not only affirmed that *N*-glycosylation is closely associated with breast cancer, but also demonstrated, as the first study to do so, that secreted *N*-glycan expression profiles recapitulate the major luminal-basal distinction observed in breast tumor tissues. Thus, it supports that the selected panel of epithelial breast cell lines serves as a reasonably accurate *in vitro* models for investigating the deregulation of the proteome and in particular the protein *N*-glycosylation in breast cancer.

The panel of breast cancer cell lines used in this study are well characterized in terms of their genome and transcriptome profiles. High quality data are often made publicly available to the scientific community by depositing data into repositories to expand current knowledge on cancer. In the light of system-wide understanding of altered glycosylation in cancer, the relevant glycosyltransferase transcriptional data were extracted from publicly available databases. Some of the *N*-glycan changes observed in the study, for example, increased sialylation and fucosylation,

correlated well to their corresponding transcript levels, while others did not, suggesting a more complex regulation involved in glycosylation alteration. However, a potential caveat in assessing external data is that the cell culturing conditions may differ from those used in this study and therefore could affect the analysis outcome by introducing another level of variability.

In the past two decades, research into global cellular proteome, and more recently the global glycome has expanded tremendously. Such efforts are in part driven by the expectations that system-wide analyses will lead to the discovery of novel cancer biomarkers and therapeutic drug targets and in part due to the expectation that such holistic molecular mapping of cells and tissues will yield fundamental insight into disease mechanisms, which was not easily achievable with previous analytical techniques. More than hundreds of discovery-based proteomics research studies have been undertaken in the past 30 years, but despite this intense research focus no reliable protein biomarkers predictive of breast cancer have been approved by authorities and translated into the clinic [144]. One reason may arise from the multifaceted nature of the proteome, which is regulated in a spatial-temporal sense and show individual specific differences. In addition, identifying unique proteome features for breast cancer has proven extreme challenging due to the heterogeneous nature of the disease, which is now recognized not as a single disease but comprised of at least 10 distinct subtypes [349]. In addition, many subtly molecular alterations are observed within the individual subtypes creating an intricate level of molecular heterogeneity. These structural complexities are also applicable to the glycome, which displays even greater diversity due to its structural heterogeneity. The lack of success of establishing specific and sensitive biomarkers has taught us that the most promising approach to accurately detect cancer may not be carried out using single biomolecule detection; the predictive power can be dramatically increased with the use of a combination of known tumor protein biomarkers [350]. Results from the protein-protein network analysis in Chapter 2 strongly

support this notion where the proteome data showed that deregulated pathways in the tumor cells were orchestrated by sets of closely related proteins.

Alternatively, as illustrated by the two following examples, combining molecular features e.g. by using the concentration of a tumor-related protein and the presence of its tumor-specific glycoforms may be an avenue to increase the biomarker features of glycosylated proteins. Serum α-fetoprotein has been widely used as a tumor marker for hepatocellular carcinomas (HCC), but its level is also increased in other benign liver diseases decreasing the specificity of the protein. Interestingly, the core fucosylated *N*-glycosylated form of human α-fetoprotein is a reasonably specific marker for the detection of HCC [351]. Similarly, detection of both the level of human prostate-specific antigen (PSA) and its *N*-glycoforms may greatly improve the early diagnosis of prostate cancer [352]. Indeed, detailed characterization of disease-specific glycosylation is now regarded as a more effective strategy for the discovery of novel biomarkers [353]. The emerging technologies for site-specific analysis of glycoproteins via developing glycoproteomics platforms will play a key role in this area.

A major disadvantage using the glycomics approach in this study was that data analysis required manual interpretation for the detailed assignment of glycan structures. This task was challenging and time-consuming due to the structural complexity, heterogeneity and non-template driven nature of glycans. The approach would have benefitted if high-throughput computational tools were available, like those that have been developed for proteomics analysis. Such platforms are urgently needed for the advancement of the glycomics field. However, developing software tools for glycomics annotation is far more complicated compared to those of proteomics because of the complexity of glycan data and limited integrative database resources. Some of these needs are being met by the program uniCarbKB that aims to provide integrated online resources to glycobiologists. However, more high-throughput computational tools with sophisticated

algorithms have to be developed for annotating glycomics data so as to expand the limited databases.

Another limitation encountered during interpretation of proteomics and glycomics data is the absence of verification studies using clinical materials such as breast tumor tissues or breast cancer plasma/serum. This limitation is partially addressed by correlating the data to clinical-based findings. Additionally, in order to ensure a comprehensive coverage of the breast cancer subtypes investigated in this project, a large panel of breast cancer cell lines were used.

While surveying a larger panel of eight breast epithelial cell lines including those studied in Chapter 2 and 3, an interesting molecular trend was observed – subcellular-specific distribution of the *N*-glycan types was consistently found on glycoproteins in the secreted and membrane fractions from all investigated cell lines. The secreted glycoproteins displayed mostly highly processed (complex type) *N*-glycans whereas the membrane glycoproteins were largely immature (high mannose) *N*-glycans. Although noted in passing in various glycomics studies of biological secreted fluids, this differential *N*-glycan processing of proteins derived from various subcellular fractions has not previously been explored.

In *Chapter 4*, a systematic investigation of the subcellular-specific *N*-glycosylation was carried out. The particular focus was directed to the secreted, cell surface and intracellular subcellular glycoproteomes of three cultured breast epithelial cells, MCF10A, MCF7 and MDA468. The strong correlation between *N*-glycan type formation and *N*-glycosylation site accessibility confirmed that subcellular-specific *N*-glycosylation arise at least in part from, and not solely determined by, differential solvent site accessibility of proteins localising to the different subcellular fractions. It is important to emphasize that the molecular relationship presented in this study is based on predominantly Golgi-residing glycosylation enzymes acting on maturely

folded proteins. The correlation therefore does not apply to immature and partially folded proteins in the ER where the initial *N*-glycan processing (trimming) of high mannose glycans occurred. As discussed previously, other factors may affect the *N*-glycan processing including protein secretion rate and ER/Golgi residence time.

There are, however, potential flaws of the approach used to measure glycosylation site solvent accessibility based on the widely-used solvent accessibility determination program called NACCESS. The software calculates the atomic accessible area when a probe (5 Å radius) is rolled around the protein surface, predicting van der Waal's interactions. PDB 3D structures obtained from X-ray crystallography were selected following a set of parameters, and high-sequence homologues were used if no PDB structures were available. Prior to measurement, any water molecules, sugars, ligands and other hetero-atoms/molecules that were not part of the core polypeptide chain were removed from the protein surface. In this study, relatively large datasets were used, which were needed to compensate for the potential inaccuracy arising from factors such as individual PDB structures, assuming glycosylation of all polypeptide chains fulfilling the basic criteria for *N*-glycosylation (i.e. presence of signal peptide and sequon and GO-categorized as a membrane/soluble protein) and the relatively simplistic solvent accessibility measurement simulating the accessibility of the processing glycosylation enzymes to the protein glycosylation site.

Higher site accessibility facilitated the presentation of more processed *N*-glycans on the cell surface and secreted glycoproteins while less glycosylation site accessibility restricted *N*-glycan processing leading to under-process *N*-glycans on proteins located predominantly in intracellular organelle membranes. The study discussed the importance of molecular interactions of the subcellular-specific *N*-glycan determinants with endogenous lectins such as collectins and siglecs with affinities for α-mannose- and sialic acid-terminating *N*-glycans, respectively, during an

immune response. Subcellular-specific *N*-glycosylation is also of high relevance in the context of breast cancer given the intimate relationship between cancer and inflammation in the tumor micro-environment [354]. In breast cancer, hyper-activation of the immune system mediated by the innate and adaptive immune cells including sialic acid-containing leukocytes and lymphocytes promotes tumor development and disease progression [355]. It is worth noting that although the analysis of secreted *N*-glycans (Part1, Chapter 3) showed a high prevalence of more processed complex type *N*-glycan in all investigated cell lines, a significant higher degree of sialylation of the complex *N*-glycans was observed for all breast cancer cells relative to the non-tumorigenic cells thereby suggesting the role of sialic acid in tumorigenesis.

The study presented in Chapter 4 also highlighted the challenges associated with isolating or enriching for the biologically very active proteome subset, the plasma membrane (cell surface) proteins. It importantly concluded that conventional ultracentrifugation methods to isolate the total membrane proteome (microsome) were not an efficient way to isolate plasma membrane proteins. In addition to capturing the cell surface proteins, the microsomal fraction also captured the much larger fraction of membrane proteins from the intracellular organelles such as ER and Golgi residing proteins. These multiple subcellular origins of the membrane protein naturally add severe complexity to the interpretation of their biological activity. In this light, future proteomic and glycomic studies can benefit by first performing a full characterization and optimization of the fractionation process. Nevertheless, to alleviate this issue, cell surface biotinylation was proven to be a highly useful method to specifically enrich this sub-population of proteins from the total cellular membrane proteins.

While modern glycomics profiling techniques have provided a means to map the *N*-glycome in a relative fast and comprehensive manner and provide clues to assess their putative functional roles in cancer progression, a major drawback is the loss of information of the carrier proteins.

System-wide analysis of intact glycoproteins or glycopeptides derived from the glycoproteome with the glycans still attached to the polypeptide carriers captures such structural information which is vital to accurately interpret the biological importance of protein *N*-glycosylation. Glycoproteomics is a term which was coined to describe the system-wide analysis of the glycoproteome in a defined system. Due to the extreme structural heterogeneity of glycopeptides in complex biological mixtures, true glycoproteomics techniques are still in their relative infancy [22] and more development is needed at all levels from sample preparation to LC-MS/MS acquisition and automated data interpretation.

In *Chapter 5*, an affinity chromatography platform utilizing lectins to enrich for tumor-relevant subsets of the glycoproteome from complex protein mixtures was developed and optimized. Lectins show affinities but rarely absolute specificity towards certain glycan determinants/epitopes and may therefore be used in targeted (and biased) approaches to enrich for glycoproteins of interest. Instead of using single lectins to enrich for glycoproteins bearing a particular glycan epitope, multiple lectins combined in an affinity chromatography system (called M-LAC) were used to capture a collection of tumor-relevant glycoproteins carrying a broader range glycan epitopes. Similar M-LAC platforms to enrich for tumorigenic glycoproteins from human serum or plasma have previously been developed where the same combination of lectins were used including Con A, Jac, and WGA to enhance the lectin–glycoprotein interactions via the glycoside cluster effects [356]. Importantly, this strategy of glycoprotein enrichment is compatible with conventional workflows for downstream LC-MS/MS based (glyco)protein identification following deglycosylation or as intact glycopeptides. The basis of Chapter 5 was to optimize the M-LAC platform to enable glycoprotein enrichment from whole cell lysates of breast cancer cell lines, but may also be applied to individual subcellular fractions. One of the technical challenges during this method optimization was lectin leaching or "column bleeding", which was related to the use of low pH mobile phases for analyte elution. To circumvent this

issue, the mode of elution was changed by utilizing saccharide elution mechanisms, which dramatically increased the glyco(protein) recovery and prevented loss of the stationary phase by lectin leaching. A possible future application for the optimized M-LAC glycoprotein enrichment technology is the isolation of membrane glycoproteins using selected lectins targeting the highly tumor-specific subsets of glycoepitopes identified in this thesis (Chapter 5). Isolating tumor-specific (sub)glycoproteomes are valuable for even deeper structural and functional characterization in particular when methodologies for glycoproteomics become more mature enabling system-wide site-specific mapping of protein *N*-glycosylation.

6.2 Future directions

It is important to bear in mind that in large scale "omics" studies, such as those discussed in this thesis, the structural knowledge derived from long lists of identified protein/glycan structures do not necessary yield answers to specific biological questions, but may instead help to formulate more intelligent questions for future studies [357]. A shortcoming of this thesis is that results of identified molecular alterations from the proteomics and glycomics studies were not further investigated using clinically relevant samples, which are essential for the translation of data from *in vitro* models towards the clinic. Hypothesis-driven investigations, based on initial "omics" based observations, are usually required to test the molecular trends or patterns observed on the system-wide level. In the context of this thesis, it would require the detailed investigations on highly specific subsets of the proteome or *N*-glycome or individual proteins such as USP14 or XPO1 or glycans to evaluate their functional roles in breast cancer biology. A key future aim is therefore to verify the observations identified in this thesis using clinically derived samples and establish their possible clinical utility.

Moreover, omics type studies, including those in performed in the thesis, do not usually provide direct evidence whether aberrant glycosylation or deregulated proteome are causing or a result of

cancer; the exact etiology of the disease clearly needs to be addressed in future studies. For example, it would be of interest to determine whether the elevated expression of α2,3-linked sialylation observed in basal subtype of breast cancer cells is a key molecular promoter of metastasis or simply a passive result of the physiological changes associated with the pathology.

The results and conclusions in this thesis were derived solely by observation made from isolated cultured cell lines. However, it is important to stress that a relative large panel of cell lines was used ensuring the inclusion of disease and genetic heterogeneity in mapping the molecular patterns. Nonetheless, it should not be ignore that the exact roles and effector functions of the identified proteins and their glycans in tumorigenesis is highly dependent on their cellular and molecular interactions in the actual tumor micro-environment. Hence, the molecular observations and functional trends described in this work clearly need to be validated in cancer tissue from biopsies or following removal by surgery.

One future direction following on from this body of work will be to bridge the existing and continuing gap between proteomics and glycomics in order to achieve an integrated understanding of the molecular events underlying fundamental aspects of human and in particular disease biology. In this light, the optimized method for lectin affinity chromatography can be further applied to different biological samples in future studies to isolate subsets of intact glycoproteins of interest for further analysis. Glycoproteomics will be pivotal in connecting the fields of proteomics and glycomics. Given this area is currently the subject of intense research interest, glycoproteomics is likely to develop at a rapid pace. Developments are bound to benefit from the significant and continuous advances in LC-MS/MS technologies facilitated by proteomics, such as the advent of high resolution and mass accuracy MS and attention to more sophisticated bioinformatics solutions being integrated into existing and novel workflows.

Specifically, the observations presented herein would benefit from glycoproteomics type investigations at the glycopeptide or ideally at the intact glycoprotein level to identify the integrity of the carrier proteins that display the tumor-associated glycan structures. Glycoproteomics data would also provide validation and possibly further insight into the documented feature of subcellular-specific *N*-glycosylation allowing us to advance our understanding of the fundamental regulatory control mechanisms associated with protein *N*-glycosylation. This, in turn, will enable us gain an even greater appreciation of the important role(s) of protein *N*-glycosylation in tumorigenesis.

6.3 Conclusion

The primary aim of this thesis was to increase our understanding of the molecular changes associated with breast cancer by specifically investigating the proteome and glycome in a global system-wide context of a panel of tumorigenic and non-tumorigenic cultured breast cell lines. These large-scale omics studies revealed that biologically significant proteins and their *N*-glycans are differentially expressed in the secretions and in the membranes of breast tumor cells relative to their healthy non-tumorigenic counterparts. In general the observed molecular alterations agreed well with observations published in the literature thereby confirming the accuracy and the usefulness of the chosen cell line panel. Importantly, a large set of new protein and glycan alterations were identified here due to the analytical depth and the dual molecular focus of this study. The molecular changes were tumor- and cancer subtype-specific; the *N*-glycomes could as such be used to delineate the common breast cancer subtypes. However, whether such cancer subtype-related structures are causing or resulting from breast tumorigenesis needs further assessment by more targeted etiology studies. While mapping these cancer-related molecular patterns, subcellular-specific protein *N*-glycosylation was identified as a more common seemingly universal feature of human cells, not only related to cancer cells. This intriguing feature could mechanistically be explained by the differential solvent accessibility to the *N*-glycosylation site of

proteins localizing to the different subcellular locations of the cell. Mapping the factors driving the spatial and temporal regulation of glycosylated proteins is crucial to build up a solid understanding of the importance of protein *N*-glycosylation in health and disease. Finally, a minor focus of this thesis was dedicated to not only use, but also further develop, a sample preparation tools for proteomics and in particular glycoproteomics. To this end, a multi-lectin glycoprotein enrichment platform was developed and optimized to enable isolation of tumor-specific subsets of the glycoproteome from breast cancer cells. It is anticipated that such isolations and further technology developments within the glycoproteomics field will enable us to get even closer to defining the exact molecular alterations associated with breast cancer. In conclusion, this thesis has provided evidence that deep system-wide molecular characterization of tumor-specific specimen may generate solid knowledge platforms from which potential biomarkers and therapeutics can be developed and from where more targeted biological questions related to the mechanistic understanding the molecular alterations in disease can be defined.

APPENDICES

MACQUARIE UNIVERSITY

Ling Lee <ling.lee@mq.edu.au>

Progress Report 4 Approved - Fanayan (Ref: 5201100040)

Ethics Secretariat <ethics.secretariat@mq.edu.au> Wed, Feb 11, 2015 at 3:03 PM
To: Dr Susan Fanayan <susan.fanayan@mq.edu.au>
Cc: Mr Ling Lee <ling.lee@mq.edu.au>, Miss Manveen Sethi <manveen.sethi@mq.edu.au>

Dear Dr Fanayan,

Title of project: "Identification and analysis of protein glycosylation changes across different cancers involving the breast, gastrointestine and colon of the human body and childhood brain cancer, for the discovery and validation of novel biomarkers for early detection and surveillance" (Ref: 5201100040)

Thank you for your Progress Report. Approval of the Progress Report has been granted, effective 11 February 2015.

Please be advised this approval covers the following period: 16/3/2014 to 11/2/2015.

NB: The following components of the project have now been completed and the students involved have been removed from the project:

Breast Cancer Component (Student: Ling Yen Lee)
Colon Cancer Component (Student: Ms Manveen kaur Sethi)

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports.

NB: A Final Report report is due on 16 March 2016.

If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report on the project.

Progress Reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. Please notify the Committee of any amendment to the project.

5. Please notify the Committee immediately in the event of any adverse effects on participants or of any unforeseen events that might affect

continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at:

http://www.research.mq.edu.au/about/research_@_macquarie/policies,_procedures_and_conduct

Yours sincerely,

Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committees

Office of the Deputy Vice Chancellor (Research)

Ethics Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 2 9850 6848
F: +61 2 9850 4465
http://www.mq.edu.au/research

CRICOS Provider Number 00002J

APPENDIX 1: List of identified secreted proteins

MW = molecular weight; BC = breast cancer; NSAF = normalized spectral abundance factor; Non-Cl = non-classical secretions
TM = transmembrane domain; Nil expression indicated by grey-shaded box

Identified secreted proteins present in all four breast epithelial cell lines										
Identified Proteins	Gene	MW (kDa)	BC database	Exosome	Signal Peptide	TM	Average NSAF			
							HMEC	MCF7	SKBR3	MDA231
14-3-3 protein beta/alpha	YWHAB	28	Yes	Yes	No	No	4.880E-03	5.004E-03	7.144E-03	5.267E-03
14-3-3 protein epsilon	YWHAE	29	Yes	Yes	No	No	5.153E-03	1.843E-02	8.511E-03	1.838E-02
14-3-3 protein eta	YWHAH	28	No	Yes	No	No	3.720E-03	4.165E-03	4.484E-03	3.749E-03
14-3-3 protein gamma	YWHAG	28	No	Yes	No	No	3.895E-03	3.596E-03	5.178E-03	2.660E-03
14-3-3 protein sigma	SFN	28	Yes	Yes	No	No	1.403E-02	4.061E-03	2.045E-03	2.420E-03
14-3-3 protein theta	YWHAQ	28	Yes	Yes	No	No	6.603E-03	4.839E-03	6.102E-03	5.311E-03
14-3-3 protein zeta/delta	YWHAZ	28	Yes	Yes	No	No	8.437E-03	7.806E-03	8.363E-03	7.802E-03
26S proteasome non-ATPase regulatory subunit 2	PSMD2	100	Yes	Yes	No	No	5.723E-05	1.557E-04	2.045E-04	2.351E-04
26S proteasome non-ATPase regulatory subunit 5	PSMD5	56	No	No	Non-Cl	No	8.297E-05	3.685E-04	1.044E-04	3.203E-04
26S proteasome non-ATPase regulatory subunit 6	PSMD6	46	No	Yes	No	No	2.846E-04	5.365E-04	2.391E-04	4.003E-04
3'(2'),5'-bisphosphate nucleotidase 1	BPNT1	33	No	No	Non-Cl	No	1.408E-04	1.232E-04	2.845E-04	3.195E-04
40S ribosomal protein S28	RPS28	8	No	Yes	No	No	1.854E-03	1.903E-03	1.429E-03	9.987E-04
40S ribosomal protein S3	RPS3	27	No	Yes	No	No	2.735E-04	8.003E-04	7.131E-04	7.131E-04
40S ribosomal protein S5	RPS5	23	No	Yes	Non-Cl	No	4.509E-04	4.875E-04	3.185E-04	8.507E-04
40S ribosomal protein SA	RPSA	33	No	Yes	Non-Cl	No	5.171E-04	1.175E-03	9.408E-04	1.055E-03
45 kDa calcium-binding protein	SDF4	42	No	Yes	Yes	No	1.928E-03	2.651E-04	5.800E-04	5.279E-04
4F2 cell-surface antigen heavy chain	SLC3A2	68	Yes	Yes	No	Yes	3.402E-04	1.337E-03	1.152E-03	6.888E-04
60S acidic ribosomal protein P0	RPLP0	34	Yes	Yes	No	No	4.590E-04	6.006E-04	4.029E-04	6.538E-04
60S ribosomal protein L12	RPL12	18	No	Yes	Non-Cl	No	8.684E-04	2.296E-03	5.972E-04	1.857E-03
6-phosphogluconate dehydrogenase, decarboxylating	PGD	53	No	Yes	No	No	6.415E-04	1.169E-03	1.127E-03	9.438E-04
6-phosphogluconolactonase	PGLS	28	No	Yes	Non-Cl	No	4.590E-04	3.305E-04	8.736E-04	8.016E-04
78 kDa glucose-regulated protein	HSPA5	72	No	Yes	Yes	No	9.948E-04	2.648E-03	2.343E-03	2.874E-03
Acetyl-CoA acetyltransferase, cytosolic	ACAT2	41	Yes	Yes	Non-Cl	No	2.401E-04	1.663E-04	5.785E-05	6.534E-04
Actin-related protein 2	ACTR2	45	No	Yes	No	No	8.661E-05	5.428E-04	5.929E-04	4.993E-04
Actin-related protein 2/3 complex subunit 3	ARPC3	21	Yes	Yes	Non-Cl	No	3.944E-04	7.731E-04	1.245E-03	1.875E-03
Actin-related protein 3	ACTR3	47	No	Yes	No	No	2.833E-04	4.908E-04	4.995E-04	4.294E-04
Adenine phosphoribosyltransferase	APRT	20	No	Yes	No	No	5.283E-04	8.953E-04	7.728E-04	4.467E-04
Adenosylhomocysteinase	AHCY	48	Yes	Yes	No	No	8.088E-04	7.677E-04	9.183E-04	1.162E-03
Adenylyl cyclase-associated protein 1	CAP1	52	Yes	Yes	No	No	7.937E-04	3.709E-04	6.265E-04	7.631E-04
ADP-ribosylation factor 1	ARF1	21	No	Yes	No	No	5.290E-04	9.134E-04	1.940E-03	1.565E-03
ADP-sugar pyrophosphatase	NUDT5	24	No	Yes	No	No	3.610E-04	2.639E-03	8.233E-04	2.087E-03
Alcohol dehydrogenase [NADP(+)]	AKR1A1	37	No	Yes	No	No	2.641E-04	1.764E-04	1.338E-03	1.242E-04
Alpha-enolase	ENO1	47	Yes	Yes	Non-Cl	No	1.590E-02	3.019E-02	3.148E-02	3.737E-02
Alpha-mannosidase 2	MAN2A1	131	No	Yes	No	Yes	3.712E-04	1.784E-04	2.818E-04	2.268E-04
Aminoacylase-1	ACY1	46	Yes	Yes	No	No	2.503E-04	7.034E-05	3.832E-04	1.418E-04
Aminopeptidase B	RNPEP	73	Yes	Yes	Non-Cl	No	1.165E-04	2.844E-04	7.883E-04	1.766E-04
Amyloid beta A4 protein	APP	87	Yes	Yes	Yes	Yes	2.571E-03	1.743E-03	2.073E-03	1.937E-03
Amyloid-like protein 2	APLP2	87	Yes	Yes	Yes	Yes	3.006E-04	7.783E-04	1.170E-03	1.079E-03
Annexin A2	ANXA2	39	No	Yes	Non-Cl	No	8.928E-03	5.013E-03	2.696E-03	7.769E-03
Annexin A3	ANXA3	36	Yes	Yes	No	No	1.106E-03	2.056E-04	5.033E-04	4.457E-04
Annexin A4	ANXA4	36	No	Yes	No	No	1.601E-04	3.236E-04	8.230E-04	1.255E-04
Annexin A5	ANXA5	36	No	Yes	Non-Cl	No	9.508E-04	1.779E-03	2.025E-03	3.450E-03
Aspartate aminotransferase, cytoplasmic	GOT1	46	No	Yes	No	No	1.460E-04	4.020E-04	5.614E-04	4.903E-04
Aspartate aminotransferase, mitochondrial	GOT2	48	No	Yes	No	No	1.387E-04	1.430E-04	5.950E-04	4.295E-04
Attractin	ATRN	159	No	Yes	No	Yes	4.595E-04	1.734E-04	2.025E-04	6.124E-05
Basal cell adhesion molecule	BCAM	67	No	Yes	Yes	Yes	8.965E-05	1.510E-03	4.983E-03	1.464E-04
Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	469	Yes	Yes	Yes	No	3.767E-03	4.958E-04	2.922E-04	7.048E-04
Beta-1,4-galactosyltransferase 1	B4GALT1	44	No	Yes	No	Yes	9.121E-04	6.560E-04	4.092E-04	2.457E-04
Beta-2-microglobulin	B2M	14	No	Yes	Yes	No	7.117E-03	3.108E-03	3.155E-03	2.701E-03
Beta-galactosidase	GLB1	76	No	Yes	Yes	No	8.679E-05	8.388E-05	2.488E-04	1.033E-04
Beta-hexosaminidase subunit alpha	HEXA	61	No	No	Yes	No	1.664E-04	3.586E-04	8.565E-04	4.248E-04
Beta-hexosaminidase subunit beta	HEXB	63	Yes	No	Yes	Yes	4.881E-04	2.899E-04	5.223E-04	6.620E-04
Calpain small subunit 1	CAPNS1	28	No	Yes	No	No	2.513E-04	3.882E-04	8.672E-04	1.121E-03
Calreticulin	CALR	48	Yes	Yes	Yes	No	8.861E-04	7.388E-04	9.070E-04	3.044E-04
Carbonyl reductase [NADPH] 1	CBR1	30	Yes	Yes	Non-Cl	No	1.419E-03	5.996E-04	1.275E-03	1.058E-03
Carboxypeptidase E	CPE	53	No	No	Yes	No	1.230E-03	3.651E-03	1.268E-03	8.446E-05
Cathepsin B	CTSB	38	Yes	Yes	Yes	No	2.986E-04	8.352E-05	1.720E-04	2.574E-04
Cathepsin L1	CTSL	38	No	No	Yes	No	1.114E-03	5.570E-04	4.906E-04	3.806E-04
Cathepsin Z	CTSZ	34	Yes	No	Yes	No	1.394E-03	1.102E-04	8.993E-04	6.278E-04
CD166 antigen	ALCAM	65	Yes	Yes	Yes	Yes	4.402E-04	5.016E-04	6.096E-04	6.130E-04
CD59 glycoprotein	CD59	14	No	Yes	Yes	No	2.605E-03	1.003E-03	5.065E-04	9.484E-04
Cell division control protein 42 homolog	CDC42	21	No	Yes	Non-Cl	No	6.131E-04	2.959E-04	5.159E-04	5.129E-04
Chitinase domain-containing protein 1	CHID1	45	No	Yes	Yes	No	1.316E-04	3.382E-04	1.161E-04	2.350E-04
Chloride intracellular channel protein 1	CLIC1	27	Yes	Yes	No	No	1.475E-03	1.237E-03	1.321E-03	2.272E-03
Chloride intracellular channel protein 4	CLIC4	29	Yes	Yes	No	Yes	6.595E-04	6.385E-04	4.338E-04	9.566E-04
Cluster of 60S ribosomal protein L11	RPL11	20	No	Yes	No	No	5.131E-04	2.409E-04	3.238E-04	4.542E-04
Cluster of Actin, cytoplasmic 1	ACTB	42	Yes	Yes	No	No	9.191E-03	6.559E-03	1.372E-02	1.022E-02
Cluster of Alpha-actinin-4	ACTN4	105	No	Yes	No	No	3.318E-03	1.082E-02	8.714E-03	1.053E-02
Cluster of AP-2 complex subunit beta	AP2B1	105	Yes	No	Non-Cl	No	1.172E-04	6.551E-04	3.160E-04	3.269E-04
Cluster of ATP-citrate synthase	ACLY	121	No	Yes	Non-Cl	No	1.084E-04	2.186E-04	2.024E-03	8.807E-04

Cluster of Clathrin heavy chain 1	CLTC	192	Yes	Yes	No	No	6.913E-05	8.675E-04	5.812E-04	1.471E-03
Cluster of Cofilin-1	CFL1	19	No	Yes	Non-Cl	No	4.758E-03	3.517E-03	2.689E-03	4.308E-03
Cluster of Cullin-associated NEDD8-dissociated protein 1	CAND1	136	No	Yes	No	No	6.999E-05	7.392E-04	4.689E-04	4.032E-04
Cluster of Elongation factor 1-alpha 1	EEF1A1	50	No	Yes	No	No	1.062E-03	1.526E-02	1.724E-03	1.252E-02
Cluster of Fibronectin	FN1	263	No	Yes	Yes	No	3.670E-02	4.710E-04	4.039E-03	3.746E-03
Cluster of Fructose-bisphosphate aldolase A	ALDOA	39	No	Yes	No	No	2.909E-03	2.259E-02	2.498E-02	9.442E-03
Cluster of Glucosamine-6-phosphate isomerase 1	GNPDA1	33	No	Yes	No	No	1.767E-04	1.649E-04	3.409E-04	1.610E-04
Cluster of Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	36	Yes	Yes	No	No	7.653E-03	5.384E-03	4.229E-03	7.387E-03
G(I)/G(S)/G(T) subunit beta-1	GNB1	37	No	Yes	No	No	1.781E-04	5.436E-04	9.072E-04	5.412E-04
Cluster of Heat shock 70 kDa protein 1A/1B	HSPA1A	70	Yes	Yes	No	No	5.904E-04	4.686E-03	2.778E-03	2.166E-03
Cluster of Heat shock cognate 71 kDa protein	HSPA8	71	Yes	Yes	No	No	1.734E-03	6.628E-03	4.409E-03	5.026E-03
Cluster of Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	38	No	No	No	No	2.245E-04	6.031E-04	5.088E-04	7.487E-04
Cluster of Latent-transforming growth factor beta-binding protein 1	LTBP1	187	No	Yes	Yes	No	9.188E-04	1.461E-04	1.563E-04	2.703E-05
Cluster of Nucleoside diphosphate kinase A	NME1	17	Yes	Yes	No	No	3.547E-03	5.667E-03	5.481E-03	3.668E-03
Cluster of Plectin	PLEC	532	No	Yes	No	No	1.090E-04	5.048E-04	5.440E-04	5.083E-04
Cluster of Protein FAM49B	FAM49B	37	Yes	Yes	No	No	2.629E-04	6.140E-04	4.489E-04	7.992E-04
Cluster of Putative heat shock protein HSP 90-alpha A5	HSP90AA5P	39	No	No	No	No	4.164E-03	2.889E-02	1.964E-02	2.405E-02
Cluster of Pyruvate kinase PKM	PKM	58	No	No	No	No	1.968E-03	3.345E-03	1.627E-03	3.109E-03
Cluster of Ras-related protein Rab-11B	RAB11B	24	No	Yes	Non-Cl	No	5.301E-04	9.609E-04	1.333E-03	1.061E-03
Cluster of Ras-related protein Rab-14	RAB14	24	No	Yes	No	No	5.005E-04	2.672E-03	4.006E-03	3.378E-03
Cluster of Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	PPP2R1A	65	No	Yes	Non-Cl	No	7.100E-05	5.958E-04	1.651E-04	2.270E-04
Cluster of Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA	36	No	Yes	No	No	1.104E-04	2.311E-04	3.021E-04	4.290E-04
Cluster of Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	PPP3CA	59	No	No	No	No	8.026E-05	5.929E-04	1.123E-04	3.332E-04
Cluster of Spliceosome RNA helicase DDX39B	DDX39B	49	No	No	No	No	1.625E-04	4.103E-04	2.032E-04	5.113E-04
Cluster of Syntenin-1	SDCBP	32	No	Yes	No	Yes	2.807E-04	3.365E-04	5.630E-04	9.070E-04
Cluster of Talin-1	TLN1	270	No	Yes	No	No	4.029E-05	4.347E-04	3.015E-04	9.282E-04
Cluster of Transforming protein RhoA	RHOA	22	Yes	Yes	Non-Cl	No	6.991E-04	7.422E-04	2.262E-03	1.473E-03
Cluster of Transgelin-2	TAGLN2	22	No	Yes	Non-Cl	No	4.470E-03	2.486E-03	4.848E-03	4.335E-03
Cluster of Tropomyosin alpha-4 chain	TPM4	29	No	Yes	No	No	1.355E-03	2.031E-03	1.888E-03	1.760E-03
Cluster of Tubulin alpha-4A chain	TUBA4A	50	No	Yes	Non-Cl	No	4.636E-03	5.154E-03	5.402E-03	8.818E-03
Cluster of Tubulin beta chain	TUBB	50	No	Yes	No	No	5.276E-03	3.494E-03	4.114E-03	8.161E-03
Cluster of Ubiquitin-40S ribosomal protein S27a	RPS27A	18	No	Yes	Non-Cl	No	1.795E-03	9.248E-03	1.463E-02	1.256E-02
Cluster of UTP--glucose-1-phosphate uridylyltransferase	UGP2	57	Yes	Yes	No	No	1.010E-04	2.709E-04	2.744E-04	1.939E-04
Clusterin	CLU	52	Yes	Yes	Yes	No	6.665E-04	4.309E-03	5.574E-03	7.722E-05
Coatomer subunit beta	COPB1	107	No	Yes	No	Yes	5.384E-05	3.993E-04	2.242E-04	2.723E-04
Coatomer subunit delta	ARCN1	57	Yes	No	No	No	2.842E-04	5.211E-04	3.866E-04	5.988E-04
Coatomer subunit epsilon	COPE	34	No	No	No	No	1.658E-04	3.069E-04	2.850E-04	2.063E-04
Coatomer subunit zeta-1	COPZ1	20	No	No	No	No	4.415E-04	6.423E-04	5.836E-04	5.584E-04
Collagen alpha-1(XII) chain	COL12A1	333	No	Yes	Yes	No	1.687E-03	2.442E-04	4.241E-04	5.749E-04
Collagen alpha-1(XVIII) chain	COL18A1	178	No	Yes	Yes	No	1.713E-04	4.652E-04	1.671E-05	2.407E-04
Complement C3	C3	187	No	Yes	Yes	No	1.429E-03	3.145E-05	1.400E-05	1.036E-04
Cystatin-C	CST3	16	Yes	Yes	Yes	No	4.283E-02	2.214E-02	3.073E-03	4.964E-03
Cytosolic non-specific dipeptidase	CNDP2	53	No	Yes	No	No	1.437E-04	3.705E-04	1.295E-03	4.360E-04
Deoxyribonuclease-2-alpha	DNASE2	40	No	No	Yes	No	2.103E-04	1.857E-04	2.583E-04	2.250E-04
Desmoglein-2	DSG2	122	Yes	Yes	Yes	Yes	1.001E-04	1.814E-04	1.922E-04	4.161E-05
Desmoplakin	DSP	332	No	Yes	No	No	2.047E-05	8.633E-05	2.836E-05	1.018E-05
Destrin	DSTN	19	No	Yes	Non-Cl	No	2.849E-03	3.323E-03	5.020E-03	1.443E-03
Disintegrin and metalloproteinase domain-containing protein 10	ADAM10	84	No	Yes	Yes	Yes	2.769E-04	2.773E-04	4.730E-04	2.315E-04
Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	93	No	No	Yes	Yes	1.061E-04	1.092E-04	8.558E-05	4.197E-05
Dystroglycan	DAG1	97	Yes	Yes	Yes	Yes	8.809E-04	1.218E-03	2.283E-03	2.515E-04
EGF-containing fibulin-like extracellular matrix protein 1	EFEMP1	55	No	Yes	Yes	No	9.273E-03	5.228E-04	6.710E-04	4.111E-04
Elongation factor 1-delta	EEF1D	31	Yes	Yes	Non-Cl	No	1.488E-04	1.553E-03	1.143E-03	1.517E-03
Elongation factor 1-gamma	EEF1G	50	No	Yes	No	No	8.316E-04	2.630E-03	2.185E-03	4.171E-03
Elongation factor 2	EEF2	95	Yes	Yes	No	No	8.814E-04	2.021E-03	1.921E-03	1.943E-03
Endoplasmic reticulum aminopeptidase 1	ERAP1	107	No	Yes	Yes	No	3.191E-04	1.465E-04	4.345E-05	2.078E-04
Endoplasmic reticulum resident protein 29	ERP29	29	No	No	Yes	No	3.348E-04	9.813E-04	1.103E-03	1.085E-03
Endoplasmic reticulum resident protein 44	ERP44	47	No	No	Yes	No	1.659E-04	1.873E-04	1.436E-04	1.557E-04
Endoplasmin	HSP90B1	92	No	Yes	Yes	No	7.256E-04	1.058E-03	1.882E-03	1.072E-03
ERO1-like protein alpha	ERO1L	54	Yes	Yes	Yes	No	2.630E-04	1.327E-04	1.733E-04	2.080E-04
Eukaryotic initiation factor 4A-I	EIF4A1	46	Yes	Yes	Non-Cl	No	8.903E-04	8.282E-04	7.875E-04	2.186E-03
Ezrin	EZR	69	No	Yes	Non-Cl	No	6.207E-04	1.199E-03	2.958E-03	1.170E-03
F-actin-capping protein subunit alpha-1	CAPZA1	33	No	Yes	No	No	5.402E-04	8.151E-04	6.868E-04	8.287E-04
Farnesyl pyrophosphate synthase	FDPS	48	Yes	Yes	No	No	6.394E-04	1.459E-04	3.895E-04	7.532E-04
Fatty acid synthase	FASN	273	Yes	Yes	No	No	1.173E-04	1.030E-03	9.623E-03	1.330E-03
Filamin-A	FLNA	281	Yes	Yes	No	No	2.469E-04	1.604E-03	7.322E-04	1.541E-03
Flavin reductase (NADPH)	BLVRB	22	No	Yes	Non-Cl	No	5.736E-04	2.647E-03	1.933E-03	6.993E-04
Fodrin alpha chain	SPTAN1	285	No	Yes	No	No	2.413E-05	5.488E-04	3.526E-04	2.571E-04
Galectin-1	LGALS1	15	Yes	Yes	No	No	6.566E-03	1.204E-03	5.634E-04	1.460E-02
Galectin-3-binding protein	LGALS3BP	65	Yes	Yes	Yes	No	2.541E-03	4.795E-04	1.562E-02	2.527E-03
Gamma-glutamyl hydrolase	GGH	36	Yes	Yes	Yes	No	2.940E-03	6.011E-03	3.711E-03	4.354E-03
Gamma-glutamylcyclotransferase	GGCT	21	No	Yes	No	No	7.491E-04	1.039E-03	1.945E-03	5.220E-04
Gamma-interferon-inducible lysosomal thiol reductase	IFI30	28	Yes	No	Yes	No	1.760E-03	1.888E-04	6.553E-04	3.894E-04
Ganglioside GM2 activator	GM2A	21	Yes	Yes	Yes	No	2.736E-03	1.444E-04	2.667E-04	1.478E-04

Glucose-6-phosphate isomerase	GPI	63	Yes	Yes	No	No	4.353E-04	4.742E-03	5.321E-03	1.716E-03
Glucosidase 2 subunit beta	PRKCSH	59	No	Yes	Yes	No	3.577E-04	4.398E-03	8.360E-04	5.563E-04
Glutathione S-transferase omega-1	GSTO1	28	No	Yes	No	No	1.435E-03	1.006E-03	8.129E-04	1.649E-03
Glutathione synthetase	GSS	52	Yes	Yes	No	No	1.045E-04	1.677E-04	1.701E-04	1.348E-04
Glycogen phosphorylase, brain form	PYGB	97	No	Yes	No	No	1.849E-04	3.085E-04	1.863E-04	4.102E-04
Golgi apparatus protein 1	GLG1	135	No	Yes	Yes	Yes	4.386E-05	7.041E-04	4.931E-05	1.278E-04
Group XV phospholipase A2	PLA2G15	47	No	No	Yes	No	1.657E-04	6.912E-05	8.460E-05	1.112E-04
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	GNG12	8	No	Yes	Non-Cl	No	9.354E-04	5.986E-04	2.584E-03	1.436E-03
Heat shock 70 kDa protein 13	HSPA13	52	No	Yes	Yes	No	2.207E-04	9.103E-05	7.467E-05	3.440E-04
Heat shock 70 kDa protein 4	HSPA4	94	Yes	Yes	No	No	3.459E-04	9.402E-04	5.863E-04	7.772E-04
Heat shock protein beta-1	HSPB1	23	Yes	Yes	No	No	1.968E-03	3.118E-03	2.451E-03	7.335E-04
Heme-binding protein 2	HEBP2	23	Yes	Yes	Non-Cl	No	6.152E-04	5.535E-04	1.182E-03	4.813E-04
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	39	No	Yes	No	No	4.593E-04	1.756E-03	1.753E-03	2.600E-03
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	37	No	Yes	No	No	1.463E-03	2.263E-03	3.256E-03	3.084E-03
Histone H4	HIST1H4A	11	No	Yes	No	No	8.843E-04	4.370E-03	3.159E-03	6.266E-03
Hornerin	HRNR	282	No	Yes	No	No	1.869E-05	1.341E-04	5.314E-05	5.491E-05
Hsp90 co-chaperone Cdc37	CDC37	44	No	Yes	No	No	1.595E-04	5.403E-04	3.226E-04	2.592E-04
Hypoxanthine-guanine phosphoribosyltransferase	HPRT1	25	No	Yes	No	No	9.711E-04	1.310E-03	9.331E-04	2.509E-03
Hypoxia up-regulated protein 1	HYOU1	111	Yes	Yes	Yes	No	7.667E-05	5.564E-04	5.941E-04	2.781E-04
Importin subunit beta-1	KPNB1	97	No	Yes	Non-Cl	No	2.169E-04	6.705E-04	3.958E-04	1.117E-03
Inactive tyrosine-protein kinase 7	PTK7	118	No	Yes	Yes	Yes	4.148E-04	2.481E-04	3.805E-05	1.883E-04
Inorganic pyrophosphatase	PPA1	33	No	Yes	No	No	2.002E-04	8.360E-04	1.767E-03	1.153E-03
Integrin alpha-2	ITGA2	129	Yes	Yes	Yes	Yes	2.813E-05	4.002E-05	4.469E-05	8.778E-05
Interleukin enhancer-binding factor 2	ILF2	43	Yes	No	No	No	1.546E-04	5.580E-04	2.230E-04	2.368E-04
Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	47	No	Yes	No	No	4.212E-04	6.172E-04	7.707E-04	5.856E-04
Isoform 2 of Basigin	BSG	29	No	Yes	Yes	Yes	1.350E-04	5.059E-04	8.430E-04	8.678E-04
Isoform 2 of Calsyntenin-1	CLSTN1	109	No	Yes	Yes	Yes	7.314E-03	1.385E-03	3.752E-03	5.747E-04
Isoform 2 of F-actin-capping protein subunit beta	CAPZB	31	No	Yes	Non-Cl	No	2.408E-04	8.924E-04	9.860E-04	8.290E-04
Isoform 2 of Glyoxalase domain-containing protein 4	GLOD4	33	No	No	Non-Cl	No	4.050E-04	3.496E-04	5.015E-04	9.317E-04
Isoform 2 of Receptor-type tyrosine-protein phosphatase kappa	PTPRK	162	No	Yes	Yes	Yes	2.452E-04	7.409E-04	2.098E-04	5.222E-05
Isoform 2 of Tropomyosin alpha-3 chain	TPM3	29	No	Yes	Non-Cl	No	1.345E-03	2.098E-03	2.509E-03	1.857E-03
Isoform 3 of Heterogeneous nuclear ribonucleoprotein K	HNRNPK	49	No	Yes	No	No	2.757E-04	9.933E-04	2.135E-04	6.846E-04
Isoform 4 of Cytosolic acyl coenzyme A thioester hydrolase	ACOT7	37	Yes	Yes	Non-Cl	No	1.813E-04	1.385E-04	3.373E-04	5.768E-04
Isoform 6 of Agrin	AGRN	215	No	Yes	Yes	No	1.172E-02	2.882E-03	5.231E-03	1.438E-03
Isoform 6 of Poly(rC)-binding protein 2	PCBP2	38	Yes	No	Non-Cl	No	6.681E-04	5.863E-04	5.230E-04	7.903E-04
Isoform 8 of Filamin-B	FLNB	282	Yes	Yes	No	No	4.755E-04	4.318E-04	2.275E-03	1.493E-03
Isoform IIb of Profilin-2	PFN2	15	No	Yes	Non-Cl	No	1.004E-03	1.393E-03	7.491E-04	1.073E-03
Keratin, type I cytoskeletal 9	KRT9	62	No	Yes	No	No	2.518E-03	2.121E-03	2.067E-03	2.137E-03
Keratin, type II cytoskeletal 5	KRT5	62	Yes	Yes	No	No	2.512E-03	3.258E-03	2.539E-03	2.651E-03
Laminin subunit alpha-5	LAMA5	400	Yes	Yes	Yes	No	6.408E-04	9.428E-04	6.189E-05	1.614E-03
Legumain	LGMN	49	No	No	Yes	No	2.551E-04	8.727E-05	3.882E-04	2.680E-04
Leukocyte elastase inhibitor	SERPINB1	43	No	Yes	Non-Cl	No	6.365E-04	3.513E-04	1.238E-04	7.018E-04
Leukotriene A-4 hydrolase	LTA4H	69	No	Yes	No	No	2.305E-04	2.567E-04	4.363E-04	2.350E-04
L-lactate dehydrogenase A chain	LDHA	37	Yes	Yes	No	Yes	1.444E-02	7.005E-03	6.551E-03	1.103E-02
Low-density lipoprotein receptor	LDLR	95	Yes	No	Yes	Yes	1.061E-03	3.356E-05	6.776E-05	1.409E-03
Lysosomal alpha-glucosidase	GAA	105	No	Yes	Yes	Yes	1.864E-04	6.843E-04	4.469E-04	4.346E-04
Lysyl oxidase homolog 2	LOXL2	87	No	Yes	Yes	No	5.422E-04	2.455E-05	8.943E-05	7.293E-04
Macrophage migration inhibitory factor	MIF	12	Yes	Yes	Non-Cl	No	4.612E-03	4.359E-03	6.305E-03	4.392E-03
Malate dehydrogenase, cytoplasmic	MDH1	36	No	Yes	No	No	5.851E-04	1.419E-03	1.410E-03	8.652E-04
Malate dehydrogenase, mitochondrial	MDH2	36	No	Yes	Non-Cl	No	4.506E-04	6.729E-04	2.839E-03	1.653E-03
Metalloproteinase inhibitor 2	TIMP2	24	No	Yes	Yes	No	1.674E-03	6.482E-04	2.132E-03	1.603E-03
Multifunctional protein ADE2	PAICS	47	No	No	No	No	3.822E-04	9.890E-04	6.289E-04	9.238E-04
Myosin light polypeptide 6	MYL6	17	Yes	No	No	No	1.084E-03	1.417E-03	1.302E-03	1.864E-03
Myotrophin	MTPN	13	No	No	No	No	5.895E-04	1.176E-03	1.576E-03	1.468E-03
N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	AGA	37	No	No	Yes	No	5.693E-04	4.360E-04	6.067E-04	6.475E-04
Nascent polypeptide-associated complex subunit alpha	NACA	23	No	Yes	No	No	2.691E-04	1.324E-03	6.749E-04	4.858E-04
Neural cell adhesion molecule L1	L1CAM	140	No	Yes	Yes	Yes	1.510E-04	2.035E-04	3.392E-04	5.038E-04
Neuropilin-1	NRP1	72	No	Yes	Yes	Yes	2.749E-04	4.430E-04	4.260E-04	1.690E-03
Neutral alpha-glucosidase AB	GANAB	107	No	Yes	Yes	Yes	2.875E-04	5.877E-03	7.897E-04	7.802E-04
Nodal modulator 3	NOMO3	134	No	Yes	Yes	Yes	6.215E-05	2.401E-04	9.039E-05	3.297E-05
Nucleobindin-1	NUCB1	54	No	Yes	Yes	No	2.843E-03	5.777E-04	5.186E-04	5.501E-04
Nucleolin	NCL	77	Yes	Yes	No	No	5.763E-05	3.203E-04	2.471E-04	3.583E-04
Omega-amidase NIT2	NIT2	31	No	No	Non-Cl	No	3.445E-04	3.790E-04	2.715E-04	2.519E-04
Palmitoyl-protein thioesterase 1	PPT1	34	Yes	No	Yes	No	3.618E-04	2.352E-04	5.933E-04	3.204E-04
Peptidyl-prolyl cis-trans isomerase A	PPIA	18	Yes	Yes	No	No	6.416E-03	6.472E-03	1.409E-02	1.022E-02
Peptidyl-prolyl cis-trans isomerase B	PPIB	24	No	Yes	No	No	4.801E-03	3.831E-03	1.105E-03	3.314E-03
Peptidyl-prolyl cis-trans isomerase FKBP1A	FKBP1A	12	No	Yes	No	No	1.262E-03	8.346E-04	2.098E-03	1.763E-03
Peptidyl-prolyl cis-trans isomerase FKBP4	FKBP4	52	No	Yes	No	No	1.433E-04	1.000E-03	9.388E-04	1.775E-04
Peroxidasin homolog	PXDN	165	No	Yes	Yes	No	7.384E-04	9.360E-04	7.264E-04	8.106E-04
Peroxiredoxin-4	PRDX4	31	Yes	Yes	Yes	No	3.850E-04	5.050E-04	3.829E-03	1.338E-03
Peroxiredoxin-5, mitochondrial	PRDX5	17	No	Yes	Non-Cl	No	1.237E-03	6.848E-04	8.960E-04	5.656E-04
Peroxiredoxin-6	PRDX6	25	No	Yes	No	No	1.454E-03	3.405E-03	1.552E-03	2.236E-03
Phosphatidylethanolamine-binding protein 1	PEBP1	21	No	Yes	Non-Cl	No	1.365E-03	2.780E-03	2.407E-03	3.016E-03
Phosphoglucomutase-1	PGM1	61	No	Yes	No	No	1.817E-04	1.773E-04	6.513E-04	5.654E-04
Phosphoglycerate kinase 1	PGK1	45	Yes	Yes	No	No	5.852E-04	6.985E-03	3.728E-03	4.378E-03

Phosphoglycerate mutase 1	PGAM1	29	No	Yes	No	No	1.098E-03	2.570E-03	3.318E-03	5.390E-03
Plasma alpha-L-fucosidase	FUCA2	54	No	Yes	Yes	No	4.241E-04	9.382E-04	5.023E-05	5.569E-04
Plexin-B2	PLXNB2	205	No	Yes	Yes	Yes	4.069E-05	3.099E-05	3.479E-04	5.358E-05
Poly(rC)-binding protein 1	PCBP1	37	Yes	Yes	Non-Cl	No	8.958E-04	6.799E-04	1.027E-03	1.166E-03
Polyadenylate-binding protein 1	PABPC1	71	No	No	No	No	2.845E-04	1.003E-03	4.588E-04	5.454E-04
Prelamin-A/C	LMNA	74	Yes	Yes	No	No	2.933E-04	8.842E-04	3.166E-03	2.723E-03
Proactivator polypeptide	PSAP	58	No	Yes	Yes	No	2.687E-03	2.184E-04	2.451E-03	1.114E-03
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	PLOD1	84	Yes	Yes	Yes	No	4.896E-04	3.011E-04	1.131E-03	6.508E-04
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2	85	No	Yes	Yes	No	1.398E-04	1.100E-04	9.434E-05	2.901E-04
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3	85	No	Yes	Yes	No	4.063E-04	2.456E-03	9.359E-04	3.897E-04
Profilin-1	PFN1	15	Yes	Yes	No	No	4.639E-02	2.773E-02	1.814E-02	4.263E-02
Programmed cell death 6-interacting protein	PDCD6IP	96	No	Yes	No	No	1.299E-04	4.148E-04	5.948E-04	4.905E-04
Proliferation-associated protein 2G4	PA2G4	44	No	Yes	No	No	8.661E-05	4.564E-04	3.828E-04	4.694E-04
Prolyl endopeptidase	PREP	81	No	No	Non-Cl	No	1.196E-04	8.693E-05	2.955E-04	2.440E-04
Prostaglandin E synthase 3	PTGES3	19	No	No	Non-Cl	No	6.040E-04	5.613E-04	9.845E-04	1.406E-03
Proteasome subunit alpha type-5	PSMA5	26	No	Yes	No	No	2.693E-04	4.758E-04	5.299E-04	5.016E-04
Proteasome subunit alpha type-6	PSMA6	27	No	Yes	No	No	7.381E-04	8.085E-04	1.359E-03	7.718E-04
Proteasome subunit alpha type-7	PSMA7	28	No	Yes	No	No	7.098E-04	1.435E-03	2.196E-03	1.698E-03
Proteasome subunit beta type-1	PSMB1	26	No	Yes	No	No	3.626E-04	9.416E-04	1.296E-03	9.820E-04
Proteasome subunit beta type-2	PSMB2	23	Yes	Yes	No	No	5.191E-04	8.037E-04	1.126E-03	7.747E-04
Proteasome subunit beta type-5	PSMB5	28	Yes	Yes	Non-Cl	No	3.080E-04	1.316E-03	1.496E-03	1.010E-03
Proteasome subunit beta type-6	PSMB6	25	Yes	Yes	Non-Cl	No	3.232E-04	1.771E-04	3.223E-04	2.666E-04
Protein disulfide-isomerase	P4HB	57	Yes	Yes	Yes	No	1.360E-03	8.702E-04	1.537E-03	3.625E-04
Protein disulfide-isomerase A3	PDIA3	57	No	Yes	Yes	No	1.848E-03	1.709E-03	2.807E-03	1.633E-03
Protein disulfide-isomerase A4	PDIA4	73	Yes	Yes	Yes	No	1.979E-04	5.298E-04	7.739E-04	2.765E-04
Protein disulfide-isomerase A6	PDIA6	48	No	Yes	Yes	No	5.537E-04	5.915E-04	5.312E-04	7.842E-04
Protein DJ-1	PARK7	20	Yes	Yes	No	No	2.886E-03	1.282E-03	3.022E-03	1.952E-03
Protein FAM3C	FAM3C	25	No	No	Yes	No	5.630E-03	1.264E-04	1.535E-03	1.270E-03
Protein S100-A11	S100A11	12	Yes	Yes	Non-Cl	No	2.193E-03	1.981E-03	2.599E-03	1.643E-03
Protein S100-A16	S100A16	12	No	Yes	Non-Cl	No	1.691E-03	3.029E-03	9.639E-03	1.113E-03
Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	25	No	No	Non-Cl	No	4.762E-04	9.208E-04	1.360E-03	1.217E-03
Purine nucleoside phosphorylase	PNP	32	No	No	No	No	8.526E-04	1.184E-03	1.668E-03	1.433E-03
Puromycin-sensitive aminopeptidase	NPEPPS	103	No	Yes	No	No	1.409E-04	4.081E-04	3.047E-04	3.572E-04
Rab GDP dissociation inhibitor alpha	GDI1	51	Yes	Yes	No	No	1.714E-04	4.032E-04	4.424E-04	2.323E-04
Rab GDP dissociation inhibitor beta	GDI2	51	Yes	Yes	No	No	4.209E-04	1.326E-03	1.425E-03	8.192E-04
Ras GTPase-activating-like protein IQGAP1	IQGAP1	189	No	Yes	No	No	4.716E-05	9.642E-04	5.865E-04	3.815E-04
Ras-related protein Rab-5C	RAB5C	23	No	Yes	Non-Cl	No	4.626E-04	6.386E-04	4.672E-03	1.841E-03
Receptor-type tyrosine-protein phosphatase F	PTPRF	213	Yes	Yes	Yes	Yes	8.711E-05	1.796E-03	2.327E-03	1.669E-04
Ribonuclease 4	RNASE4	17	No	No	Yes	No	2.393E-03	1.217E-03	2.594E-03	4.068E-03
Ribonuclease inhibitor	RNH1	50	No	Yes	Non-Cl	No	4.591E-04	3.704E-04	3.649E-04	4.259E-04
Ribonuclease T2	RNASET2	29	No	No	Yes	No	6.908E-04	2.984E-04	6.860E-04	6.539E-04
Roundabout homolog 1	ROBO1	181	No	No	Yes	Yes	2.627E-05	1.646E-04	8.128E-05	4.572E-05
S-adenosylmethionine synthase isoform type-2	MAT2A	44	No	No	Non-Cl	No	2.001E-04	1.798E-04	2.070E-04	2.042E-04
Semaphorin-7A	SEMA7A	75	No	Yes	Yes	Yes	8.589E-05	7.129E-05	1.900E-04	2.508E-04
Serpin B6	SERPINB6	43	Yes	Yes	No	No	1.791E-04	3.913E-04	9.270E-05	3.845E-04
Sialate O-acetylesterase	SIAE	58	No	Yes	Yes	No	1.123E-04	4.990E-04	4.428E-04	1.546E-04
Single-stranded DNA-binding protein, mitochondrial	SSBP1	17	No	Yes	No	No	5.341E-04	3.504E-04	9.420E-04	7.005E-04
Soluble calcium-activated nucleotidase 1	CANT1	45	No	No	No	Yes	6.800E-04	3.915E-04	1.182E-03	2.879E-04
Stanniocalcin-1	STC1	28	Yes	Yes	Yes	No	4.820E-04	2.571E-03	1.174E-04	3.739E-04
Sulfhydryl oxidase 1	QSOX1	83	No	Yes	Yes	Yes	1.059E-03	2.249E-03	2.480E-03	1.387E-03
Superoxide dismutase [Cu-Zn]	SOD1	16	Yes	Yes	No	No	7.790E-04	1.397E-03	2.190E-03	1.962E-03
Synaptic vesicle membrane protein VAT-1 homolog	VAT1	42	No	Yes	No	No	1.104E-04	1.697E-04	8.869E-05	4.847E-04
Syndecan-4	SDC4	22	Yes	Yes	Yes	Yes	2.106E-03	3.668E-03	1.704E-03	1.611E-03
T-complex protein 1 subunit theta	CCT8	60	No	Yes	No	No	1.257E-04	9.084E-04	7.508E-04	6.628E-04
Thioredoxin domain-containing protein 17	TXNDC17	14	No	Yes	No	No	1.045E-03	1.240E-02	3.947E-03	1.718E-02
Thioredoxin domain-containing protein 5	TXNDC5	48	No	Yes	Yes	No	3.296E-04	2.409E-04	4.302E-04	3.739E-04
Thrombospondin-1	THBS1	129	Yes	Yes	Yes	No	1.329E-02	9.128E-03	2.712E-03	1.169E-02
Transaldolase	TALDO1	38	No	Yes	No	No	4.383E-04	6.186E-04	1.223E-03	8.872E-04
Transferrin receptor protein 1	TFRC	85	Yes	Yes	Yes	Yes	8.862E-05	6.007E-04	3.037E-03	1.237E-03
Transforming growth factor-beta-induced protein ig-h3	TGFBI	75	No	Yes	Yes	No	2.929E-03	3.064E-04	3.901E-04	9.282E-04
Transitional endoplasmic reticulum ATPase	VCP	89	No	Yes	No	No	3.386E-04	9.540E-04	7.267E-04	1.022E-03
Transketolase	TKT	68	No	Yes	No	No	5.845E-04	1.113E-03	2.198E-03	1.580E-03
Transmembrane protein 132A	TMEM132A	110	No	No	Yes	Yes	1.248E-04	3.202E-04	5.884E-04	9.061E-05
Triosephosphate isomerase	TPI1	31	Yes	Yes	Non-Cl	No	7.784E-03	1.857E-02	2.494E-02	1.421E-02
Tripeptidyl-peptidase 1	TPP1	61	No	Yes	Yes	No	3.218E-04	4.211E-05	2.592E-04	2.150E-04
Ubiquitin-conjugating enzyme E2 N	UBE2N	17	Yes	Yes	Non-Cl	No	8.700E-04	6.193E-04	7.684E-04	8.709E-04
Ubiquitin-conjugating enzyme E2 variant 2	UBE2V2	16	No	Yes	Non-Cl	No	1.041E-03	1.248E-03	1.613E-03	3.067E-03
Ubiquitin-like modifier-activating enzyme 1	UBA1	118	No	Yes	Non-Cl	No	3.146E-04	9.895E-04	9.615E-04	5.443E-04
Vacuolar protein sorting-associated protein 29	VPS29	21	No	No	No	No	4.379E-04	3.129E-04	5.404E-04	3.164E-04
Vasorin	VASN	72	No	Yes	Yes	Yes	7.915E-05	7.239E-04	9.688E-04	2.736E-04
Vesicular integral-membrane protein VIP36	LMAN2	40	No	Yes	Yes	Yes	1.079E-03	5.052E-04	2.973E-04	5.508E-04
Vinculin	VCL	124	No	Yes	No	No	3.050E-04	5.219E-04	1.451E-03	8.482E-04
WD repeat-containing protein 1	WDR1	66	No	Yes	Non-Cl	No	1.391E-04	5.799E-04	4.687E-04	6.951E-04
Xylosyltransferase 2	XYLT2	97	No	No	Yes	Yes	7.815E-05	2.037E-04	1.352E-04	8.656E-05
Identified secreted proteins present breast cancer cell lines only										
Identified Proteins	**Gene**	**MW**	**BC database**	**Exosome**	**Signal Peptide**	**TM**	**Average NSAF**			
							HMEC	**MCF7**	**SKBR3**	**MDA231**
10 kDa heat shock protein, mitochondrial	HSPE1	11	No	Yes	Non-Cl	No		1.248E-03	4.302E-03	6.279E-03

116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	109	No	No	No	No		1.028E-04	7.172E-05	1.551E-04
26S protease regulatory subunit 10B	PSMC6	44	No	Yes	Non-Cl	No		2.196E-04	8.960E-05	2.236E-04
26S protease regulatory subunit 6A	PSMC3	49	No	No	No	No		3.348E-04	1.821E-04	3.027E-04
26S protease regulatory subunit 7	PSMC2	49	No	Yes	Non-Cl	No		9.827E-05	9.443E-05	1.999E-04
26S protease regulatory subunit 8	PSMC5	46	No	Yes	No	No		7.068E-05	9.828E-05	2.700E-04
26S proteasome non-ATPase regulatory subunit 11	PSMD11	47	No	Yes	No	No		2.152E-04	2.200E-04	2.879E-04
26S proteasome non-ATPase regulatory subunit 13	PSMD13	43	No	Yes	Non-Cl	No		2.274E-04	1.862E-04	2.764E-04
26S proteasome non-ATPase regulatory subunit 3	PSMD3	61	No	Yes	No	No		2.143E-04	1.635E-04	1.954E-04
40S ribosomal protein S21	RPS21	9	No	Yes	Non-Cl	No		1.253E-03	1.201E-03	1.319E-03
4-trimethylaminobutyraldehyde dehydrogenase	ALDH9A1	54	Yes	Yes	No	No		1.341E-04	1.530E-04	1.050E-04
5'-nucleotidase domain-containing protein 1	NT5DC1	52	No	No	Non-Cl	No		1.678E-04	1.532E-04	7.659E-05
60 kDa heat shock protein, mitochondrial	HSPD1	61	No	Yes	No	No		2.161E-04	6.083E-05	2.232E-04
60 kDa SS-A/Ro ribonucleoprotein	TROVE2	61	No	No	No	No		7.970E-05	4.328E-05	7.473E-05
7,8-dihydro-8-oxoguanine triphosphatase	NUDT1	23	Yes	No	No	No		2.149E-04	3.241E-04	2.664E-04
Acidic leucine-rich nuclear phosphoprotein 32 family member B	ANP32B	29	No	Yes	No	No		6.992E-04	5.036E-04	2.305E-04
Actin-like protein 6A	ACTL6A	47	No	Yes	No	No		7.711E-05	9.375E-05	1.623E-04
Actin-related protein 2/3 complex subunit 1B	ARPC1B	41	Yes	Yes	No	No		2.423E-04	2.963E-04	2.466E-04
Actin-related protein 2/3 complex subunit 2	ARPC2	34	No	Yes	No	No		5.219E-04	5.824E-04	6.499E-04
Actin-related protein 2/3 complex subunit 4	ARPC4	20	No	Yes	Non-Cl	No		1.015E-03	1.022E-03	8.240E-04
Actin-related protein 2/3 complex subunit 5	ARPC5	16	Yes	Yes	Non-Cl	No		5.958E-04	5.415E-04	7.255E-04
Activator of 90 kDa heat shock protein ATPase homolog 1	AHSA1	38	No	Yes	No	No		2.236E-04	3.618E-04	1.729E-04
Acylamino-acid-releasing enzyme	APEH	81	Yes	Yes	No	Yes		1.688E-04	4.039E-04	3.898E-05
Adenylate kinase 2, mitochondrial	AK2	26	Yes	Yes	Non-Cl	No		1.398E-04	7.838E-04	2.883E-04
Adenylate kinase isoenzyme 1	AK1	22	No	Yes	No	No		3.966E-04	6.028E-04	1.341E-03
Adenylosuccinate synthetase isozyme 2	ADSS	50	No	No	No	No		1.033E-04	1.793E-04	3.413E-04
Aflatoxin B1 aldehyde reductase member 2	AKR7A2	40	Yes	Yes	Non-Cl	No		2.509E-04	2.604E-04	2.897E-04
Alanine--tRNA ligase, cytoplasmic	AARS	107	Yes	No	No	No		8.303E-04	1.496E-04	1.489E-04
Alpha/beta hydrolase domain-containing protein 14B	ABHD14B	22	No	Yes	Non-Cl	No		4.519E-04	1.895E-03	1.291E-03
Alpha-2-HS-glycoprotein	AHSG	39	No	Yes	Yes	No		1.698E-03	2.454E-03	2.843E-03
Alpha-galactosidase A	GLA	49	Yes	No	Yes	No		5.541E-05	8.125E-05	8.062E-05
Alpha-parvin	PARVA	42	No	No	Non-Cl	No		1.150E-04	1.235E-04	1.720E-04
Alpha-soluble NSF attachment protein	NAPA	33	No	Yes	Non-Cl	No		2.103E-04	4.344E-04	2.738E-04
Anamorsin	CIAPIN1	34	No	No	Non-Cl	No		1.369E-04	2.996E-04	2.406E-04
Annexin A6	ANXA6	76	No	Yes	No	No		8.487E-05	7.954E-04	2.142E-04
Arginine--tRNA ligase, cytoplasmic	RARS	75	No	Yes	No	No		3.612E-04	9.749E-05	1.313E-04
Asparagine synthetase [glutamine-hydrolyzing]	ASNS	64	Yes	Yes	No	No		2.794E-04	9.400E-05	2.996E-04
Asparagine--tRNA ligase, cytoplasmic	NARS	63	No	Yes	No	No		7.754E-05	4.249E-05	8.410E-05
Aspartate--tRNA ligase, cytoplasmic	DARS	57	No	Yes	No	No		4.744E-05	4.682E-05	1.971E-04
Aspartyl aminopeptidase	DNPEP	52	No	Yes	No	No		3.967E-04	4.536E-04	9.739E-05
ATP-dependent RNA helicase A	DHX9	141	Yes	Yes	No	No		1.671E-04	1.282E-04	2.056E-04
Bifunctional glutamate/proline--tRNA ligase	EPRS	171	Yes	Yes	No	No		7.258E-05	6.517E-05	6.514E-05
Bifunctional purine biosynthesis protein PURH	ATIC	65	No	Yes	No	No		7.003E-04	4.504E-04	3.029E-04
Biliverdin reductase A	BLVRA	33	Yes	Yes	Non-Cl	No		3.706E-04	5.475E-04	3.301E-04
BolA-like protein 2	BOLA2	10	No	No	Non-Cl	No		1.049E-03	1.288E-03	1.406E-03
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	102	Yes	Yes	No	No		1.472E-04	1.801E-04	1.484E-04
CAD protein	CAD	243	Yes	Yes	No	No		2.816E-05	9.369E-05	1.168E-04
Calcyclin-binding protein	CACYBP	26	No	Yes	Non-Cl	No		1.084E-03	7.138E-04	1.543E-03
Calpain-1 catalytic subunit	CAPN1	82	No	Yes	Non-Cl	No		2.681E-04	2.518E-04	9.834E-05
Calsyntenin-3	CLSTN3	106	No	No	Yes	Yes		7.892E-04	5.517E-04	7.896E-05
Catalase	CAT	60	No	Yes	No	No		6.308E-05	8.929E-05	1.421E-04
Cation-independent mannose-6-phosphate receptor	IGF2R	274	No	Yes	Yes	Yes		5.353E-05	2.643E-04	7.103E-04
Citrate synthase, mitochondrial	CS	52	Yes	Yes	No	No		7.133E-05	1.747E-04	1.620E-04
Cleavage and polyadenylation specificity factor subunit 5	NUDT21	26	No	No	No	No		4.625E-04	1.801E-04	2.285E-04
Cluster of Alpha-2-macroglobulin	A2M	163	No	No	Yes	No		5.067E-05	3.052E-04	4.607E-04
Cluster of AP-2 complex subunit alpha-1	AP2A1	108	No	Yes	No	No		2.327E-04	1.064E-04	1.008E-04
Cluster of Chromobox protein homolog 3	CBX3	21	Yes	No	Non-Cl	No		1.352E-03	1.030E-03	7.569E-04
Cluster of Coatomer subunit gamma-1	COPG1	98	No	No	No	No		3.311E-04	1.468E-04	1.659E-04
Cluster of Cytoplasmic FMR1-interacting protein 1	CYFIP1	145	Yes	Yes	No	No		3.096E-04	5.305E-04	1.158E-04
Cluster of Disco-interacting protein 2 homolog B	DIP2B	171	No	Yes	Non-Cl	No		7.579E-05	9.986E-05	1.855E-05
Cluster of Drebrin-like protein	DBNL	48	No	Yes	Non-Cl	No		3.956E-04	1.030E-03	1.206E-04
Cluster of Dynein light chain 1, cytoplasmic	DYNLL1	10	No	No	No	No		2.648E-03	1.395E-03	2.198E-03
Cluster of Ephrin type-A receptor 4	EPHA4	110	No	No	Yes	Yes		2.733E-04	1.467E-04	1.871E-04
Cluster of GTPase NRas	NRAS	21	Yes	Yes	No	No		1.078E-03	1.188E-03	6.716E-04
Cluster of Guanine nucleotide-binding protein G(i) subunit alpha-2	GNAI2	40	No	Yes	No	No		2.015E-04	1.789E-04	4.874E-04
Cluster of Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	49	No	No	No	No		5.394E-04	3.378E-04	5.280E-04
Cluster of Isoform 2 of Protein SET	SET	32	No	Yes	No	No		1.815E-04	3.207E-04	2.773E-04
ribonucleoprotein A/B	HNRNPAB	31	No	No	No	No		7.714E-04	4.730E-04	5.219E-04
Cluster of Isoform 5 of Thioredoxin reductase 1, cytoplasmic	TXNRD1	55	Yes	No	No	No		6.522E-05	4.578E-04	2.928E-04
Cluster of Isoform VEGF183 of Vascular endothelial growth factor A	VEGFA	24	No	No	Yes	No		3.434E-04	3.818E-04	5.687E-04
Cluster of Mitogen-activated protein kinase 1	MAPK1	41	No	No	No	No		1.830E-04	1.778E-04	2.880E-04
Cluster of Nuclease-sensitive element-binding protein 1	YBX1	36	No	Yes	Non-Cl	No		1.129E-03	3.397E-04	7.518E-04
Cluster of Nucleolysin TIAR	TIAL1	42	Yes	No	No	No		1.014E-04	1.389E-04	1.236E-04
Cluster of Plasma membrane calcium-transporting ATPase 1	ATP2B1	139	No	Yes	No	Yes		1.056E-04	4.624E-05	4.577E-05

Cluster of Ras-related C3 botulinum toxin substrate 1	RAC1	21	No	Yes	No	No		5.895E-04	9.354E-04	8.120E-04
Cluster of Ras-related protein Rab-2A	RAB2A	24	No	Yes	No	No		5.598E-04	5.263E-04	1.413E-03
Cluster of Ras-related protein Rab-6A	RAB6A	24	No	Yes	Non-Cl	No		4.128E-04	1.002E-03	8.609E-04
Cluster of Ras-related protein Ral-A	RALA	24	No	Yes	No	No		4.357E-04	1.440E-03	1.973E-04
Cluster of Ras-related protein Rap-1A	RAP1A	21	No	No	No	No		1.166E-03	1.787E-03	1.228E-03
Cluster of RNA-binding motif protein, X chromosome	RBMX	42	No	No	No	No		8.744E-04	4.032E-04	5.875E-04
Cluster of Septin-11	SEPT11	49	No	No	No	No		1.439E-04	8.239E-05	1.479E-04
Cluster of Serine/threonine-protein phosphatase PP1-beta catalytic subunit	PPP1CB	37	No	Yes	Non-Cl	No		9.378E-04	9.238E-04	5.992E-04
Cluster of Small ubiquitin-related modifier 2	SUMO2	11	No	No	Non-Cl	No		3.038E-04	7.954E-04	5.452E-04
Cluster of Spectrin beta chain, non-erythrocytic 1	SPTBN1	275	No	Yes	No	No		2.494E-04	1.349E-04	2.132E-04
Cluster of Tumor protein D54	TPD52L2	22	No	No	Non-Cl	No		1.374E-04	3.629E-04	5.876E-04
Coatomer subunit alpha	COPA	138	No	Yes	No	No		2.164E-04	2.544E-04	9.904E-05
Coatomer subunit beta'	COPB2	102	Yes	Yes	No	No		2.936E-04	4.559E-04	2.480E-04
Collagen alpha-1(V) chain	COL5A1	184	Yes	Yes	Yes	No		1.521E-04	5.069E-05	1.317E-04
COP9 signalosome complex subunit 1	GPS1	56	No	No	No	No		1.450E-04	1.055E-04	1.178E-04
COP9 signalosome complex subunit 2	COPS2	52	No	No	No	No		2.142E-04	6.506E-05	1.692E-04
COP9 signalosome complex subunit 3	COPS3	48	No	Yes	Non-Cl	No		2.600E-04	1.089E-04	1.367E-04
COP9 signalosome complex subunit 4	COPS4	46	No	Yes	Non-Cl	No		3.736E-04	2.595E-04	2.270E-04
COP9 signalosome complex subunit 8	COPS8	23	No	Yes	No	No		1.362E-04	2.236E-04	2.213E-04
Coronin-1B	CORO1B	54	No	Yes	No	No		2.318E-04	3.444E-04	1.307E-04
Cullin-2	CUL2	87	Yes	No	No	No		2.551E-05	3.125E-05	1.072E-04
Cyclin-dependent kinase 2	CDK2	34	Yes	No	No	No		1.108E-04	7.813E-05	1.337E-04
Cystatin-B	CSTB	11	Yes	Yes	No	No		1.502E-03	1.966E-03	2.288E-03
Cysteine and glycine-rich protein 1	CSRP1	21	No	Yes	No	No		3.215E-04	1.097E-03	4.144E-04
Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	532	No	Yes	No	No		9.806E-05	1.624E-04	2.319E-04
Cytosol aminopeptidase	LAP3	56	No	Yes	No	No		7.238E-05	2.704E-04	1.886E-04
D-dopachrome decarboxylase	DDT	13	No	Yes	No	No		5.489E-03	1.287E-02	5.770E-03
Dipeptidyl peptidase 2	DPP7	54	No	No	Yes	No		4.864E-05	1.896E-04	8.172E-05
Dipeptidyl peptidase 3	DPP3	83	No	Yes	No	No		3.958E-04	1.578E-04	7.086E-05
Disintegrin and metalloproteinase domain-containing protein 9	ADAM9	72	No	Yes	Yes	Yes		4.032E-05	6.384E-05	3.537E-05
DNA damage-binding protein 1	DDB1	127	Yes	Yes	Non-Cl	No		1.915E-04	2.637E-04	1.665E-04
DNA replication licensing factor MCM2	MCM2	102	Yes	No	No	No		3.150E-05	3.223E-05	6.312E-05
DNA replication licensing factor MCM6	MCM6	93	Yes	No	No	No		1.212E-04	7.734E-05	8.425E-05
DNA-(apurinic or apyrimidinic site) lyase	APEX1	36	Yes	No	Non-Cl	No		1.343E-04	1.096E-04	4.331E-04
DnaJ homolog subfamily A member 1	DNAJA1	45	No	Yes	No	No		8.373E-05	1.177E-04	1.296E-04
Dynein light chain roadblock-type 1	DYNLRB1	11	No	No	Non-Cl	No		5.003E-04	7.838E-04	5.395E-04
E3 ubiquitin-protein ligase UBR4	UBR4	574	No	No	No	Yes		8.315E-06	6.664E-06	4.254E-05
Early endosome antigen 1	EEA1	162	No	Yes	No	No		1.244E-04	1.161E-04	6.986E-05
EF-hand domain-containing protein D2	EFHD2	27	No	No	No	No		1.771E-04	4.588E-04	4.349E-04
EH domain-containing protein 1	EHD1	61	No	Yes	No	No		2.116E-04	8.805E-04	3.051E-04
EH domain-containing protein 4	EHD4	61	No	Yes	No	No		7.854E-05	4.055E-04	8.438E-05
ELAV-like protein 1	ELAVL1	36	Yes	No	No	No		3.039E-04	3.177E-04	3.896E-04
Elongation factor 1-beta	EEF1B2	25	No	No	Non-Cl	No		1.032E-03	6.667E-04	1.181E-03
Enhancer of rudimentary homolog	ERH	12	No	No	Non-Cl	No		1.001E-03	1.693E-03	1.156E-03
Enolase-phosphatase E1	ENOPH1	29	No	No	Non-Cl	No		2.545E-04	2.888E-04	1.100E-04
Enoyl-CoA delta isomerase 1, mitochondrial	ECI1	33	No	No	Non-Cl	No		1.578E-04	3.801E-04	2.093E-04
Epididymis-specific alpha-mannosidase	MAN2B2	114	No	No	Yes	No		3.783E-05	4.584E-05	6.229E-05
Erythrocyte band 7 integral membrane protein	STOM	32	No	Yes	No	Yes		1.833E-04	1.809E-04	1.201E-04
Ethanolamine-phosphate cytidylyltransferase	PCYT2	44	No	No	No	No		4.927E-05	1.653E-04	1.027E-04
Eukaryotic initiation factor 4A-II	EIF4A2	46	No	Yes	No	No		3.617E-04	4.855E-04	1.219E-03
Eukaryotic initiation factor 4A-III	EIF4A3	47	No	Yes	Non-Cl	No		1.723E-04	2.402E-04	2.945E-04
Eukaryotic peptide chain release factor subunit 1	ETF1	49	No	No	No	No		4.669E-04	2.646E-04	3.837E-04
Eukaryotic translation initiation factor 2 subunit 1	EIF2S1	36	No	Yes	No	No		2.556E-04	1.468E-04	3.475E-04
Eukaryotic translation initiation factor 3 subunit B	EIF3B	92	No	Yes	No	No		2.268E-04	1.347E-04	1.830E-04
Eukaryotic translation initiation factor 3 subunit I	EIF3I	37	No	Yes	Non-Cl	No		4.076E-04	2.871E-04	3.733E-04
Eukaryotic translation initiation factor 3 subunit K	EIF3K	25	No	No	Non-Cl	No		2.196E-04	1.599E-04	2.122E-04
Eukaryotic translation initiation factor 4B	EIF4B	69	No	No	Non-Cl	No		2.163E-04	1.042E-04	7.571E-05
Eukaryotic translation initiation factor 4H	EIF4H	27	No	Yes	Non-Cl	No		1.164E-04	1.858E-04	1.869E-04
Eukaryotic translation initiation factor 5A-1-like	EIF5AL1	17	No	No	No	No		2.924E-03	1.187E-03	4.920E-03
Exosome complex component RRP41	EXOSC4	26	No	No	No	No		1.339E-04	9.503E-05	9.494E-05
Exportin-1	XPO1	123	No	Yes	Non-Cl	No		1.287E-04	1.673E-04	3.785E-04
Exportin-2	CSE1L	110	Yes	Yes	No	No		1.902E-04	2.587E-04	7.935E-04
Exportin-5	XPO5	136	No	No	No	No		1.560E-05	2.880E-05	1.910E-05
Exportin-T	XPOT	110	No	No	Non-Cl	No		1.622E-04	1.692E-04	1.556E-04
F-actin-capping protein subunit alpha-2	CAPZA2	33	No	Yes	No	No		2.683E-04	5.056E-04	1.817E-04
Far upstream element-binding protein 1	FUBP1	68	No	No	No	No		5.312E-04	1.004E-04	1.878E-04
Fructose-bisphosphate aldolase C	ALDOC	39	No	Yes	No	No		1.239E-03	1.749E-03	3.665E-04
Fumarylacetoacetase	FAH	46	Yes	Yes	Non-Cl	No		3.161E-04	5.439E-04	3.570E-04
GDP-L-fucose synthase	TSTA3	36	Yes	Yes	No	No		5.920E-05	1.441E-04	1.080E-04
General vesicular transport factor p115	USO1	108	No	No	No	No		6.409E-05	7.855E-05	6.624E-05
Glucose-6-phosphate 1-dehydrogenase	G6PD	59	No	Yes	No	No		1.658E-03	1.418E-03	6.041E-04
Glucosylceramidase	GBA	60	No	Yes	Yes	Yes		9.695E-05	2.714E-04	1.729E-04
Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 1	GFPT1	79	No	Yes	No	No		1.972E-04	5.894E-05	1.324E-04
Glutamine--tRNA ligase	QARS	88	No	No	No	No		1.226E-04	8.280E-05	2.164E-04
Glutaredoxin-3	GLRX3	37	No	No	No	No		3.117E-04	1.891E-04	2.602E-04

Glycine--tRNA ligase	GARS	83	Yes	Yes	Yes	No		3.203E-04	4.071E-04	3.601E-04
Glycolipid transfer protein	GLTP	24	No	No	No	No		2.942E-04	3.842E-04	4.948E-04
GMP synthase [glutamine-hydrolyzing]	GMPS	77	No	No	No	No		3.421E-05	1.030E-04	8.360E-05
Golgi membrane protein 1	GOLM1	45	No	No	Yes	Yes		8.234E-05	4.049E-04	1.138E-04
Golgin subfamily A member 7	GOLGA7	16	No	Yes	Non-CI	No		2.786E-04	3.340E-04	2.095E-04
Growth factor receptor-bound protein 2	GRB2	25	Yes	Yes	No	No		6.331E-04	7.372E-04	3.718E-04
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	35	Yes	Yes	No	No		3.749E-04	3.700E-04	6.690E-04
Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase MESH1	HDDC3	20	No	No	Non-CI	No		2.652E-04	2.594E-04	2.584E-04
HD domain-containing protein 2	HDDC2	23	No	No	Non-CI	No		1.860E-04	3.130E-04	1.700E-04
Heat shock protein 105 kDa	HSPH1	97	Yes	Yes	No	No		3.817E-04	3.861E-04	6.714E-04
Heat shock protein 75 kDa, mitochondrial	TRAP1	80	Yes	Yes	Non-CI	No		4.037E-04	3.704E-04	4.666E-04
Heme-binding protein 1	HEBP1	21	Yes	Yes	No	No		3.506E-04	3.404E-04	5.494E-04
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	46	No	Yes	No	No		3.093E-04	5.271E-04	4.050E-04
Heterogeneous nuclear ribonucleoprotein L	HNRNPL	64	No	Yes	No	No		1.047E-04	5.971E-05	1.277E-04
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	70	Yes	No	No	No		3.132E-04	6.484E-05	2.604E-04
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	34	Yes	Yes	No	No		1.752E-03	8.325E-04	1.486E-03
Hexokinase-1	HK1	102	Yes	No	No	No		8.265E-05	2.529E-04	8.816E-05
Histidine--tRNA ligase, cytoplasmic	HARS	57	Yes	No	No	No		3.171E-04	2.870E-04	1.600E-04
Histone H2B type F-S	H2BFS	14	No	No	No	No		1.213E-03	6.033E-04	2.435E-03
Hsc70-interacting protein	ST13	41	Yes	Yes	Non-CI	No		4.504E-04	9.493E-05	1.101E-04
Hsp70-binding protein 1	HSPBP1	39	No	No	No	No		1.563E-04	1.125E-04	1.421E-04
Immunoglobulin superfamily member 8	IGSF8	65	No	Yes	Yes	Yes		9.192E-05	1.031E-03	3.738E-05
Importin subunit alpha-1	KPNA2	58	No	No	Non-CI	No		2.073E-04	2.515E-04	2.617E-04
Importin-5	IPO5	124	No	Yes	Non-CI	No		1.910E-04	3.168E-04	7.932E-04
Importin-7	IPO7	120	No	No	No	No		4.384E-04	2.413E-04	6.510E-04
Importin-9	IPO9	116	No	No	No	No		3.651E-05	3.301E-05	4.942E-05
Inhibin beta B chain	INHBB	45	No	Yes	Yes	No		1.171E-04	1.725E-04	7.031E-05
Integrin alpha-V	ITGAV	116	No	Yes	Yes	Yes		2.996E-04	1.330E-04	2.219E-05
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	106	No	Yes	Yes	No		1.309E-04	1.908E-04	4.077E-04
Interleukin enhancer-binding factor 3	ILF3	95	Yes	No	No	No		3.071E-04	1.053E-04	1.735E-04
Interleukin-1 receptor accessory protein	IL1RAP	65	No	Yes	Yes	Yes		9.155E-05	1.020E-04	6.068E-05
Isoamyl acetate-hydrolyzing esterase 1 homolog	IAH1	28	No	No	No	No		9.650E-05	1.878E-04	2.332E-04
Isoform 2 of Apoptosis inhibitor 5	API5	57	No	No	No	No		2.072E-04	1.215E-04	7.673E-05
Isoform 2 of Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	DUT	18	No	Yes	No	No		8.451E-04	1.635E-04	7.106E-04
Isoform 3 of Cysteine--tRNA ligase, cytoplasmic	CARS	95	Yes	No	No	No		1.401E-04	3.070E-05	2.219E-04
Isoform 3 of Dynactin subunit 1	DCTN1	137	Yes	Yes	No	No		1.372E-04	4.535E-05	5.866E-05
Isoform 3 of Seizure 6-like protein 2	SEZ6L2	92	No	No	Yes	Yes		3.836E-04	4.420E-05	5.739E-05
Isoform 3 of Unconventional myosin-Ic	MYO1C	120	No	Yes	No	No		1.031E-04	7.869E-04	1.471E-04
Isoform 7 of Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	155	No	Yes	No	No		5.675E-05	6.497E-05	5.034E-05
subunit alpha isoforms short	GNAS	44	Yes	Yes	Non-CI	No		5.446E-04	5.177E-04	7.352E-05
Junction plakoglobin	JUP	82	No	Yes	No	No		5.289E-04	2.586E-04	8.530E-05
Keratin, type I cytoskeletal 18	KRT18	48	Yes	Yes	No	No		3.053E-02	3.700E-03	4.436E-04
Keratin, type II cytoskeletal 8	KRT8	54	Yes	Yes	No	No		2.130E-02	8.903E-03	1.791E-03
Kinectin	KTN1	156	No	No	No	Yes		1.012E-04	1.248E-04	7.283E-05
Kynureninase	KYNU	52	Yes	No	No	No		2.087E-03	8.861E-05	1.123E-04
Lamina-associated polypeptide 2, isoform alpha	TMPO	75	No	No	No	No		7.520E-05	1.004E-04	1.333E-04
Lamin-B1	LMNB1	66	No	No	No	No		2.271E-04	3.571E-04	2.668E-04
LanC-like protein 1	LANCL1	45	Yes	No	Non-CI	No		7.137E-05	2.906E-04	7.172E-05
Large neutral amino acids transporter small subunit 1	SLC7A5	55	Yes	Yes	No	Yes		5.054E-04	5.580E-04	1.922E-04
Leucyl-cystinyl aminopeptidase	LNPEP	117	No	No	No	Yes		5.615E-05	1.919E-04	1.187E-04
LIM and SH3 domain protein 1	LASP1	30	Yes	No	Non-CI	No		7.837E-04	1.069E-03	1.790E-03
Lupus La protein	SSB	47	No	No	No	No		1.664E-03	2.151E-04	5.660E-04
L-xylulose reductase	DCXR	26	No	Yes	Non-CI	No		4.830E-04	6.441E-04	9.533E-05
Lysine--tRNA ligase	KARS	68	Yes	No	No	No		4.532E-04	4.940E-05	4.837E-05
Lysosomal protective protein	CTSA	54	No	Yes	Yes	No		1.990E-04	7.726E-04	4.802E-04
Lysosomal Pro-X carboxypeptidase	PRCP	56	No	Yes	Yes	No		2.113E-04	5.743E-04	2.895E-04
m7GpppX diphosphatase	DCPS	39	Yes	No	Non-CI	No		1.133E-04	1.749E-04	6.799E-05
Macrophage-capping protein	CAPG	38	Yes	Yes	No	No		3.552E-04	7.179E-04	1.922E-03
Meteorin-like protein	METRNL	34	No	Yes	Yes	No		6.700E-04	5.456E-04	1.468E-04
Methionine adenosyltransferase 2 subunit beta	MAT2B	38	No	No	Non-CI	No		1.130E-04	8.629E-05	8.646E-05
Methylosome protein 50	WDR77	37	No	No	Non-CI	No		2.083E-04	3.229E-04	2.364E-04
Mitotic checkpoint protein BUB3	BUB3	37	No	Yes	Non-CI	No		1.731E-04	1.604E-04	2.291E-04
Myristoylated alanine-rich C-kinase substrate	MARCKS	32	No	Yes	No	No		1.866E-04	4.190E-04	4.698E-04
N(G),N(G)-dimethylarginine dimethylaminohydrolase 1	DDAH1	31	No	Yes	No	No		1.491E-04	8.972E-04	5.670E-04
N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	DDAH2	30	No	Yes	No	No		2.490E-04	1.411E-04	1.223E-04
Na(+)/H(+) exchange regulatory cofactor NHE-RF1	SLC9A3R1	39	Yes	Yes	No	No		1.234E-03	1.313E-03	1.125E-04
N-acetyl-D-glucosamine kinase	NAGK	37	No	Yes	Non-CI	No		4.723E-04	3.404E-04	1.511E-04
N-acetylgalactosaminyltransferase 7	GALNT7	75	No	No	No	Yes		6.409E-05	3.571E-05	1.316E-04
N-acetylglucosamine-6-sulfatase	GNS	62	Yes	No	Yes	No		4.335E-05	1.687E-04	4.488E-04
Nardilysin	NRD1	132	No	No	No	No		9.891E-05	1.208E-04	1.062E-04
Nck-associated protein 1	NCKAP1	129	No	Yes	Non-CI	No		1.263E-04	1.750E-04	3.097E-05
NEDD8-conjugating enzyme Ubc12	UBE2M	21	No	No	Non-CI	No		1.825E-04	1.282E-04	3.761E-04
Neogenin	NEO1	160	No	No	Yes	Yes		7.128E-04	3.015E-04	1.107E-04
Neuroblast differentiation-associated protein AHNAK	AHNAK	629	No	Yes	Non-CI	No		9.097E-05	1.242E-04	2.371E-05
Neurogenic locus notch homolog protein 2	NOTCH2	265	No	No	Yes	Yes		4.787E-05	8.956E-05	3.303E-05

Neutral amino acid transporter B(0)	SLC1A5	57	No	Yes	No	Yes		1.654E-03	9.104E-04	4.386E-04
NHP2-like protein 1	NHP2L1	14	No	No	No	No		3.384E-04	3.194E-04	3.175E-04
Niban-like protein 1	FAM129B	84	No	Yes	No	No		3.225E-04	2.478E-04	1.461E-04
NSFL1 cofactor p47	NSFL1C	41	No	No	No	No		7.712E-05	2.201E-04	9.441E-05
Nuclear migration protein nudC	NUDC	38	No	No	No	No		6.740E-04	4.050E-04	5.031E-04
Nuclear mitotic apparatus protein 1	NUMA1	238	Yes	No	No	No		2.036E-04	2.860E-04	1.173E-04
Nucleoprotein TPR	TPR	267	No	No	No	No		8.456E-05	1.327E-04	4.885E-05
Nucleosome assembly protein 1-like 1	NAP1L1	45	Yes	No	No	No		2.669E-04	7.452E-05	5.312E-04
NudC domain-containing protein 1	NUDCD1	67	Yes	No	No	No		1.139E-04	1.291E-04	1.192E-04
Obg-like ATPase 1	OLA1	45	No	Yes	No	No		8.363E-04	5.429E-04	5.832E-04
PDZ and LIM domain protein 1	PDLIM1	36	Yes	No	No	No		1.724E-04	2.668E-04	6.321E-04
Peroxiredoxin-2	PRDX2	22	No	Yes	Non-Cl	No		2.123E-03	1.891E-03	9.323E-04
Phosphatidylinositol transfer protein beta isoform	PITPNB	32	No	No	No	No		1.037E-04	2.368E-04	2.122E-04
Phosphoacetylglucosamine mutase	PGM3	60	Yes	No	Non-Cl	No		1.831E-04	1.196E-04	1.598E-04
Phosphoglucomutase-2	PGM2	68	No	No	No	No		1.389E-04	1.889E-04	1.507E-04
Phosphoribosylformylglycinamidine synthase	PFAS	145	No	Yes	No	No		3.941E-05	1.479E-04	2.082E-04
Phosphoserine phosphatase	PSPH	25	Yes	No	No	No		1.468E-04	1.795E-04	2.805E-04
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	45	No	Yes	No	No		4.759E-04	2.602E-04	3.935E-04
Platelet-activating factor acetylhydrolase IB subunit alpha	PAFAH1B1	47	No	Yes	No	No		3.483E-04	2.828E-04	4.226E-04
Platelet-activating factor acetylhydrolase IB subunit gamma	PAFAH1B3	26	No	No	No	No		1.662E-04	1.988E-04	2.242E-04
Podocalyxin	PODXL	59	No	Yes	Yes	Yes		6.840E-05	6.246E-05	1.556E-04
Poly(ADP-ribose) glycohydrolase ARH3	ADPRHL2	39	No	No	No	No		1.187E-04	2.894E-04	2.544E-04
Polypeptide N-acetylgalactosaminyltransferase 10	GALNT10	69	Yes	No	Yes	Yes		4.786E-05	1.835E-04	1.438E-04
Polypyrimidine tract-binding protein 1	PTBP1	57	No	Yes	No	No		1.973E-04	3.092E-04	4.321E-04
Prefoldin subunit 2	PFDN2	17	Yes	No	Non-Cl	No		5.227E-04	5.615E-04	6.344E-04
Prefoldin subunit 3	VBP1	23	No	No	Non-Cl	No		3.122E-04	1.785E-04	2.656E-04
Probable ATP-dependent RNA helicase DDX6	DDX6	54	No	No	No	No		3.460E-04	4.820E-05	1.797E-04
Probable ubiquitin carboxyl-terminal hydrolase FAF-X	USP9X	292	No	No	No	No		2.985E-05	5.172E-05	2.032E-05
Pro-cathepsin H	CTSH	37	No	No	Yes	No		9.233E-04	4.478E-04	3.807E-04
Programmed cell death protein 5	PDCD5	14	Yes	Yes	Non-Cl	No		7.167E-04	6.045E-04	7.816E-04
Programmed cell death protein 6	PDCD6	22	No	Yes	Non-Cl	No		1.729E-04	8.500E-04	3.305E-04
Proliferating cell nuclear antigen	PCNA	29	Yes	Yes	Non-Cl	No		5.455E-04	5.093E-04	6.397E-04
Proteasome activator complex subunit 1	PSME1	29	No	No	No	No		1.742E-03	1.332E-03	6.973E-04
Proteasome activator complex subunit 2	PSME2	27	Yes	Yes	No	No		1.445E-03	1.311E-03	7.717E-04
Proteasome activator complex subunit 3	PSME3	30	No	No	No	No		3.730E-04	3.186E-04	7.925E-04
Proteasome assembly chaperone 3	PSMG3	13	No	No	Non-Cl	No		2.316E-04	2.377E-04	4.288E-04
Proteasome subunit alpha type-1	PSMA1	30	No	Yes	No	No		1.246E-03	1.038E-03	7.453E-04
Proteasome subunit alpha type-2	PSMA2	26	No	Yes	No	No		9.356E-04	1.885E-03	1.032E-03
Proteasome subunit alpha type-3	PSMA3	28	Yes	Yes	No	No		1.357E-03	3.263E-03	9.016E-04
Proteasome subunit alpha type-4	PSMA4	29	Yes	Yes	No	No		1.097E-03	1.108E-03	7.315E-04
Proteasome subunit beta type-3	PSMB3	23	No	Yes	No	No		1.070E-03	5.964E-04	3.944E-04
Proteasome subunit beta type-4	PSMB4	29	Yes	Yes	Non-Cl	No		4.854E-04	6.559E-04	4.169E-04
Proteasome subunit beta type-7	PSMB7	30	No	Yes	No	No		2.212E-04	3.986E-04	4.785E-04
Proteasome-associated protein ECM29 homolog	KIAA0368	204	No	No	No	No		6.429E-05	5.042E-05	1.154E-04
Protein arginine N-methyltransferase 1	PRMT1	42	No	No	No	No		1.223E-03	5.638E-04	6.706E-04
Protein diaphanous homolog 1	DIAPH1	141	No	No	No	No		9.312E-05	7.785E-05	7.770E-05
Protein NDRG1	NDRG1	43	Yes	Yes	Non-Cl	No		2.278E-04	1.494E-03	8.778E-05
Protein RCC2	RCC2	56	No	Yes	No	No		2.098E-04	1.223E-04	2.102E-04
Protein transport protein Sec23A	SEC23A	86	No	No	Non-Cl	No		2.120E-04	2.280E-04	3.160E-04
Protein transport protein Sec23B	SEC23B	86	No	No	Non-Cl	No		2.907E-04	4.707E-04	2.477E-04
Protein transport protein Sec24C	SEC24C	118	No	No	No	No		6.938E-05	1.437E-04	4.742E-05
Protein transport protein Sec24D	SEC24D	113	No	No	No	No		1.841E-05	2.810E-05	2.220E-05
Pterin-4-alpha-carbinolamine dehydratase	PCBD1	12	No	No	No	No		5.439E-04	2.187E-03	6.618E-04
Putative deoxyribonuclease TATDN1	TATDN1	34	No	No	No	No		4.025E-04	2.544E-04	7.744E-05
Putative GTP cyclohydrolase 1 type 2 NIF3L1	NIF3L1	42	No	No	No	No		1.374E-04	2.782E-04	1.984E-04
Putative phospholipase B-like 2	PLBD2	65	No	No	Yes	No		4.036E-05	2.259E-04	2.651E-04
Pyridoxal kinase	PDXK	35	No	Yes	Non-Cl	No		3.209E-04	5.601E-04	7.371E-05
Radixin	RDX	69	Yes	Yes	No	No		9.295E-04	1.543E-03	9.770E-04
Ran GTPase-activating protein 1	RANGAP1	64	No	No	No	No		1.051E-04	5.854E-05	1.383E-04
Ras GTPase-activating protein-binding protein 1	G3BP1	52	No	No	No	No		9.119E-05	9.925E-05	2.225E-04
Ras-related protein Rab-21	RAB21	24	No	Yes	Non-Cl	No		2.319E-04	3.829E-04	5.918E-04
Ras-related protein Rab-7a	RAB7A	23	No	Yes	No	No		8.270E-04	7.607E-04	1.087E-03
Receptor-type tyrosine-protein phosphatase S	PTPRS	217	No	No	Yes	Yes		1.245E-04	7.767E-05	8.349E-05
Rho GDP-dissociation inhibitor 1	ARHGDIA	23	No	Yes	No	No		9.315E-04	1.398E-03	1.664E-03
Rho-associated protein kinase 2	ROCK2	161	No	Yes	No	No		1.051E-04	7.450E-05	4.629E-05
RuvB-like 1	RUVBL1	50	No	Yes	No	No		2.702E-04	2.804E-04	3.421E-04
RuvB-like 2	RUVBL2	51	No	Yes	No	No		1.627E-04	2.261E-04	2.499E-04
Secernin-1	SCRN1	46	No	No	Non-Cl	No		4.942E-04	4.032E-04	3.486E-04
Sepiapterin reductase	SPR	28	No	Yes	Non-Cl	No		1.622E-04	1.342E-04	1.998E-04
Serine/threonine-protein kinase OSR1	OXSR1	58	No	Yes	No	No		4.468E-05	1.558E-04	6.480E-05
Serine/threonine-protein phosphatase 2A activator	PPP2R4	41	No	No	No	No		2.255E-04	2.274E-04	2.079E-04
Serine/threonine-protein phosphatase 5	PPP5C	57	No	No	No	No		1.996E-04	2.333E-04	1.951E-04
Serine-threonine kinase receptor-associated protein	STRAP	38	Yes	No	No	No		1.757E-04	1.163E-04	2.470E-04
Serine--tRNA ligase, cytoplasmic	SARS	59	No	Yes	No	No		5.344E-04	7.893E-05	2.811E-04
Sialic acid synthase	NANS	40	No	No	Non-Cl	No		4.218E-04	6.642E-04	1.464E-04
Signal recognition particle 9 kDa protein	SRP9	10	No	No	Non-Cl	No		5.521E-04	7.539E-04	5.349E-04

Small glutamine-rich tetratricopeptide repeat-containing protein alpha	SGTA	34	No	No	Non-Cl	No		1.820E-04	3.715E-04	4.232E-04
Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	113	No	Yes	No	Yes		4.786E-04	1.498E-03	2.929E-04
Sodium/potassium-transporting ATPase subunit beta-3	ATP1B3	32	Yes	Yes	No	Yes		5.606E-04	5.628E-04	1.240E-04
Sorbitol dehydrogenase	SORD	38	Yes	Yes	Non-Cl	No		2.790E-04	2.947E-04	2.577E-04
Sortilin	SORT1	92	No	Yes	Yes	Yes		1.117E-03	1.680E-03	2.009E-04
Spermidine synthase	SRM	34	No	No	No	No		3.458E-04	4.832E-04	3.250E-04
S-phase kinase-associated protein 1	SKP1	19	No	No	Non-Cl	No		3.215E-04	2.887E-04	2.106E-04
Splicing factor 3B subunit 3	SF3B3	136	No	No	Non-Cl	No		2.693E-04	2.597E-04	1.796E-04
Staphylococcal nuclease domain-containing protein 1	SND1	102	No	Yes	No	No		1.360E-04	3.775E-04	1.961E-04
Stathmin	STMN1	17	Yes	Yes	Non-Cl	No		5.362E-04	7.029E-04	1.588E-03
Stress-70 protein, mitochondrial	HSPA9	74	No	Yes	No	No		9.815E-05	1.025E-04	3.820E-04
Stress-induced-phosphoprotein 1	STIP1	63	No	Yes	No	No		8.541E-04	2.550E-04	5.414E-04
SUMO-activating enzyme subunit 1	SAE1	38	Yes	No	No	No		1.800E-04	1.666E-04	1.324E-04
SUMO-conjugating enzyme UBC9	UBE2I	18	No	No	Non-Cl	No		6.006E-04	5.187E-04	3.683E-04
Suppressor of G2 allele of SKP1 homolog	SUGT1	41	No	No	No	No		1.566E-04	2.231E-04	3.963E-04
Tax1-binding protein 3	TAX1BP3	14	No	Yes	Non-Cl	No		4.977E-04	5.189E-04	6.038E-04
T-complex protein 1 subunit alpha	TCP1	60	Yes	Yes	Non-Cl	No		3.763E-04	2.724E-04	4.056E-04
T-complex protein 1 subunit beta	CCT2	57	Yes	Yes	No	No		8.721E-04	4.590E-04	7.892E-04
T-complex protein 1 subunit delta	CCT4	58	Yes	Yes	Non-Cl	No		4.928E-04	2.151E-04	5.674E-04
T-complex protein 1 subunit epsilon	CCT5	60	Yes	Yes	No	No		9.100E-04	3.558E-04	5.862E-04
T-complex protein 1 subunit eta	CCT7	59	No	Yes	No	No		2.710E-04	1.176E-04	3.291E-04
T-complex protein 1 subunit gamma	CCT3	61	No	Yes	No	No		5.279E-04	3.084E-04	8.053E-04
T-complex protein 1 subunit zeta	CCT6A	58	Yes	Yes	No	No		1.341E-04	9.820E-05	2.605E-04
Testin	TES	48	Yes	No	No	No		4.514E-05	8.608E-04	2.208E-04
Thimet oligopeptidase	THOP1	79	No	No	No	No		6.167E-05	1.610E-04	1.180E-04
Thioredoxin	TXN	12	Yes	Yes	No	No		4.317E-03	3.126E-03	5.944E-03
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	28	Yes	Yes	Non-Cl	No		1.121E-04	5.420E-04	7.187E-04
Thioredoxin-like protein 1	TXNL1	32	No	No	No	No		2.300E-04	3.415E-04	5.179E-04
Threonine--tRNA ligase, cytoplasmic	TARS	83	Yes	Yes	No	No		3.740E-04	3.129E-04	9.721E-04
THUMP domain-containing protein 1	THUMPD1	39	No	No	No	No		1.210E-04	1.158E-04	1.308E-04
TIP41-like protein	TIPRL	31	No	No	No	No		3.662E-04	3.210E-04	4.447E-04
Toll-interacting protein	TOLLIP	30	No	Yes	No	No		1.393E-04	1.491E-04	8.489E-05
Transcription elongation factor B polypeptide 1	TCEB1	12	Yes	Yes	No	No		5.525E-04	7.326E-04	6.693E-04
Transcription elongation factor B polypeptide 2	TCEB2	13	Yes	No	Non-Cl	No		9.634E-04	8.820E-04	8.339E-04
Translationally-controlled tumor protein	TPT1	20	Yes	No	Non-Cl	No		7.714E-04	1.279E-03	1.395E-03
Translin	TSN	26	Yes	Yes	No	No		7.683E-04	8.457E-04	6.580E-04
Transportin-1	TNPO1	102	No	No	No	No		7.453E-05	8.524E-05	1.674E-04
Trifunctional purine biosynthetic protein adenosine-3	GART	108	Yes	Yes	Non-Cl	No		5.661E-05	7.981E-05	1.597E-04
Tryptophan--tRNA ligase, cytoplasmic	WARS	53	Yes	Yes	No	No		3.651E-04	4.981E-05	3.429E-04
Tubulin-folding cofactor B	TBCB	27	No	No	Non-Cl	No		3.492E-04	3.578E-04	4.022E-04
Tubulin-specific chaperone A	TBCA	13	No	Yes	No	No		1.626E-03	9.786E-04	1.441E-03
Twinfilin-1	TWF1	40	No	No	Non-Cl	No		3.112E-04	3.336E-04	2.483E-04
Tyrosine-protein phosphatase non-receptor type 11	PTPN11	68	No	No	No	No		1.746E-04	1.653E-04	1.456E-04
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	245	No	No	No	No		3.574E-05	4.102E-05	5.716E-05
Ubiquitin carboxyl-terminal hydrolase 14	USP14	56	Yes	No	Non-Cl	No		4.700E-04	2.608E-04	3.030E-04
Ubiquitin carboxyl-terminal hydrolase 5	USP5	96	No	No	No	No		2.710E-04	3.320E-04	1.752E-04
Ubiquitin carboxyl-terminal hydrolase 7	USP7	128	No	No	No	No		1.164E-04	1.058E-04	6.826E-05
Ubiquitin carboxyl-terminal hydrolase isozyme L3	UCHL3	26	Yes	No	Non-Cl	No		2.071E-04	6.019E-04	1.007E-03
Ubiquitin carboxyl-terminal hydrolase isozyme L5	UCHL5	38	Yes	No	No	No		1.448E-04	7.130E-05	1.406E-04
Ubiquitin fusion degradation protein 1 homolog	UFD1L	35	Yes	No	Non-Cl	No		1.545E-04	5.268E-04	3.951E-04
Ubiquitin-conjugating enzyme E2 K	UBE2K	22	No	No	Non-Cl	No		3.559E-04	5.521E-04	5.754E-04
Ubiquitin-conjugating enzyme E2 variant 1	UBE2V1	16	No	No	Non-Cl	No		7.428E-04	1.342E-03	1.487E-03
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	177	No	No	No	Yes		1.127E-04	2.955E-04	2.958E-05
UMP-CMP kinase	CMPK1	22	No	Yes	Non-Cl	No		2.654E-04	2.122E-03	1.940E-03
Uncharacterized protein C9orf142	C9orf142	22	No	No	Non-Cl	No		2.072E-04	2.865E-04	2.268E-04
UPF0160 protein MYG1, mitochondrial	C12orf10	42	No	No	No	No		2.024E-04	2.174E-04	1.699E-04
UV excision repair protein RAD23 homolog B	RAD23B	43	No	No	No	No		2.794E-04	2.576E-04	1.133E-04
Vacuolar protein sorting-associated protein 26A	VPS26A	38	No	No	No	No		1.160E-04	3.377E-04	1.769E-04
Vacuolar protein sorting-associated protein 26B	VPS26B	39	No	No	No	No		9.876E-05	3.607E-04	6.923E-05
Vacuolar protein sorting-associated protein 35	VPS35	92	No	Yes	No	No		1.421E-04	2.339E-04	1.155E-04
Valine--tRNA ligase	VARS	140	Yes	No	No	No		6.759E-05	9.210E-05	7.340E-05
Vasodilator-stimulated phosphoprotein	VASP	40	No	Yes	No	No		3.859E-04	1.504E-04	6.988E-04
Vesicle-associated membrane protein-associated protein A	VAPA	28	No	Yes	No	Yes		5.146E-04	4.887E-04	5.556E-04
X-ray repair cross-complementing protein 5	XRCC5	83	No	Yes	No	No		2.206E-04	3.580E-04	3.695E-04
X-ray repair cross-complementing protein 6	XRCC6	70	No	Yes	No	No		2.327E-04	3.332E-04	2.381E-04
Zinc finger protein ZPR1	ZPR1	51	No	No	No	No		7.207E-05	8.800E-05	8.854E-05

Identified secreted proteins present in HMEC only

Identified Proteins	Gene	MW	BC database	Exosome	Signal Peptide	TM	Average NSAF			
							HMEC	MCF7	SKBR3	MDA231
72 kDa type IV collagenase	MMP2	74	Yes	Yes	Yes	No	5.369E-04			
Aldehyde dehydrogenase family 1 member A3	ALDH1A3	56	Yes	No	No	No	2.239E-04			
Amphiregulin	AREG	28	Yes	Yes	Yes	Yes	7.469E-03			
Annexin A8	ANXA8	37	No	Yes	No	No	1.203E-03			
Antileukoproteinase	SLPI	14	Yes	Yes	Yes	No	8.646E-03			
Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase	GCNT1	50	No	No	Yes	Yes	2.809E-04			

Bone morphogenetic protein 1	BMP1	111	No	No	Yes	No	6.489E-04			
Cadherin-13	CDH13	78	No	Yes	Yes	No	4.526E-04			
Cadherin-3	CDH3	91	Yes	No	Yes	Yes	2.827E-04			
Calmodulin-like protein 3	CALML3	17	No	Yes	Non-Cl	No	6.173E-04			
Carboxypeptidase B2	CPB2	48	No	No	Yes	No	1.192E-04			
Ceroid-lipofuscinosis neuronal protein 5	CLN5	41	No	No	Yes	No	1.838E-04			
Cluster of C-X-C motif chemokine 3	CXCL3	11	Yes	No	Yes	No	8.427E-03			
Cluster of Laminin subunit alpha-3	LAMA3	367	Yes	Yes	Yes	No	5.326E-03			
Cluster of Metallothionein-2	MT2A	6	Yes	No	No	No	2.631E-03			
Cluster of Serum amyloid A-2 protein	SAA2	14	No	Yes	Yes	No	6.979E-03			
Cocaine esterase	CES2	62	No	No	Yes	No	8.696E-05			
Collagen alpha-1(XVII) chain	COL17A1	150	Yes	No	No	Yes	4.039E-04			
Collagen alpha-2(V) chain	COL5A2	145	Yes	Yes	Yes	No	5.563E-05			
Complement factor I	CFI	66	No	Yes	Yes	No	6.575E-04			
CUB domain-containing protein 1	CDCP1	93	No	No	Yes	Yes	5.002E-05			
C-X-C motif chemokine 10	CXCL10	11	Yes	No	Yes	No	7.098E-04			
Desmocollin-3	DSC3	100	Yes	No	Yes	Yes	1.597E-03			
Dickkopf-related protein 3	DKK3	38	Yes	Yes	Yes	No	1.973E-03			
Elafin	PI3	12	No	No	Yes	No	6.547E-04			
Endonuclease domain-containing 1 protein	ENDOD1	55	No	No	Yes	Yes	1.298E-04			
Ephrin-B1	EFNB1	38	Yes	Yes	Yes	Yes	3.716E-04			
Ephrin-B2	EFNB2	37	No	No	Yes	Yes	3.576E-04			
Epididymal secretory protein E1	NPC2	17	No	No	Yes	No	5.598E-04			
Fibroblast growth factor-binding protein 1	FGFBP1	26	Yes	Yes	Yes	No	2.204E-02			
Fructose-2,6-bisphosphatase TIGAR	TIGAR	30	No	No	No	No	1.915E-04			
Granulocyte colony-stimulating factor	CSF3	22	No	Yes	Yes	No	2.467E-04			
Haptoglobin	HP	45	Yes	Yes	Yes	No	3.995E-04			
Ig alpha-1 chain C region	IGHA1	38	No	Yes	No	No	3.521E-04			
Ig gamma-1 chain C region	IGHG1	36	No	Yes	No	No	4.622E-04			
Ig gamma-2 chain C region	IGHG2	36	No	Yes	No	No	7.292E-04			
Immunoglobulin lambda-like polypeptide 5	IGLL5	23	No	No	Non-Cl	No	5.319E-04			
Inactive serine protease PAMR1	PAMR1	80	No	No	Yes	No	1.755E-04			
Inhibin beta A chain	INHBA	47	Yes	Yes	Yes	No	8.948E-04			
Inter-alpha-trypsin inhibitor heavy chain H4	ITIH4	103	Yes	Yes	Yes	No	6.163E-04			
Kallikrein-10	KLK10	30	Yes	No	Yes	No	2.533E-03			
Kallikrein-5	KLK5	32	Yes	No	Yes	No	3.672E-03			
Kallikrein-7	KLK7	28	Yes	No	Yes	No	2.790E-03			
Kallikrein-8	KLK8	28	Yes	No	Yes	No	9.845E-04			
glucosaminyltransferase	B3GNT5	44	No	No	No	Yes	1.351E-04			
Laminin subunit alpha-1	LAMA1	337	No	No	Yes	No	5.397E-04			
Laminin subunit beta-3	LAMB3	130	No	Yes	Yes	No	3.561E-02			
Leucine-rich repeat transmembrane protein FLRT3	FLRT3	73	No	No	Yes	Yes	4.474E-04			
Matrix metalloproteinase-28	MMP28	59	No	No	Yes	No	1.338E-04			
Matrix metalloproteinase-9	MMP9	78	Yes	No	Yes	No	1.105E-04			
Melanoma-derived growth regulatory protein	MIA	15	No	No	Yes	No	1.922E-03			
Microfibrillar-associated protein 5	MFAP5	20	Yes	No	Yes	No	1.522E-03			
Neuropilin-2	NRP2	105	Yes	No	Yes	Yes	5.485E-05			
Nidogen-1	NID1	136	No	Yes	Yes	No	1.165E-04			
Nidogen-2	NID2	151	No	Yes	Yes	No	3.726E-04			
Parathyroid hormone-related protein	PTHLH	20	No	No	Yes	No	1.335E-03			
Plasminogen activator inhibitor 2	SERPINB2	47	No	No	No	No	3.135E-04			
Platelet-derived growth factor subunit A	PDGFA	24	No	Yes	Yes	No	7.156E-04			
Polypeptide N-acetylgalactosaminyltransferase 3	GALNT3	73	Yes	No	No	Yes	1.759E-04			
Polypeptide N-acetylgalactosaminyltransferase 5	GALNT5	106	No	Yes	Yes	Yes	1.192E-04			
Pro-neuregulin-1, membrane-bound isoform	NRG1	70	Yes	No	No	Yes	1.657E-04			
ProSAAS	PCSK1N	27	No	No	Yes	No	6.890E-04			
Protein delta homolog 2	DLK2	41	No	No	Yes	Yes	2.632E-04			
Protein FAM3A	FAM3A	25	No	No	No	Yes	1.914E-04			
Protein S100-A2	S100A2	11	Yes	No	No	No	2.190E-03			
Protein Wnt-5a	WNT5A	42	No	Yes	Yes	Yes	8.980E-05			
Prothymosin alpha	PTMA	12	Yes	No	No	No	1.717E-03			
Protocadherin Fat 2	FAT2	479	No	Yes	Yes	Yes	3.252E-04			
Receptor-type tyrosine-protein phosphatase zeta	PTPRZ1	255	No	No	Yes	Yes	8.169E-05			
Secreted frizzled-related protein 1	SFRP1	35	Yes	Yes	Yes	No	4.634E-04			
Serpin B13	SERPINB13	44	No	Yes	No	No	2.275E-04			
Serpin B5	SERPINB5	42	Yes	Yes	No	No	5.310E-03			
Serpin B7	SERPINB7	43	No	No	No	Yes	2.690E-03			
Stromelysin-2	MMP10	54	No	Yes	Yes	No	3.487E-04			
domain-containing protein 1	SVEP1	390	No	Yes	Yes	No	2.324E-05			
Thrombospondin type-1 domain-containing protein 4	THSD4	112	No	Yes	Yes	No	1.157E-04			
Tissue factor pathway inhibitor 2	TFPI2	27	Yes	No	Yes	No	1.231E-03			
Tolloid-like protein 1	TLL1	115	No	No	Yes	No	1.215E-04			
Tumor necrosis factor receptor superfamily member 6B	TNFRSF6B	33	No	No	Yes	No	2.290E-03			
Tyrosine-protein phosphatase non-receptor type substrate 1	SIRPA	55	No	Yes	Yes	Yes	1.525E-04			
Vascular endothelial growth factor C	VEGFC	47	No	Yes	Yes	No	3.193E-04			
Versican core protein	VCAN	265	No	Yes	Yes	No	1.887E-04			
Identified secreted proteins present in MCF7 only										

Identified Proteins	Gene	MW	BC database	Exosome	Signal Peptide	TM	Average NSAF			
							HMEC	MCF7	SKBR3	MDA231
Apoptosis regulator BAX	BAX	21	Yes	Yes	No	Yes	2.626E-04		3.309E-04	5.989E-04
Calcium-binding protein 39	CAB39	40	No	Yes	No	No	1.769E-04		2.565E-04	1.676E-04
Cluster of Fibrillin-1	FBN1	312	Yes	No	Yes	No	7.561E-04		1.849E-05	8.313E-05
Cluster of Fibulin-1	FBLN1	77	Yes	No	Yes	No	1.421E-03		3.427E-03	9.753E-05
Cluster of Myosin regulatory light chain 12B	MYL12B	20	No	No	No	No	1.169E-03		8.040E-04	9.059E-04
Complement C1r subcomponent	C1R	80	No	Yes	Yes	No	1.340E-04		1.485E-04	1.068E-04
Complement factor B	CFB	86	No	Yes	Yes	No	2.629E-04		8.463E-05	3.757E-05
Connective tissue growth factor	CTGF	38	Yes	No	Yes	No	6.010E-04		2.650E-04	3.264E-03
Cysteine-rich motor neuron 1 protein	CRIM1	114	No	No	Yes	Yes	1.730E-04		2.264E-05	1.057E-04
Dipeptidyl peptidase 1	CTSC	52	Yes	Yes	Yes	No	1.340E-03		4.921E-04	4.592E-04
Eukaryotic translation initiation factor 6	EIF6	27	No	Yes	Non-Cl	No	2.649E-04		6.602E-04	8.244E-04
Exostosin-1	EXT1	86	Yes	No	No	Yes	1.690E-04		5.457E-05	1.618E-04
Exostosin-2	EXT2	82	No	Yes	No	Yes	1.063E-04		1.057E-04	2.020E-04
GTP-binding nuclear protein Ran	RAN	24	No	Yes	Non-Cl	No	3.974E-04		1.063E-04	4.299E-04
Hepatocyte growth factor receptor	MET	156	Yes	No	Yes	Yes	2.527E-04		3.327E-05	1.494E-04
Insulin-like growth factor-binding protein 7	IGFBP7	29	Yes	Yes	Yes	No	1.557E-02		3.477E-04	2.019E-03
Isoform Gamma of Poliovirus receptor	PVR	39	No	No	Yes	No	2.655E-04		2.218E-04	1.800E-04
Keratin, type II cytoskeletal 7	KRT7	51	Yes	Yes	No	No	3.263E-04		4.797E-03	2.356E-04
Laminin subunit beta-1	LAMB1	198	No	Yes	Yes	No	8.435E-04		5.262E-05	5.547E-04
L-lactate dehydrogenase B chain	LDHB	37	Yes	Yes	No	No	2.884E-03		2.095E-03	3.485E-03
Myosin-9	MYH9	227	Yes	Yes	No	No	3.507E-04		7.644E-04	8.785E-04
PDZ and LIM domain protein 5	PDLIM5	64	No	No	No	No	8.828E-05		2.924E-04	1.250E-04
Pigment epithelium-derived factor	SERPINF1	46	Yes	Yes	Yes	No	1.929E-03		5.570E-05	2.073E-04
Plasminogen activator inhibitor 1	SERPINE1	45	Yes	Yes	Yes	No	1.890E-02		9.448E-03	8.647E-03
Plastin-3	PLS3	71	No	No	No	No	4.136E-04		7.565E-04	4.930E-04
Protein CYR61	CYR61	42	Yes	No	Yes	No	1.475E-03		3.519E-04	1.740E-03
Protocadherin Fat 1	FAT1	506	No	No	Yes	Yes	8.180E-05		2.953E-04	1.892E-05
Tubulointerstitial nephritis antigen-like	TINAGL1	52	No	Yes	Yes	No	1.941E-03		5.092E-04	1.600E-04
Vimentin	VIM	54	Yes	Yes	Non-Cl	No	2.965E-03		2.223E-04	8.172E-03
Zyxin	ZYX	61	Yes	No	Non-Cl	No	7.153E-05		1.418E-04	1.113E-04

Identified secreted proteins present in SKBR3 only

Identified Proteins	Gene	MW	BC database	Exosome	Signal Peptide	TM	Average NSAF			
							HMEC	MCF7	SKBR3	MDA231
Actin-related protein 2/3 complex subunit 5-like protein	ARPC5L	17	No	Yes	Non-Cl	No	2.835E-04	2.158E-04		3.029E-04
Activated RNA polymerase II transcriptional coactivator p15	SUB1	14	No	Yes	Non-Cl	No	3.466E-04	6.714E-04		3.642E-04
Acyl-protein thioesterase 1	LYPLA1	25	Yes	No	Non-Cl	No	2.192E-04	3.069E-04		4.007E-04
Alpha-N-acetylglucosaminidase	NAGLU	82	No	Yes	Yes	No	1.511E-04	1.538E-04		6.226E-05
Annexin A1	ANXA1	39	Yes	Yes	Non-Cl	No	1.560E-03	3.993E-04		4.653E-03
Apolipoprotein E	APOE	36	Yes	Yes	Yes	No	4.048E-04	1.372E-04		1.825E-04
Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe	LFNG	42	Yes	Yes	Yes	No	2.746E-04	2.504E-04		1.672E-04
Biotinidase	BTD	61	No	No	Yes	No	2.022E-04	6.537E-04		1.697E-04
CD44 antigen	CD44	82	Yes	Yes	Yes	Yes	1.612E-03	2.729E-04		1.111E-03
Cluster of HLA class I histocompatibility antigen, A-2 alpha chain	HLA-A	41	No	Yes	Yes	Yes	1.831E-03	2.726E-03		6.151E-03
Collagen alpha-1(VI) chain	COL6A1	109	No	Yes	Yes	No	1.283E-04	1.329E-04		8.947E-04
Coronin-1C	CORO1C	53	Yes	Yes	Non-Cl	No	2.160E-04	6.741E-04		8.493E-04
C-type mannose receptor 2	MRC2	167	No	Yes	Yes	Yes	1.613E-04	9.290E-05		6.635E-05
Dickkopf-related protein 1	DKK1	29	Yes	Yes	Yes	No	2.603E-03	1.898E-03		1.515E-04
Di-N-acetylchitobiase	CTBS	44	No	No	Yes	No	2.923E-04	1.615E-04		1.653E-04
Extracellular matrix protein 1	ECM1	61	Yes	Yes	Yes	No	4.545E-04	1.147E-04		2.375E-03
Fascin	FSCN1	55	Yes	Yes	No	No	1.449E-03	2.098E-04		3.038E-04
Follistatin-related protein 3	FSTL3	28	No	No	Yes	No	1.725E-03	1.639E-04		8.745E-05
Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	B3GAT3	37	No	Yes	Yes	Yes	4.336E-04	1.133E-04		1.381E-04
Galectin-3	LGALS3	26	Yes	Yes	No	No	3.980E-04	3.707E-03		7.830E-04
Glypican-1	GPC1	62	No	Yes	Yes	Yes	2.335E-03	9.412E-04		7.624E-04
Granulins	GRN	64	No	Yes	Yes	No	7.616E-04	2.087E-04		3.305E-04
High mobility group protein B1	HMGB1	25	Yes	No	No	No	5.180E-04	2.031E-03		1.711E-03
Hydroxymethylglutaryl-CoA synthase, cytoplasmic	HMGCS1	57	No	No	No	No	1.325E-04	4.487E-04		6.651E-05
Insulin-like growth factor-binding protein 4	IGFBP4	28	Yes	Yes	Yes	No	6.819E-04	1.140E-03		2.926E-03
Integrin alpha-3	ITGA3	117	Yes	Yes	Yes	Yes	1.818E-04	8.618E-05		2.525E-04
Integrin beta-1	ITGB1	88	No	Yes	Yes	Yes	3.566E-04	1.856E-04		3.123E-04
monooxygenase	PAM	108	No	Yes	Yes	Yes	2.807E-04	2.050E-04		5.296E-05
Isoform Alpha-6X1A of Integrin alpha-6	ITGA6	119	No	Yes	Yes	Yes	2.666E-04	2.956E-05		7.162E-05
Laminin subunit beta-2	LAMB2	196	Yes	Yes	Yes	No	8.998E-05	7.267E-04		6.439E-05
Laminin subunit gamma-1	LAMC1	178	No	Yes	Yes	No	1.922E-03	1.083E-03		1.119E-03
Major prion protein	PRNP	28	Yes	Yes	Yes	Yes	4.214E-03	7.576E-05		9.142E-04
Metalloproteinase inhibitor 1	TIMP1	23	Yes	Yes	Yes	No	2.335E-02	3.332E-03		1.839E-03
Microtubule-associated protein 4	MAP4	121	No	No	No	No	5.825E-05	9.435E-05		1.497E-04
N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	B3GNT1	47	No	Yes	Yes	Yes	3.691E-04	2.406E-04		8.418E-05
Neuronal cell adhesion molecule	NRCAM	144	No	No	Yes	Yes	2.117E-04	1.035E-03		3.117E-05
Nucleobindin-2	NUCB2	50	No	Yes	Yes	No	2.034E-04	6.311E-04		1.239E-04
Peptidyl-prolyl cis-trans isomerase FKBP3	FKBP3	25	No	No	Non-Cl	No	1.553E-04	5.295E-04		2.806E-04
Polypeptide N-acetylgalactosaminyltransferase 2	GALNT2	65	Yes	No	Yes	Yes	2.833E-04	1.330E-04		7.252E-04
Serine protease 23	PRSS23	43	Yes	Yes	Yes	No	2.357E-04	8.780E-05		5.854E-04

Serpin H1	SERPINH1	46	No	No	Yes	No	1.053E-04	1.703E-04		3.719E-04
Signal recognition particle 14 kDa protein	SRP14	15	Yes	No	No	No	4.435E-04	5.891E-04		5.467E-04
Sulfhydryl oxidase 2	QSOX2	78	No	No	Yes	Yes	1.829E-04	4.135E-05		1.240E-04
Torsin-1B	TOR1B	38	No	Yes	Yes	Yes	3.624E-04	1.129E-04		1.732E-04
V-type proton ATPase subunit S1	ATP6AP1	52	Yes	Yes	Yes	Yes	5.929E-04	3.404E-04		9.898E-05
Identified secreted proteins present in MDA231 only										
Identified Proteins	**Gene**	**MW**	**BC database**	**Exosome**	**Signal Peptide**	**TM**	**Average NSAF**			
							HMEC	**MCF7**	**SKBR3**	**MDA231**
Angiopoietin-related protein 4	ANGPTL4	45	Yes	Yes	Yes	No	9.263E-04	3.616E-04	7.220E-05	
Apolipoprotein D	APOD	21	Yes	Yes	Yes	No	3.109E-04	1.771E-04	2.456E-04	
Beta-mannosidase	MANBA	101	No	No	Yes	No	3.722E-04	1.941E-04	3.183E-04	
Calumenin	CALU	37	Yes	Yes	Yes	No	4.193E-04	4.056E-04	1.251E-03	
Carbonic anhydrase 12	CA12	39	Yes	No	Yes	Yes	2.494E-04	5.614E-04	3.621E-04	
Cathepsin L2	CTSV	37	No	No	Yes	No	1.616E-03	8.526E-05	8.629E-05	
CD9 antigen	CD9	25	No	Yes	No	Yes	2.377E-03	9.771E-04	5.429E-04	
Cluster of Isoform 2 of Tropomyosin beta chain	TPM2	33	Yes	No	No	No	9.362E-04	1.172E-03	1.265E-03	
Cluster of Kunitz-type protease inhibitor 1	SPINT1	58	No	No	Yes	Yes	1.016E-03	8.204E-04	1.800E-03	
Cluster of Lipolysis-stimulated lipoprotein receptor	LSR	71	No	Yes	No	Yes	1.813E-03	1.219E-03	5.970E-04	
dCTP pyrophosphatase 1	DCTPP1	19	No	No	Non-Cl	No	2.859E-04	6.107E-04	6.466E-04	
Epithelial discoidin domain-containing receptor 1	DDR1	101	Yes	Yes	Yes	Yes	1.230E-04	1.874E-04	1.953E-04	
G-protein coupled receptor 126	GPR126	137	Yes	No	Yes	Yes	6.649E-05	1.011E-04	1.139E-04	
Insulin-like growth factor-binding protein 2	IGFBP2	35	Yes	No	Yes	No	2.028E-02	3.473E-03	1.423E-04	
Junctional adhesion molecule A	F11R	33	No	Yes	Yes	Yes	6.043E-04	1.284E-03	1.245E-03	
Neutrophil gelatinase-associated lipocalin	LCN2	23	Yes	Yes	Yes	No	9.993E-04	2.188E-04	1.550E-02	
Poliovirus receptor-related protein 1	PVRL1	57	No	No	Yes	Yes	3.159E-04	1.850E-04	1.244E-04	
Prostaglandin F2 receptor negative regulator	PTGFRN	99	No	Yes	Yes	Yes	2.607E-04	7.057E-04	1.196E-03	
Prostasin	PRSS8	36	No	Yes	Yes	No	1.892E-04	2.082E-04	2.379E-04	
Protein canopy homolog 2	CNPY2	21	No	No	Yes	No	2.891E-04	1.315E-04	6.381E-04	
Protein S100-A14	S100A14	12	Yes	Yes	Non-Cl	No	8.907E-04	1.873E-03	7.895E-03	
Rho GTPase-activating protein 1	ARHGAP1	50	No	Yes	No	No	1.394E-04	8.681E-05	1.331E-04	
Semaphorin-4B	SEMA4B	92	No	No	Yes	Yes	5.323E-04	2.334E-04	6.988E-05	
Sortilin-related receptor	SORL1	248	No	Yes	Yes	Yes	5.784E-05	1.313E-04	2.870E-04	
Tumor-associated calcium signal transducer 2	TACSTD2	36	Yes	Yes	No	Yes	4.464E-04	6.920E-04	1.223E-03	
Identified secreted proteins present in at least two cell lines										
Identified Proteins	**Gene**	**MW**	**BC database**	**Exosome**	**Signal Peptide**	**TM**	**Average NSAF**			
							HMEC	**MCF7**	**SKBR3**	**MDA231**
1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	ADI1	21	No	No	No	No		3.466E-04		
1,4-alpha-glucan-branching enzyme	GBE1	80	No	Yes	No	No		1.211E-04		1.811E-04
phosphodiesterase beta-3	PLCB3	139	No	No	No	No		8.126E-05	2.303E-04	
phosphodiesterase beta-4	PLCB4	134	No	No	No	No			2.453E-05	
phosphodiesterase gamma-1	PLCG1	149	Yes	No	No	No		2.190E-05		5.369E-05
2',3'-cyclic-nucleotide 3'-phosphodiesterase	CNP	45	No	Yes	No	No			5.389E-04	
2,4-dienoyl-CoA reductase, mitochondrial	DECR1	36	No	Yes	Non-Cl	No			1.741E-04	1.192E-04
26S protease regulatory subunit 4	PSMC1	49	No	No	No	No		5.439E-05		6.563E-05
26S protease regulatory subunit 6B	PSMC4	47	Yes	Yes	Non-Cl	No		5.725E-05		
26S proteasome non-ATPase regulatory subunit 10	PSMD10	24	Yes	No	No	No				1.783E-04
26S proteasome non-ATPase regulatory subunit 12	PSMD12	53	Yes	Yes	No	No		1.571E-04		1.143E-04
26S proteasome non-ATPase regulatory subunit 7	PSMD7	37	Yes	Yes	No	No			1.085E-04	1.241E-04
26S proteasome non-ATPase regulatory subunit 9	PSMD9	25	No	No	Non-Cl	No		2.996E-04		5.432E-04
3-hydroxyisobutyrate dehydrogenase, mitochondrial	HIBADH	35	No	No	Non-Cl	No			3.805E-04	2.913E-04
3-hydroxyisobutyryl-CoA hydrolase, mitochondrial	HIBCH	43	No	No	Non-Cl	No				9.028E-05
40S ribosomal protein S12	RPS12	15	Yes	No	Non-Cl	No			3.108E-04	2.188E-04
40S ribosomal protein S13	RPS13	17	No	Yes	Non-Cl	No		3.761E-04		3.087E-04
40S ribosomal protein S14	RPS14	16	Yes	Yes	Non-Cl	No		5.015E-04		3.837E-04
40S ribosomal protein S16	RPS16	16	Yes	Yes	Non-Cl	No		6.474E-04		5.188E-04
40S ribosomal protein S17	RPS17	16	No	No	Non-Cl	No		2.093E-04		
40S ribosomal protein S18	RPS18	18	Yes	Yes	No	No		4.086E-04		4.637E-04
40S ribosomal protein S19	RPS19	16	No	Yes	Non-Cl	No		4.272E-04		
40S ribosomal protein S2	RPS2	31	No	Yes	Non-Cl	No		2.577E-04		
40S ribosomal protein S20	RPS20	13	No	Yes	Non-Cl	No		5.550E-04	7.307E-04	
40S ribosomal protein S27	RPS27	9	No	No	No	No		6.806E-04		4.786E-04
40S ribosomal protein S3a	RPS3A	30	No	Yes	Non-Cl	No		5.557E-04		2.424E-04
40S ribosomal protein S6	RPS6	29	No	No	No	No		1.524E-04		
40S ribosomal protein S8	RPS8	24	No	Yes	No	No		2.032E-04		
5'-3' exoribonuclease 2	XRN2	109	No	No	No	No		3.021E-05		
5'-nucleotidase	NT5E	63	No	Yes	Yes	No				3.124E-04
60S ribosomal protein L10a	RPL10A	25	No	Yes	Non-Cl	No		3.051E-04		3.989E-04
60S ribosomal protein L13	RPL13	24	No	No	No	No		1.791E-04		
60S ribosomal protein L14	RPL14	23	No	Yes	No	No				1.327E-04
60S ribosomal protein L23a	RPL23A	18	Yes	No	No	No		2.131E-04		
60S ribosomal protein L3	RPL3	46	No	Yes	Non-Cl	No		8.139E-05		
60S ribosomal protein L30	RPL30	13	No	Yes	No	No		4.162E-04		
60S ribosomal protein L5	RPL5	34	Yes	Yes	No	No		2.555E-04	1.377E-04	
60S ribosomal protein L6	RPL6	33	No	Yes	No	No		2.478E-04		1.204E-04
6-phosphofructokinase type C	PFKP	86	No	Yes	No	No		7.272E-05		2.339E-04
6-phosphofructokinase, liver type	PFKL	85	Yes	Yes	No	No		7.281E-05	1.130E-04	
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	45	No	Yes	Non-Cl	No				2.317E-04
Acid ceramidase	ASAH1	45	Yes	Yes	Yes	No	1.510E-04		6.479E-04	
Acid sphingomyelinase-like phosphodiesterase 3a	SMPDL3A	51	No	No	Yes	No			1.032E-04	

Acid sphingomyelinase-like phosphodiesterase 3b	SMPDL3B	51	No	Yes	Yes	No		1.356E-04	1.908E-04	
Acidic leucine-rich nuclear phosphoprotein 32 family member A	ANP32A	29	No	No	No	No		7.207E-04	3.252E-04	
Acidic leucine-rich nuclear phosphoprotein 32 family member E	ANP32E	31	No	No	No	No				1.278E-04
Aconitate hydratase, mitochondrial	ACO2	85	No	No	No	No				9.566E-05
Acyl-CoA-binding protein	DBI	10	No	Yes	Non-Cl	No			5.973E-04	
Acyl-coenzyme A thioesterase 2, mitochondrial	ACOT2	53	No	No	No	No			6.069E-05	
Acylphosphatase-1	ACYP1	11	No	No	No	No		2.417E-04		2.323E-04
Adapter molecule crk	CRK	34	No	Yes	No	No			1.350E-04	
Adenosine kinase	ADK	41	Yes	Yes	No	No		6.505E-05	9.492E-05	
Adenylate kinase 4, mitochondrial	AK4	25	No	No	Non-Cl	No				1.027E-04
Adenylosuccinate lyase	ADSL	55	Yes	Yes	Non-Cl	No		1.086E-04		1.438E-04
ADP-ribosylation factor 4	ARF4	21	No	Yes	Non-Cl	No				7.042E-04
ADP-ribosylation factor 5	ARF5	21	No	Yes	Non-Cl	No			6.072E-04	
ADP-ribosylation factor 6	ARF6	20	No	Yes	Non-Cl	No			3.611E-04	
Adseverin	SCIN	80	No	Yes	No	No			1.547E-04	
AH receptor-interacting protein	AIP	38	Yes	No	No	No				1.048E-04
Aldehyde dehydrogenase family 16 member A1	ALDH16A1	85	No	Yes	Non-Cl	No		2.964E-05		
Aldehyde dehydrogenase, mitochondrial	ALDH2	56	No	Yes	Non-Cl	No			1.455E-04	7.793E-05
Aldose 1-epimerase	GALM	38	No	Yes	No	No			2.195E-04	
Aldose reductase	AKR1B1	36	Yes	Yes	No	No				1.825E-04
Allograft inflammatory factor 1-like	AIF1L	17	No	Yes	Non-Cl	No		2.833E-04		
Alpha-(1,6)-fucosyltransferase	FUT8	67	Yes	Yes	No	Yes		7.457E-05		
Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	MGAT1	51	No	Yes	No	Yes	1.355E-04	4.270E-05		
Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A	MGAT5	85	No	No	No	Yes	1.955E-04	1.097E-04		
Alpha-1-antichymotrypsin	SERPINA3	48	Yes	Yes	Yes	No		8.156E-03		
Alpha-1-antitrypsin	SERPINA1	47	No	Yes	Yes	No	2.701E-03	4.995E-04		
Alpha-2-macroglobulin-like protein 1	A2ML1	161	No	Yes	Yes	No	3.453E-04		2.036E-04	
Alpha-mannosidase 2x	MAN2A2	131	No	No	No	Yes		3.300E-05	8.144E-05	
Alpha-N-acetylgalactosaminidase	NAGA	47	Yes	No	Yes	No				1.262E-04
Amidophosphoribosyltransferase	PPAT	57	No	No	No	No		4.598E-05		
Aminoacyl tRNA synthase complex-interacting multifunctional protein 1	AIMP1	34	No	Yes	Non-Cl	No		1.978E-04		
Amphoterin-induced protein 2	AMIGO2	58	No	No	No	Yes	1.497E-04	4.645E-04		
Angio-associated migratory cell protein	AAMP	47	Yes	No	Non-Cl	No		7.659E-05		
Angiotensin-converting enzyme	ACE	150	No	Yes	Yes	Yes		9.770E-05		
Annexin A11	ANXA11	54	Yes	Yes	Non-Cl	No			3.689E-04	
Annexin A7	ANXA7	53	No	Yes	Non-Cl	No			4.038E-04	5.882E-05
Anterior gradient protein 2 homolog	AGR2	20	Yes	Yes	Yes	No		2.441E-04	9.568E-04	
AP-1 complex subunit gamma-1	AP1G1	91	No	No	No	No		1.494E-04		
AP-1 complex subunit mu-1	AP1M1	49	No	Yes	Non-Cl	No		2.141E-04		1.235E-04
AP-1 complex subunit mu-2	AP1M2	48	Yes	No	No	No		3.703E-04		
Apolipoprotein A-I	APOA1	31	Yes	Yes	Yes	No	2.455E-04	1.608E-04		
Apolipoprotein B-100	APOB	516	No	Yes	Yes	No			6.386E-06	1.639E-05
Apoptosis-associated speck-like protein containing a CARD	PYCARD	22	Yes	No	Non-Cl	No	2.539E-04	4.415E-04		
Apoptosis-inducing factor 1, mitochondrial	AIFM1	66	No	No	No	Yes		2.167E-04	3.212E-04	
Arfaptin-1	ARFIP1	42	No	Yes	No	No		7.634E-05		
Argininosuccinate lyase	ASL	52	No	Yes	No	No		1.420E-04	3.268E-04	
Argininosuccinate synthase	ASS1	47	No	Yes	No	No		4.700E-04	2.119E-04	
Arrestin domain-containing protein 1	ARRDC1	46	No	Yes	Non-Cl	No		2.951E-04	5.244E-04	
Arylsulfatase A	ARSA	54	Yes	No	Yes	No			6.875E-05	
Aspartyl/asparaginyl beta-hydroxylase	ASPH	86	No	No	No	Yes				3.775E-05
Astrocytic phosphoprotein PEA-15	PEA15	15	Yes	No	Non-Cl	No			5.888E-04	6.591E-04
Ataxin-10	ATXN10	53	No	No	No	No		6.998E-05		1.094E-04
ATP synthase subunit beta, mitochondrial	ATP5B	57	No	Yes	Non-Cl	No		7.257E-05	2.417E-04	
ATPase ASNA1	ASNA1	39	No	Yes	Non-Cl	No		8.229E-05		
ATP-binding cassette sub-family A member 12	ABCA12	293	No	No	No	Yes		1.471E-05		
ATP-binding cassette sub-family E member 1	ABCE1	67	No	No	No	No		4.690E-05		3.883E-05
Band 4.1-like protein 1	EPB41L1	99	No	No	No	No			1.059E-04	
Band 4.1-like protein 2	EPB41L2	113	No	Yes	No	No		3.814E-05		
Beta-1,4 N-acetylgalactosaminyltransferase 1	B4GALNT1	59	No	No	Yes	No		3.524E-05		
Beta-galactoside alpha-2,6-sialyltransferase 1	ST6GAL1	47	Yes	No	No	Yes	1.258E-04		8.706E-05	
Beta-lactamase-like protein 2	LACTB2	33	No	No	No	No			7.081E-04	2.417E-04
Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1	PAPSS1	71	No	No	No	No		4.563E-05		
Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	DAK	59	No	Yes	No	No		2.060E-04	1.209E-04	
Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1	NDST1	101	No	No	No	Yes			1.979E-04	
Bis(5'-nucleosyl)-tetraphosphatase [asymmetrical]	NUDT2	17	Yes	No	Non-Cl	No		1.304E-04		3.135E-04
Bleomycin hydrolase	BLMH	53	Yes	Yes	Non-Cl	No			2.677E-04	1.266E-04
Bone morphogenetic protein 7	BMP7	49	Yes	No	Yes	No		7.701E-04		
Brain acid soluble protein 1	BASP1	23	No	Yes	No	No		4.183E-04		2.042E-04
Brain-specific angiogenesis inhibitor 1-associated protein 2	BAIAP2	61	No	Yes	No	No		5.228E-05	3.257E-04	
Brain-specific serine protease 4	PRSS22	34	No	No	Yes	No		2.525E-04	1.290E-04	

Branched-chain-amino-acid aminotransferase, cytosolic	BCAT1	43	No	No	No	No				2.383E-04
Brefeldin A-inhibited guanine nucleotide-exchange protein 2	ARFGEF2	202	No	No	No	No			2.601E-05	
BRISC complex subunit Abro1	FAM175B	47	No	No	No	No		4.526E-05		
BRO1 domain-containing protein BROX	BROX	46	No	No	No	No			8.557E-05	1.123E-04
BTB/POZ domain-containing protein KCTD12	KCTD12	36	No	No	Non-CI	No	4.449E-04			3.541E-04
Cadherin-1	CDH1	97	Yes	Yes	Yes	Yes	5.925E-04	6.164E-04		
Cadherin-11	CDH11	88	Yes	No	Yes	Yes				2.889E-05
Cadherin-4	CDH4	100	No	No	Yes	Yes	6.584E-05			1.758E-04
Calcineurin B homologous protein 1	CHP1	22	No	No	Non-CI	No			2.071E-04	
Calcineurin-like phosphoesterase domain-containing protein 1	CPPED1	36	No	No	No	No			3.159E-04	1.815E-04
Calcium and integrin-binding protein 1	CIB1	22	Yes	Yes	Non-CI	No			1.493E-04	
Calcium-binding and coiled-coil domain-containing protein 2	CALCOCO2	52	No	No	Non-CI	No				6.436E-05
Calcyphosin	CAPS	21	No	Yes	Non-CI	No			1.037E-03	
Calmodulin	CALM1	17	No	Yes	Non-CI	No		1.617E-04		
Calnexin	CANX	68	Yes	No	Yes	Yes			7.842E-05	
Calpain-2 catalytic subunit	CAPN2	80	No	Yes	Non-CI	No			1.072E-04	2.069E-04
Calponin-2	CNN2	34	No	No	No	No				3.920E-04
Calponin-3	CNN3	36	Yes	No	Non-CI	No				6.841E-04
Calretinin	CALB2	32	Yes	No	No	No				2.125E-04
Calsyntenin-2	CLSTN2	107	No	No	Yes	Yes		4.684E-04		
cAMP-dependent protein kinase type I-alpha regulatory subunit	PRKAR1A	43	Yes	No	No	No		1.495E-04		2.124E-04
cAMP-dependent protein kinase type II-alpha regulatory subunit	PRKAR2A	46	No	Yes	Non-CI	No			1.029E-04	1.006E-04
Caprin-1	CAPRIN1	78	No	No	No	No		7.377E-05		
Carbohydrate sulfotransferase 14	CHST14	43	No	Yes	No	Yes		1.395E-04		
Carbonyl reductase [NADPH] 3	CBR3	31	Yes	Yes	Non-CI	No				7.057E-04
Carboxymethylenebutenolidase homolog	CMBL	28	No	Yes	No	No		4.624E-04		
Carboxypeptidase A4	CPA4	47	Yes	No	Yes	No	8.571E-03		9.094E-04	
Carboxypeptidase D	CPD	153	No	Yes	Yes	Yes			1.565E-04	7.111E-05
Carboxypeptidase Q	CPQ	52	No	No	Yes	No				4.854E-05
Cartilage intermediate layer protein 2	CILP2	126	No	No	Yes	No			2.536E-05	
Casein kinase II subunit beta	CSNK2B	25	No	Yes	Non-CI	No		1.324E-04		
Caspase-14	CASP14	28	No	Yes	No	No		1.962E-04		
Caspase-3	CASP3	32	Yes	No	No	No			1.881E-04	2.073E-04
Catenin alpha-1	CTNNA1	100	Yes	Yes	No	No		2.672E-04		5.087E-05
Catenin beta-1	CTNNB1	85	Yes	Yes	Non-CI	No		5.049E-04		5.159E-05
Catenin delta-1	CTNND1	108	No	Yes	No	No		6.862E-05	4.256E-05	
CCA tRNA nucleotidyltransferase 1, mitochondrial	TRNT1	50	Yes	No	No	No		4.379E-05		
CD109 antigen	CD109	162	No	Yes	Yes	No	5.613E-04	2.227E-04		
CD276 antigen	CD276	57	No	Yes	Yes	Yes		5.405E-04		
CD63 antigen	CD63	26	Yes	Yes	No	Yes		9.399E-04		
CD82 antigen	CD82	30	Yes	Yes	No	Yes			1.196E-03	1.940E-04
CD99 antigen	CD99	19	No	No	Yes	Yes		4.106E-04		
Cell cycle and apoptosis regulator protein 2	CCAR2	103	No	No	No	No		3.637E-05	3.757E-05	
Cell surface glycoprotein MUC18	MCAM	72	No	Yes	Yes	Yes				1.525E-04
Cellular nucleic acid-binding protein	CNBP	19	Yes	No	Non-CI	No		1.340E-04		4.201E-04
Cellular retinoic acid-binding protein 1	CRABP1	16	Yes	No	Non-CI	No		1.376E-03		
Cellular retinoic acid-binding protein 2	CRABP2	16	No	Yes	Non-CI	No		2.025E-03	8.403E-04	
Ceruloplasmin	CP	122	Yes	Yes	Yes	No			2.036E-04	
Charged multivesicular body protein 4b	CHMP4B	25	No	Yes	No	No				2.320E-04
Chloride intracellular channel protein 3	CLIC3	27	No	Yes	Non-CI	No			2.706E-04	
Choline transporter-like protein 1	SLC44A1	73	No	Yes	No	Yes		1.363E-04	5.305E-05	
Choline transporter-like protein 2	SLC44A2	80	No	Yes	No	Yes		2.432E-04	4.053E-04	
Chondroitin sulfate proteoglycan 4	CSPG4	251	Yes	Yes	Yes	Yes				3.226E-04
Claudin-3	CLDN3	23	No	Yes	No	Yes		3.901E-04	2.909E-04	
Cleavage stimulation factor subunit 1	CSTF1	48	No	No	Non-CI	No			6.801E-05	
Cleavage stimulation factor subunit 3	CSTF3	83	Yes	No	No	No		3.338E-05		
Cluster of 60S ribosomal protein L26	RPL26	17	No	No	No	No		1.311E-04		
Cluster of Aldo-keto reductase family 1 member C2	AKR1C2	37	Yes	No	No	No	2.896E-04	1.170E-04		
Cluster of Cadherin EGF LAG seven-pass G-type receptor 2	CELSR2	317	Yes	No	Yes	Yes		3.468E-05	2.379E-05	
Cluster of Collagen alpha-2(VI) chain	COL6A2	109	No	Yes	Yes	No		3.686E-05		2.271E-04
Cluster of Complement C4-B	C4B	193	Yes	Yes	Yes	No			3.036E-05	8.260E-05
Cluster of Creatine kinase U-type, mitochondrial	CKMT1A	47	No	No	No	Yes			3.363E-04	
Cluster of Cullin-4B	CUL4B	104	No	Yes	No	No				3.163E-05
Cluster of Dual specificity mitogen-activated protein kinase kinase 1	MAP2K1	43	No	No	No	No				2.059E-04
Cluster of Dynamin-2	DNM2	98	No	Yes	No	No		7.667E-05		
Cluster of Fructose-1,6-bisphosphatase 1	FBP1	37	Yes	Yes	Non-CI	No		3.355E-03	7.016E-04	
Cluster of Hepatoma-derived growth factor	HDGF	27	Yes	No	No	No		9.972E-05	9.775E-05	
Cluster of Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	40	No	No	No	No		2.637E-04		1.370E-04
Cluster of Histone H1.2	HIST1H1C	21	No	No	No	No				9.301E-04
Cluster of Histone H2A type 1	HIST1H2AG	14	No	Yes	Non-CI	No				7.562E-04
Cluster of Importin subunit alpha-4	KPNA3	58	No	No	Non-CI	No		1.285E-04	6.690E-05	

Cluster of Latent-transforming growth factor beta-binding protein 3	LTBP3	139	No	Yes	Yes	Yes				1.111E-04
Cluster of Low-density lipoprotein receptor-related protein 6	LRP6	180	No	Yes	Yes	No			2.535E-05	
Cluster of Membrane cofactor protein	CD46	44	No	Yes	Yes	Yes		8.423E-05	2.820E-04	
Cluster of NEDD4-like E3 ubiquitin-protein ligase WWP1	WWP1	105	Yes	No	No	No			3.126E-05	
Cluster of NKG2D ligand 2	ULBP2	27	No	No	Yes	Yes	1.676E-03	3.891E-04		
Cluster of Probable ATP-dependent RNA helicase DDX17	DDX17	80	No	No	Yes	No		1.177E-04		7.935E-05
Cluster of Protein jagged-1	JAG1	134	No	No	Yes	Yes	2.673E-04	5.026E-05		
Cluster of Protein mago nashi homolog	MAGOH	17	Yes	No	No	No			3.648E-04	7.184E-04
Cluster of Protein phosphatase 1B	PPM1B	53	No	No	No	No		1.380E-04	6.083E-05	
Cluster of Protein scribble homolog	SCRIB	175	No	Yes	No	No		4.371E-05	6.464E-05	
Cluster of Proto-oncogene tyrosine-protein kinase Src	SRC	60	Yes	Yes	No	No		5.384E-05	8.629E-05	
Cluster of Ras-related protein Rap-2c	RAP2C	21	No	Yes	Non-Cl	No		8.774E-04	4.798E-04	
Cluster of Ribose-phosphate pyrophosphokinase 1	PRPS1	35	No	No	No	No		7.526E-05		1.444E-04
Cluster of Semaphorin-3F	SEMA3F	88	No	No	Yes	No		5.279E-04	8.546E-04	
Cluster of Sushi repeat-containing protein SRPX	SRPX	52	Yes	No	Yes	No	1.423E-03			3.608E-04
Cluster of Thiosulfate sulfurtransferase/rhodanese-like domain-containing protein 1	TSTD1	13	No	No	Non-Cl	No			4.531E-04	
Cluster of Ubiquilin-2	UBQLN2	66	No	No	No	No		6.784E-05	3.731E-05	
Clustered mitochondria protein homolog	CLUH	147	No	No	No	No		1.841E-05	8.449E-05	
Coactosin-like protein	COTL1	16	No	Yes	Non-Cl	No			6.181E-04	2.533E-03
Coagulation factor V	F5	252	No	Yes	Yes	No			1.816E-05	
Cochlin	COCH	59	No	No	Yes	No		3.415E-05		
Cohesin subunit SA-2	STAG2	141	No	No	No	No		2.295E-05		
Cold shock domain-containing protein E1	CSDE1	89	No	No	No	No		3.541E-05	5.124E-05	
Collagen alpha-1(I) chain	COL1A1	139	Yes	Yes	Yes	No		1.930E-05		
Collagen alpha-1(VII) chain	COL7A1	295	No	Yes	Yes	No	2.283E-04			3.119E-04
Collagen alpha-1(XIII) chain	COL13A1	70	No	No	No	Yes				4.860E-05
Collagen alpha-2(I) chain	COL1A2	129	Yes	Yes	Yes	No		3.415E-05		
Collagen alpha-2(IV) chain	COL4A2	168	No	Yes	Yes	No	8.140E-05			3.537E-04
Collagen alpha-5(IV) chain	COL4A5	161	Yes	No	Yes	No		1.008E-04		
Collagen and calcium-binding EGF domain-containing protein 1	CCBE1	44	No	No	Yes	No				7.135E-05
Complement C1s subcomponent	C1S	77	Yes	Yes	Yes	No				3.343E-05
Complement component 1 Q subcomponent-binding protein, mitochondrial	C1QBP	31	No	No	No	No				4.501E-04
Complement decay-accelerating factor	CD55	41	No	Yes	Yes	No		2.119E-04		
Complement factor H	CFH	139	No	Yes	Yes	No				2.585E-04
Condensin complex subunit 1	NCAPD2	157	No	No	No	No				1.635E-05
Contactin-1	CNTN1	113	No	No	Yes	Yes	2.616E-04		9.231E-05	
Contactin-3	CNTN3	113	No	No	Yes	Yes			6.194E-05	
COP9 signalosome complex subunit 5	COPS5	38	No	No	No	No		7.099E-05		
Copine-3	CPNE3	60	No	Yes	Non-Cl	No			8.645E-05	
Copine-8	CPNE8	63	No	Yes	No	No				9.214E-05
Coproporphyrinogen-III oxidase, mitochondrial	CPOX	50	No	No	Non-Cl	No				6.438E-05
Core histone macro-H2A.1	H2AFY	40	No	Yes	No	No			6.306E-05	
Coronin-1A	CORO1A	51	Yes	Yes	No	Yes			2.145E-04	
Creatine kinase B-type	CKB	43	No	Yes	No	No			2.441E-04	
Crk-like protein	CRKL	34	No	No	Non-Cl	No			1.923E-04	
CTP synthase 1	CTPS1	67	No	No	No	No			3.969E-05	7.812E-05
CTP synthase 2	CTPS2	66	No	No	No	No		6.486E-05		
C-type lectin domain family 11 member A	CLEC11A	36	No	Yes	Yes	No				8.941E-05
Cullin-3	CUL3	89	No	Yes	No	No		4.899E-05		
C-X-C motif chemokine 16	CXCL16	28	No	Yes	Yes	Yes	6.247E-04	5.071E-04		
Cyclin-dependent kinase 1	CDK1	34	No	No	No	No			9.705E-05	7.744E-05
Cyclin-dependent kinase 7	CDK7	39	Yes	No	No	No				8.372E-05
Cystathionine beta-synthase	CBS	61	No	No	Non-Cl	No		8.606E-05		
Cystatin-F	CST7	16	No	No	Yes	No				5.575E-04
Cystatin-M	CST6	17	Yes	No	Yes	No	4.564E-04		5.464E-04	
Cysteine and histidine-rich domain-containing protein 1	CHORDC1	37	Yes	No	Non-Cl	No			2.462E-04	2.779E-04
Cysteine-rich protein 2	CRIP2	22	Yes	No	Non-Cl	No			5.010E-04	
Cytochrome c oxidase subunit 6B1	COX6B1	10	No	No	No	No			3.409E-04	
Cytokine receptor-like factor 1	CRLF1	46	No	No	Yes	No		2.366E-04		
Cytokine receptor-like factor 3	CRLF3	50	Yes	No	No	No			5.196E-05	
Cytoplasmic aconitate hydratase	ACO1	98	No	Yes	No	No		4.269E-05		4.493E-05
Cytoplasmic dynein 1 intermediate chain 2	DYNC1I2	71	No	No	No	No		5.176E-05	6.458E-05	
Cytoskeleton-associated protein 5	CKAP5	226	No	No	No	No		3.681E-05		3.106E-05
Cytosolic purine 5'-nucleotidase	NT5C2	65	No	No	No	No		5.076E-05		
D-3-phosphoglycerate dehydrogenase	PHGDH	57	Yes	Yes	No	No		1.246E-04		
DCN1-like protein 1	DCUN1D1	30	No	No	No	No		7.252E-05	3.364E-04	
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	ECH1	36	No	Yes	Non-Cl	No		1.755E-04	1.579E-04	
Delta-aminolevulinic acid dehydratase	ALAD	36	No	Yes	Non-Cl	No			1.777E-04	
Deoxycytidylate deaminase	DCTD	20	Yes	No	Non-Cl	No				7.772E-04
Deoxyhypusine synthase	DHPS	41	Yes	No	Non-Cl	No		1.560E-04		
Deoxyribonuclease-1	DNASE1	31	No	No	Yes	No			1.230E-04	
Dermcidin	DCD	11	No	Yes	Yes	No				6.272E-04
Desmocollin-2	DSC2	100	Yes	Yes	Yes	Yes	2.646E-04		2.701E-04	

Desmoglein-1	DSG1	114	No	No	Yes	Yes		2.715E-05		
Developmentally-regulated GTP-binding protein 1	DRG1	41	Yes	No	No	No		1.680E-04		
Dihydrolipoyl dehydrogenase, mitochondrial	DLD	54	No	Yes	Non-Cl	No			1.498E-04	9.157E-05
Dihydropyrimidinase-related protein 2	DPYSL2	62	Yes	Yes	No	No		5.850E-05		
Dipeptidyl peptidase 9	DPP9	98	No	No	No	No			7.322E-05	4.659E-05
Diphosphoinositol polyphosphate phosphohydrolase 1	NUDT3	19	No	No	No	No			1.704E-04	
Diphosphomevalonate decarboxylase	MVD	43	No	No	No	No		9.446E-05		
Disks large homolog 1	DLG1	100	Yes	Yes	No	No		2.100E-04	2.506E-04	
DNA-dependent protein kinase catalytic subunit	PRKDC	469	No	Yes	No	No			4.371E-05	8.446E-05
DNA-directed RNA polymerases I, II, and III subunit RPABC3	POLR2H	17	No	No	No	No			3.485E-04	3.833E-04
DnaJ homolog subfamily C member 3	DNAJC3	58	No	Yes	Yes	No				9.247E-05
DnaJ homolog subfamily C member 5	DNAJC5	22	No	No	No	No		1.427E-04		
DnaJ homolog subfamily C member 7	DNAJC7	56	No	Yes	No	No		3.802E-05		8.227E-05
DnaJ homolog subfamily C member 9	DNAJC9	30	No	No	No	No				1.124E-04
Double-stranded RNA-binding protein Staufen homolog 1	STAU1	63	No	Yes	No	No		6.587E-05	4.995E-05	
Drebrin	DBN1	71	No	No	No	No		4.294E-05		3.584E-05
D-tyrosyl-tRNA(Tyr) deacylase 1	DTD1	23	No	No	Non-Cl	No		4.787E-04		1.373E-04
Dual specificity protein phosphatase 23	DUSP23	17	Yes	No	Non-Cl	No		1.884E-04		
Dual specificity protein phosphatase 3	DUSP3	20	No	No	Non-Cl	No				3.412E-04
Dynactin subunit 2	DCTN2	44	No	Yes	No	No		1.781E-04		
Dynactin subunit 5	DCTN5	20	No	No	No	No				1.592E-04
E3 ubiquitin-protein ligase HUWE1	HUWE1	482	No	Yes	No	No		2.581E-05	5.049E-05	
E3 ubiquitin-protein ligase RNF213	RNF213	591	No	Yes	No	No			6.624E-05	
Echinoderm microtubule-associated protein-like 2	EML2	71	Yes	No	No	No		3.653E-05	6.273E-05	
Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1	105	Yes	No	No	Yes		1.324E-04		
EF-hand domain-containing protein D1	EFHD1	27	No	No	No	No			3.659E-04	
EGF-containing fibulin-like extracellular matrix protein 2	EFEMP2	49	No	Yes	Yes	No				7.748E-05
EGF-like repeat and discoidin I-like domain-containing protein 3	EDIL3	54	No	Yes	Yes	No				5.993E-04
EH domain-containing protein 2	EHD2	61	No	Yes	No	No				7.468E-05
Electron transfer flavoprotein subunit alpha, mitochondrial	ETFA	35	No	Yes	No	No				1.049E-04
Ena/VASP-like protein	EVL	45	Yes	No	Non-Cl	No		1.497E-04		
Endoglin	ENG	68	No	Yes	Yes	Yes				7.792E-05
Endophilin-B2	SH3GLB2	44	No	No	No	No		1.948E-04		
Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	MAN1B1	80	No	No	No	Yes		1.495E-04		
Endothelial cell-selective adhesion molecule	ESAM	41	No	No	Yes	Yes				1.918E-04
Endothelial differentiation-related factor 1	EDF1	16	No	No	Non-Cl	No		4.810E-04		3.108E-04
Endothelial protein C receptor	PROCR	27	Yes	No	Yes	Yes	1.938E-03			6.790E-04
Ephrin type-A receptor 7	EPHA7	112	No	No	Yes	Yes		9.054E-05		
Ephrin type-B receptor 2	EPHB2	117	No	Yes	Yes	Yes		6.302E-05		3.844E-05
Ephrin type-B receptor 4	EPHB4	108	Yes	Yes	Yes	Yes		1.435E-04	1.173E-04	
Ephrin-A1	EFNA1	24	No	No	Yes	No		3.224E-04	1.013E-03	
Epidermal growth factor receptor	EGFR	134	Yes	Yes	Yes	Yes			3.159E-04	
Epidermal growth factor receptor kinase substrate 8-like protein 1	EPS8L1	80	No	Yes	No	No		2.651E-05		
Epidermal growth factor receptor kinase substrate 8-like protein 2	EPS8L2	81	No	Yes	No	No			8.106E-05	
Epiplakin	EPPK1	556	No	Yes	No	No		8.790E-05	1.092E-04	
Epithelial cell adhesion molecule	EPCAM	35	No	No	Yes	Yes		1.558E-03	7.810E-04	
Epsilon-sarcoglycan	SGCE	50	No	No	No	Yes	2.140E-04		7.976E-05	
Ester hydrolase C11orf54	C11orf54	35	No	Yes	No	No			9.306E-05	1.287E-04
Ethylmalonyl-CoA decarboxylase	ECHDC1	34	Yes	No	No	No				1.507E-04
Eukaryotic translation elongation factor 1 epsilon-1	EEF1E1	20	No	Yes	Non-Cl	No		1.910E-04		
Eukaryotic translation initiation factor 1A, X-chromosomal	EIF1AX	16	No	No	No	No				2.420E-04
Eukaryotic translation initiation factor 1b	EIF1B	13	No	No	Non-Cl	No				3.068E-04
Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	51	No	Yes	No	No		3.021E-04		1.953E-04
Eukaryotic translation initiation factor 3 subunit A	EIF3A	167	No	Yes	No	No		7.287E-05		2.490E-05
Eukaryotic translation initiation factor 3 subunit C-like protein	EIF3CL	105	No	No	No	No		7.292E-05		3.169E-05
Eukaryotic translation initiation factor 3 subunit E	EIF3E	52	No	Yes	No	No		1.819E-04		
Eukaryotic translation initiation factor 3 subunit F	EIF3F	38	No	No	Non-Cl	No		1.319E-04		1.458E-04
Eukaryotic translation initiation factor 3 subunit G	EIF3G	36	Yes	No	Non-Cl	No		1.198E-04	2.367E-04	
Eukaryotic translation initiation factor 3 subunit H	EIF3H	40	No	Yes	Non-Cl	No		1.209E-04		
Eukaryotic translation initiation factor 3 subunit L	EIF3L	67	No	No	No	No		1.255E-04		4.078E-05
Eukaryotic translation initiation factor 3 subunit M	EIF3M	43	No	No	No	No		2.422E-04		3.558E-04
Eukaryotic translation initiation factor 4E	EIF4E	25	No	Yes	Non-Cl	No		1.320E-04		1.594E-04
Eukaryotic translation initiation factor 5	EIF5	49	No	No	No	No		1.666E-04		
Exocyst complex component 4	EXOC4	111	No	Yes	No	No			3.528E-05	
Exportin-4	XPO4	130	No	No	Non-Cl	No		2.046E-05		
Exportin-7	XPO7	124	No	No	Non-Cl	No			4.226E-05	
Extended synaptotagmin-1	ESYT1	123	No	No	No	Yes		2.168E-05		1.049E-04
Extracellular serine/threonine protein kinase FAM20C	FAM20C	66	No	Yes	Yes	No				3.923E-05
Extracellular sulfatase Sulf-1	SULF1	101	No	No	Yes	No		4.345E-05		
Extracellular superoxide dismutase [Cu-Zn]	SOD3	26	Yes	Yes	Yes	No				3.350E-04
FACT complex subunit SPT16	SUPT16H	120	No	No	No	No		3.646E-05		1.271E-04

FACT complex subunit SSRP1	SSRP1	81	Yes	No	No	No				1.713E-04
Far upstream element-binding protein 2	KHSRP	73	No	No	No	No		2.673E-05	4.098E-05	
FAS-associated factor 1	FAF1	74	No	No	No	No			4.510E-05	
Fatty acid-binding protein, epidermal	FABP5	15	Yes	No	No	No	5.873E-03			1.191E-03
FERM, RhoGEF and pleckstrin domain-containing protein 1	FARP1	119	Yes	No	No	No			4.365E-05	
Ferritin family homolog 3	FERMT3	76	No	Yes	No	No				4.277E-05
Ferritin heavy chain	FTH1	21	No	Yes	No	No		5.782E-04		9.450E-04
Ferritin light chain	FTL	20	No	Yes	No	No				8.898E-04
Fibroblast growth factor receptor 4	FGFR4	88	Yes	No	Yes	Yes		3.551E-05	4.368E-05	
Fibulin-2	FBLN2	127	No	No	Yes	No			6.208E-04	
Filaggrin-2	FLG2	248	No	Yes	No	No		1.200E-05	1.437E-05	
Filamin-C	FLNC	291	No	Yes	No	No				8.137E-04
Flap endonuclease 1	FEN1	43	No	Yes	No	No		1.254E-04		1.829E-04
Flotillin-2	FLOT2	47	Yes	Yes	No	No		1.219E-04	5.440E-05	
FLYWCH family member 2	FLYWCH2	15	No	No	Non-Cl	No		1.709E-04		
Follistatin	FST	38	No	Yes	Yes	No	1.100E-02			2.182E-04
Follistatin-related protein 1	FSTL1	35	No	No	Yes	No	2.265E-03			1.328E-03
FRAS1-related extracellular matrix protein 2	FREM2	351	No	No	Yes	Yes		1.549E-04	4.390E-04	
Fumarate hydratase, mitochondrial	FH	55	No	Yes	Non-Cl	No			9.007E-05	
Gamma-enolase	ENO2	47	No	Yes	Non-Cl	No		2.581E-03		5.650E-03
GDNF family receptor alpha-1	GFRA1	51	Yes	Yes	Yes	Yes		1.429E-03		
Geranylgeranyl transferase type-2 subunit alpha	RABGGTA	65	No	No	No	No			5.083E-05	
Geranylgeranyl transferase type-2 subunit beta	RABGGTB	37	No	No	No	No			1.230E-04	
Gigaxonin	GAN	68	No	No	No	No		8.806E-05		
Glia-derived nexin	SERPINE2	44	No	Yes	Yes	No	7.843E-03			3.630E-04
Glucosamine 6-phosphate N-acetyltransferase	GNPNAT1	21	No	No	Non-Cl	No		3.351E-04	3.169E-04	
Glutamate dehydrogenase 1, mitochondrial	GLUD1	61	Yes	Yes	No	No			9.514E-05	6.245E-05
Glutamate--cysteine ligase regulatory subunit	GCLM	31	Yes	No	Non-Cl	No				1.688E-04
Glutaminyl-peptide cyclotransferase	QPCT	41	Yes	Yes	Yes	No	1.429E-03	4.487E-04		
Glutathione reductase, mitochondrial	GSR	56	No	Yes	No	No		1.909E-04	2.219E-04	
Glutathione S-transferase kappa 1	GSTK1	25	No	Yes	No	No				1.282E-04
Glutathione S-transferase Mu 1	GSTM1	26	Yes	Yes	No	No				6.072E-04
Glutathione S-transferase Mu 3	GSTM3	27	Yes	Yes	No	No			9.288E-04	4.877E-04
Glutathione S-transferase P	GSTP1	23	No	Yes	Non-Cl	No	6.440E-03			1.262E-03
Glutathione S-transferase theta-1	GSTT1	27	Yes	No	No	No		1.969E-04		
Glycerol-3-phosphate dehydrogenase 1-like protein	GPD1L	38	No	No	No	No		1.763E-04		
Glycogen debranching enzyme	AGL	175	No	No	No	No		3.431E-05	1.480E-04	
Glycogen phosphorylase, liver form	PYGL	97	No	Yes	No	No			1.029E-04	4.165E-04
Glycogenin-1	GYG1	39	No	No	No	No		5.430E-05		
Glycoprotein endo-alpha-1,2-mannosidase-like protein	MANEAL	51	No	No	Yes	Yes		9.347E-05		
Glycosaminoglycan xylosylkinase	FAM20B	46	No	No	Yes	Yes	1.691E-04			1.271E-04
Glycylpeptide N-tetradecanoyltransferase 1	NMT1	57	No	No	No	No				9.256E-05
Glypican-4	GPC4	62	Yes	Yes	Yes	Yes		4.259E-04		
Golgi resident protein GCP60	ACBD3	61	No	No	Non-Cl	No			1.092E-04	
Golgi to ER traffic protein 4 homolog	GET4	37	No	No	Non-Cl	No		5.812E-05	7.024E-05	
Golgin subfamily A member 3	GOLGA3	167	Yes	No	No	No		2.533E-05		
Growth arrest-specific protein 6	GAS6	80	No	Yes	Yes	No	6.971E-04			6.416E-05
Growth/differentiation factor 15	GDF15	34	Yes	Yes	Yes	No		9.811E-04		
Growth-regulated alpha protein	CXCL1	11	Yes	No	Yes	No	8.819E-03			2.149E-04
GTP:AMP phosphotransferase AK3, mitochondrial	AK3	26	No	No	Non-Cl	No				4.586E-04
GTP-binding protein Rheb	RHEB	20	Yes	Yes	Non-Cl	No			1.265E-04	4.422E-04
Guanidinoacetate N-methyltransferase	GAMT	26	No	No	Non-Cl	No				1.462E-04
Guanine nucleotide exchange factor VAV2	VAV2	101	No	No	No	No		4.304E-05		
Guanine nucleotide-binding protein subunit alpha-11	GNA11	42	Yes	Yes	No	No			1.288E-04	
Guanine nucleotide-binding protein subunit alpha-13	GNA13	44	No	Yes	No	No		1.264E-04	3.888E-04	
Guanylate kinase	GUK1	22	No	No	Non-Cl	No			2.075E-04	1.453E-04
HEAT repeat-containing protein 6	HEATR6	129	No	No	No	No		4.771E-05		
Heat shock 70 kDa protein 4L	HSPA4L	95	No	No	No	No			1.099E-04	
Hemicentin-1	HMCN1	613	No	No	Yes	No	9.492E-05	1.772E-05		
Hemoglobin subunit alpha	HBA1	15	No	Yes	No	No	1.932E-03	4.319E-04		
Hemopexin	HPX	52	No	Yes	Yes	No	1.203E-02		1.001E-04	
Heparan-sulfate 6-	HS6ST1	48	No	No	Yes	Yes		1.622E-04	1.134E-04	
Hepatocyte growth factor activator	HGFAC	71	No	No	Yes	No			4.400E-05	
HERV-MER_4q12 provirus ancestral Env polyprotein	ERVMER34-1	64	No	No	Yes	Yes	6.110E-04	8.394E-05		
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	37	No	No	No	No				2.316E-04
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	78	No	No	No	No		7.096E-05		1.030E-04
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	71	No	No	No	No			6.304E-05	1.929E-04
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	91	No	No	No	No		5.104E-05		
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	85	No	No	No	No			6.257E-05	
Hexokinase-2	HK2	102	Yes	No	No	No			1.145E-04	4.356E-05
High mobility group protein B2	HMGB2	24	No	No	No	No		6.807E-04		8.572E-04
High mobility group protein B3	HMGB3	23	Yes	No	No	No		4.764E-04		1.448E-04
Hippocalcin-like protein 1	HPCAL1	22	No	Yes	Non-Cl	No				5.085E-04
Histidine triad nucleotide-binding protein 1	HINT1	14	No	Yes	Non-Cl	No		4.898E-04		6.396E-04
Histone H3.1	HIST1H3A	15	No	Yes	Non-Cl	No				4.228E-04
Histone-binding protein RBBP4	RBBP4	48	No	No	Non-Cl	No		1.331E-04		
Histone-binding protein RBBP7	RBBP7	48	Yes	No	Non-Cl	No		1.228E-04		
Histone-lysine N-methyltransferase SETD7	SETD7	41	No	No	Non-Cl	No		6.478E-05		

Hyaluronan and proteoglycan link protein 3	HAPLN3	41	No	Yes	Yes	No				1.292E-04
Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	HADH	34	No	No	No	No			1.665E-04	4.023E-04
Hydroxyacylglutathione hydrolase, mitochondrial	HAGH	29	Yes	No	Non-Cl	No			9.518E-05	
ICOS ligand	ICOSLG	33	No	No	Yes	Yes		2.803E-04	1.338E-04	
Iduronate 2-sulfatase	IDS	62	Yes	No	Yes	No			4.265E-05	
Immunity-related GTPase family Q protein	IRGQ	63	No	No	No	No			8.392E-05	4.594E-05
Immunoglobulin superfamily member 3	IGSF3	135	No	Yes	Yes	Yes		1.421E-04	2.350E-04	
Importin-4	IPO4	119	No	No	No	No		9.676E-05		3.687E-05
Inhibin alpha chain	INHA	40	No	No	Yes	No			1.758E-04	
Inorganic pyrophosphatase 2, mitochondrial	PPA2	38	No	No	No	No		7.099E-05		
Inosine-5'-monophosphate dehydrogenase 1	IMPDH1	55	No	No	No	No			9.059E-05	
Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	56	Yes	Yes	No	No			1.927E-04	1.793E-04
Inositol monophosphatase 1	IMPA1	30	No	No	Non-Cl	No				2.294E-04
Inositol monophosphatase 3	IMPAD1	39	No	No	No	Yes				2.247E-04
Inositol-3-phosphate synthase 1	ISYNA1	61	No	No	Non-Cl	No			9.338E-05	
Insulin-degrading enzyme	IDE	118	No	No	Yes	Yes			3.372E-05	6.245E-05
Insulin-like growth factor 1 receptor	IGF1R	155	Yes	No	Yes	Yes		1.739E-05		
Insulin-like growth factor-binding protein 1	IGFBP1	28	No	No	Yes	No				7.334E-04
Insulin-like growth factor-binding protein 5	IGFBP5	31	Yes	No	Yes	No		7.332E-04	9.120E-03	
Insulin-like growth factor-binding protein 6	IGFBP6	25	Yes	Yes	Yes	No	2.151E-03			1.898E-04
Insulin-like growth factor-binding protein-like 1	IGFBPL1	29	No	No	Yes	No		2.738E-04		4.762E-04
Integrin alpha-5	ITGA5	115	No	Yes	Yes	Yes		4.980E-05		4.945E-05
Integrin beta-4	ITGB4	202	Yes	Yes	Yes	Yes	2.355E-04	3.396E-05		
Integrin beta-5	ITGB5	88	Yes	Yes	Yes	Yes			6.583E-05	
Integrin beta-6	ITGB6	86	No	No	Yes	Yes		6.025E-05	2.955E-05	
Intercellular adhesion molecule 1	ICAM1	58	Yes	Yes	Yes	Yes		3.572E-05		
Intercellular adhesion molecule 5	ICAM5	97	Yes	Yes	Yes	Yes		2.590E-05		
Interferon-induced transmembrane protein 3	IFITM3	15	Yes	Yes	No	Yes		2.882E-04		6.521E-04
Interleukin-11	IL11	21	No	Yes	Yes	No				6.669E-04
Interleukin-6 receptor subunit alpha	IL6R	40	No	No	Yes	Yes			6.159E-05	
Interstitial collagenase	MMP1	54	Yes	Yes	No	No	2.782E-03			7.367E-04
Involucrin	IVL	68	No	Yes	No	No			9.156E-04	
Isoaspartyl peptidase/L-asparaginase	ASRGL1	32	No	No	No	No			9.415E-05	
Isochorismatase domain-containing protein 1	ISOC1	32	No	No	No	No		4.618E-04	2.329E-04	
Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	51	Yes	No	Non-Cl	No			9.007E-05	
Isoform 1 of Four and a half LIM domains protein 1	FHL1	32	Yes	No	Non-Cl	No				1.234E-04
Isoform 2 of Afadin	MLLT4	206	Yes	Yes	No	No		2.591E-05	9.858E-05	
Isoform 2 of Alpha-aminoadipic semialdehyde dehydrogenase	ALDH7A1	55	No	Yes	Non-Cl	No			1.405E-04	6.400E-05
Isoform 2 of Dehydrogenase/reductase SDR family member 2, mitochondrial	DHRS2	31	Yes	No	No	No			2.728E-04	
Isoform 2 of Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	GSPT1	69	No	No	No	No			4.603E-05	
Isoform 2 of Extracellular matrix protein FRAS1	FRAS1	444	No	No	Yes	Yes	4.108E-05		2.622E-04	
Isoform 2 of Extracellular sulfatase Sulf-2	SULF2	100	Yes	No	Yes	No	2.115E-04	1.113E-03		
Isoform 2 of Isopentenyl-diphosphate Delta-isomerase 1	IDI1	32	Yes	No	No	No		1.245E-04		2.788E-04
Isoform 2 of Latent-transforming growth factor beta-binding protein 4	LTBP4	166	No	No	Yes	No	1.038E-04			2.130E-05
Isoform 2 of Matrilin-2	MATN2	105	No	No	Yes	No		1.347E-04		5.463E-05
Isoforms 1/2/3/5	MACF1	620	No	No	No	No		5.114E-06		6.226E-06
Isoform 2 of Nebulette	NEBL	31	Yes	Yes	No	No		6.535E-05		
Isoform 2 of Poly(U)-binding-splicing factor PUF60	PUF60	58	No	No	No	No		1.026E-04		9.343E-05
Isoform 2 of Protein unc-45 homolog A	UNC45A	102	No	Yes	Non-Cl	No		1.364E-04		
Isoform 2 of Tumor protein D52	TPD52	20	Yes	No	Non-Cl	No		1.677E-04	2.018E-03	
Isoform 3 of Leucine-rich repeat flightless-interacting protein 1	LRRFIP1	83	No	No	No	No		2.914E-05		
Isoform 3 of Polyadenylate-binding protein-interacting protein 1	PAIP1	42	No	No	Non-Cl	No		1.690E-04		
Isoform A of Endothelin-converting enzyme 1	ECE1	86	Yes	Yes	No	Yes		6.735E-05		
Isoleucine--tRNA ligase, cytoplasmic	IARS	145	Yes	Yes	No	No		9.399E-05		5.947E-05
Isoleucine--tRNA ligase, mitochondrial	IARS2	114	No	No	No	No				3.426E-05
Kallikrein-6	KLK6	27	Yes	No	Yes	Yes	1.973E-03	1.555E-04		
Keratin, type I cytoskeletal 19	KRT19	44	Yes	Yes	Non-Cl	No		6.808E-03	6.801E-03	
Keratinocyte proline-rich protein	KPRP	64	No	Yes	No	No		1.401E-04	1.384E-04	
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	48	No	No	No	No				7.728E-05
KIF1-binding protein	KIAA1279	72	No	No	No	No			6.606E-05	
Kinesin-1 heavy chain	KIF5B	110	Yes	No	No	No		2.311E-04		2.827E-04
Kinesin-like protein KIF23	KIF23	110	Yes	No	Non-Cl	No		2.025E-04		
Kinetochore-associated protein 1	KNTC1	251	No	No	No	No				2.627E-05
Kit ligand	KITLG	31	No	No	Yes	Yes		4.000E-04		
Kynurenine--oxoglutarate transaminase 3	CCBL2	51	No	No	Non-Cl	No		6.272E-05	1.036E-04	
Lactadherin	MFGE8	43	Yes	Yes	Yes	No	5.557E-04			3.876E-04
Lactosylceramide 4-alpha-galactosyltransferase	A4GALT	41	No	No	No	Yes		1.736E-04		
Lactotransferrin	LTF	78	Yes	Yes	Yes	No				3.394E-04
Lamin-B2	LMNB2	68	No	No	No	No			1.656E-04	3.679E-04
Laminin subunit alpha-2	LAMA2	344	No	No	Yes	No			7.357E-06	
Laminin subunit alpha-4	LAMA4	203	No	Yes	Yes	No				3.505E-05
Laminin subunit gamma-2	LAMC2	131	Yes	Yes	Yes	No	1.479E-02	1.607E-05		

L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase	AASDHPPT	36	No	Yes	Non-Cl	No			1.513E-04	
LanC-like protein 2	LANCL2	51	No	No	No	No			7.745E-05	
Large proline-rich protein BAG6	BAG6	119	No	No	No	No		3.372E-05	2.057E-05	
Latent-transforming growth factor beta-binding protein 2	LTBP2	195	No	Yes	Yes	No	8.795E-04			2.526E-05
Latexin	LXN	26	No	No	No	No		5.792E-04	2.407E-04	
Latrophilin-1	LPHN1	163	No	No	Yes	Yes		1.320E-04		
Lethal(2) giant larvae protein homolog 2	LLGL2	113	No	No	No	No		1.067E-04	2.175E-04	
Leucine zipper transcription factor-like protein 1	LZTFL1	35	Yes	No	Non-Cl	No				7.663E-05
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	158	No	Yes	No	No				4.175E-05
Leucine-rich repeat-containing protein 47	LRRC47	63	No	No	Non-Cl	No		1.067E-04		
Leucine-rich repeat-containing protein 59	LRRC59	35	No	No	No	Yes		9.328E-05		1.115E-04
Leucine--tRNA ligase, cytoplasmic	LARS	134	No	No	No	No		2.808E-05		
LIM domain and actin-binding protein 1	LIMA1	85	No	Yes	No	No			3.862E-05	
Lipopolysaccharide-responsive and beige-like anchor protein	LRBA	319	Yes	No	No	Yes			2.235E-05	
Low affinity cationic amino acid transporter 2	SLC7A2	72	No	No	No	Yes		5.018E-05		
Low molecular weight phosphotyrosine protein phosphatase	ACP1	18	No	Yes	Non-Cl	No		4.270E-04	2.175E-04	
Ly6/PLAUR domain-containing protein 3	LYPD3	36	No	No	Yes	Yes		8.480E-04	4.411E-04	
Lysosomal acid phosphatase	ACP2	48	No	Yes	Yes	Yes			6.930E-05	
Lysosomal alpha-mannosidase	MAN2B1	114	No	No	Yes	No			9.247E-05	2.344E-04
Lysosome-associated membrane glycoprotein 2	LAMP2	45	Yes	Yes	Yes	Yes				1.554E-04
Lysyl oxidase homolog 4	LOXL4	84	No	Yes	Yes	No				1.138E-04
Macrophage colony-stimulating factor 1	CSF1	60	No	Yes	Yes	Yes				5.327E-04
Macrophage colony-stimulating factor 1 receptor	CSF1R	108	Yes	No	Yes	Yes			1.674E-04	
MAGUK p55 subfamily member 6	MPP6	61	No	No	Non-Cl	No			7.448E-05	
Major vault protein	MVP	99	Yes	Yes	No	No			5.799E-05	3.883E-05
Maleylacetoacetate isomerase	GSTZ1	24	Yes	No	Non-Cl	No		3.296E-04		
MAM domain-containing protein 2	MAMDC2	78	No	Yes	Yes	No				7.664E-05
Mammalian ependymin-related protein 1	EPDR1	25	No	No	Yes	No				6.462E-04
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	MAN1A1	73	No	Yes	Yes	Yes	1.308E-04	3.416E-04		
MARCKS-related protein	MARCKSL1	20	No	Yes	No	No		3.185E-04	8.353E-04	
Matrilin-3	MATN3	53	No	No	Yes	No		5.859E-05		
Matrix Gla protein	MGP	12	No	Yes	Yes	No		3.206E-04		
Matrix metalloproteinase-14	MMP14	66	Yes	No	Yes	Yes	8.619E-05			2.072E-04
Melanotransferrin	MFI2	80	No	Yes	Yes	No	2.483E-04			8.564E-05
Membrane-bound transcription factor site-1 protease	MBTPS1	118	Yes	No	Yes	Yes		9.921E-05		
Metalloproteinase inhibitor 4	TIMP4	26	Yes	No	Yes	No				1.289E-04
Metastasis-suppressor KiSS-1	KISS1	15	No	No	Yes	No				6.247E-04
Methionine--tRNA ligase, cytoplasmic	MARS	101	No	Yes	Non-Cl	No			3.202E-05	
Methylthioribulose-1-phosphate dehydratase	APIP	27	No	No	No	No			1.682E-04	1.908E-04
MHC class I polypeptide-related sequence A	MICA	43	Yes	Yes	Yes	Yes		7.335E-05		
Microfibrillar-associated protein 2	MFAP2	21	Yes	No	Yes	No			2.884E-04	7.545E-04
Microtubule-associated protein RP/EB family member 1	MAPRE1	30	No	No	No	No		1.772E-04		2.807E-04
Midkine	MDK	16	No	No	Yes	No	5.786E-04	9.983E-04		
Mitochondrial import inner membrane translocase subunit Tim13	TIMM13	11	No	No	Non-Cl	No			2.418E-04	
Mitogen-activated protein kinase 14	MAPK14	34	No	No	No	No				1.122E-04
Mitotic spindle assembly checkpoint protein MAD2A	MAD2L1	24	Yes	No	Non-Cl	No				1.691E-04
MOB kinase activator 1A	MOB1A	25	No	No	Non-Cl	No		3.286E-04		5.084E-04
MOB-like protein phocein	MOB4	26	No	No	Non-Cl	No		2.339E-04		1.802E-04
Moesin	MSN	68	Yes	Yes	No	No	6.480E-04			1.746E-03
Monocarboxylate transporter 4	SLC16A3	49	Yes	Yes	No	Yes			2.133E-04	1.992E-04
MRG/MORF4L-binding protein	MRGBP	22	No	No	No	No		9.395E-05		
Mth938 domain-containing protein	AAMDC	13	No	No	No	No		1.571E-04		2.835E-04
Mucin-1	MUC1	122	Yes	Yes	Yes	Yes		3.391E-05		
Mucin-16	MUC16	2353	No	Yes	No	Yes			1.179E-05	
Mucin-5B	MUC5B	596	No	No	Yes	No		3.066E-05		
Multiple epidermal growth factor-like domains protein 8	MEGF8	303	No	Yes	Yes	Yes		1.161E-05	1.654E-05	
Multiple inositol polyphosphate phosphatase 1	MINPP1	55	No	No	Yes	Yes	8.402E-05	2.920E-04		
Multivesicular body subunit 12A	MVB12A	29	No	No	Non-Cl	No		1.043E-04	8.593E-05	
Myeloid-associated differentiation marker	MYADM	35	No	Yes	No	Yes			7.133E-05	
Myoferlin	MYOF	235	Yes	No	No	Yes				1.568E-04
Myosin-14	MYH14	228	No	Yes	No	No		5.956E-05	1.985E-04	
Na(+)/H(+) exchange regulatory cofactor NHE-RF2	SLC9A3R2	37	No	Yes	Non-Cl	No			4.413E-04	
N-acetylgalactosamine kinase	GALK2	50	No	No	No	No		5.093E-05		
N-acetylgalactosamine-6-sulfatase	GALNS	58	No	No	Yes	No	1.278E-04	1.640E-04		
N-acetylglucosamine-1-phosphotransferase subunit gamma	GNPTG	34	No	Yes	Yes	No		1.090E-04	9.669E-05	
N-acetylmuramoyl-L-alanine amidase	PGLYRP2	62	No	No	Yes	No	2.377E-04	5.771E-05		
N-acetylserotonin O-methyltransferase-like protein	ASMTL	69	No	No	No	No		3.024E-05		
NAD(P)H-hydrate epimerase	APOA1BP	32	No	No	Yes	No			3.218E-04	9.937E-05
NADP-dependent malic enzyme	ME1	64	No	Yes	No	No		2.969E-04	1.011E-04	
NADPH--cytochrome P450 reductase	POR	77	No	No	No	Yes			8.494E-05	
N-alpha-acetyltransferase 25, NatB auxiliary subunit	NAA25	112	No	No	No	No				6.538E-05
N-alpha-acetyltransferase 50	NAA50	19	No	No	No	No				2.711E-04
NEDD8-activating enzyme E1 catalytic subunit	UBA3	52	No	No	Non-Cl	No			6.263E-05	

NEDD8-activating enzyme E1 regulatory subunit	NAE1	60	No	No	Non-Cl	No			5.397E-05	7.610E-05
Nephronectin	NPNT	62	No	Yes	Yes	No		6.819E-04		
Netrin receptor UNC5C	UNC5C	103	Yes	No	Yes	Yes		2.059E-05		
Neural cell adhesion molecule 2	NCAM2	93	No	No	Yes	Yes		4.874E-04		
Neurogenic locus notch homolog protein 3	NOTCH3	244	No	No	Yes	Yes		6.984E-05	4.520E-05	
Neuroligin-2	NLGN2	91	No	No	Yes	Yes		8.506E-05		2.754E-05
Neurolysin, mitochondrial	NLN	81	No	No	No	No		6.790E-05		
Neurosecretory protein VGF	VGF	67	No	No	Yes	No		1.926E-04		
Neuroserpin	SERPINI1	46	Yes	No	Yes	No	1.439E-04		2.379E-04	
Neutral amino acid transporter A	SLC1A4	56	No	Yes	No	Yes		6.279E-05		
Nicastrin	NCSTN	78	No	Yes	Yes	Yes		7.369E-05		4.817E-05
Nicotinamide phosphoribosyltransferase	NAMPT	56	No	Yes	No	No		7.762E-05		2.126E-04
Nicotinate phosphoribosyltransferase	NAPRT1	58	No	Yes	No	No			4.741E-04	
Nicotinate-nucleotide pyrophosphorylase [carboxylating]	QPRT	31	Yes	Yes	Yes	No		1.438E-04	5.675E-04	
Niemann-Pick C1 protein	NPC1	142	No	Yes	Yes	Yes				5.365E-05
Noelin	OLFM1	55	Yes	No	Yes	No		1.601E-04		
Non-histone chromosomal protein HMG-14	HMGN1	11	Yes	No	No	No		7.967E-04		
Non-POU domain-containing octamer-binding protein	NONO	54	No	Yes	No	No		1.616E-04		
Non-specific lipid-transfer protein	SCP2	15	No	No	Non-Cl	No		5.206E-05		
N-sulphoglucosamine sulphohydrolase	SGSH	57	No	No	Yes	No			4.611E-04	1.729E-04
N-terminal Xaa-Pro-Lys N-methyltransferase 1	NTMT1	25	No	No	No	No				1.031E-04
Nuclear autoantigenic sperm protein	NASP	85	Yes	No	No	No		3.614E-05		
Nuclear cap-binding protein subunit 1	NCBP1	92	No	No	No	No			2.907E-05	
Nuclear receptor 2C2-associated protein	NR2C2AP	16	No	No	Non-Cl	No			2.109E-04	3.297E-04
Nuclear receptor coactivator 5	NCOA5	66	No	No	No	No				5.973E-05
Nuclear receptor-binding protein	NRBP1	60	No	No	No	No		7.104E-05		
Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	NUCKS1	27	No	No	No	No		1.774E-04		
Nucleophosmin	NPM1	33	No	Yes	Non-Cl	No		4.479E-04		6.452E-04
Nucleoside diphosphate kinase 3	NME3	19	Yes	No	Yes	No		1.094E-03	2.747E-04	
Nucleotide exchange factor SIL1	SIL1	52	No	No	Yes	No	4.022E-04	2.890E-04		
Occludin	OCLN	59	No	No	No	Yes		5.529E-05		
Olfactomedin-like protein 2A	OLFML2A	73	No	No	Yes	No				5.264E-05
Olfactomedin-like protein 3	OLFML3	46	No	No	Yes	No				3.833E-04
Oligoribonuclease, mitochondrial	REXO2	27	Yes	No	Non-Cl	No	2.193E-04			2.185E-04
Opioid growth factor receptor	OGFR	73	No	No	No	No			4.257E-05	
Ovarian cancer-associated gene 2 protein	OVCA2	24	No	No	Non-Cl	No				2.285E-04
Oxysterol-binding protein 1	OSBP	89	Yes	No	No	No			8.616E-05	5.689E-05
Paraspeckle component 1	PSPC1	59	No	No	No	No		5.445E-05		4.447E-05
PCTP-like protein	STARD10	33	No	No	No	No		8.931E-04	3.991E-04	
PDZ and LIM domain protein 7	PDLIM7	50	No	Yes	No	No				7.492E-05
PDZ domain-containing protein GIPC1	GIPC1	36	No	Yes	Non-Cl	No		3.146E-04	1.741E-04	
Pentraxin-related protein PTX3	PTX3	42	Yes	Yes	Yes	No				2.495E-03
Peptidyl-prolyl cis-trans isomerase FKBP10	FKBP10	64	No	No	Yes	No			1.696E-04	
Peptidyl-prolyl cis-trans isomerase FKBP2	FKBP2	16	No	No	Yes	No			7.845E-04	
Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	PIN1	18	Yes	No	Non-Cl	No				1.772E-04
Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4	PIN4	17	No	No	Non-Cl	No		2.897E-04		3.960E-04
Perilipin-3	PLIN3	47	No	No	Non-Cl	No		2.307E-04		3.997E-04
Peripheral plasma membrane protein CASK	CASK	105	No	No	No	Yes		6.109E-05		
Periplakin	PPL	205	No	Yes	No	No		8.099E-05	2.356E-05	
Persulfide dioxygenase ETHE1, mitochondrial	ETHE1	28	No	No	No	No	2.010E-04			1.137E-04
PEST proteolytic signal-containing nuclear protein	PCNP	19	No	No	Non-Cl	No				2.919E-04
Phenylalanine--tRNA ligase alpha subunit	FARSA	58	No	No	No	No		1.027E-04		
Phenylalanine--tRNA ligase beta subunit	FARSB	66	No	No	No	No		5.579E-05		
Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein	PREX1	186	Yes	No	No	No			1.737E-05	
Phosphatidylinositol transfer protein alpha isoform	PITPNA	32	No	No	No	No			8.688E-05	
Phosphatidylinositol-binding clathrin assembly protein	PICALM	71	Yes	No	No	No		1.598E-04		1.676E-04
Phosphoglucomutase-like protein 5	PGM5	62	No	No	No	No			1.344E-04	
Phosphoglycolate phosphatase	PGP	34	No	No	No	No		2.242E-04	2.527E-04	
Phospholipase D3	PLD3	55	No	Yes	No	Yes		5.811E-05		8.276E-05
Phospholipid transfer protein	PLTP	55	Yes	Yes	Yes	No	3.295E-04		8.236E-04	
Phosphomannomutase 2	PMM2	28	No	No	Non-Cl	No		3.297E-04	2.814E-04	
Phosphopantothenate--cysteine ligase	PPCS	34	No	No	No	No		1.074E-04	1.686E-04	
Phosphoribosyl pyrophosphate synthase-associated protein 2	PRPSAP2	41	No	No	No	No		6.426E-05		1.568E-04
Phosphoribosyltransferase domain-containing protein 1	PRTFDC1	26	No	No	No	No				1.806E-04
Phosphoserine aminotransferase	PSAT1	40	Yes	Yes	No	No			1.748E-04	5.751E-04
Pirin	PIR	32	Yes	No	No	No		1.971E-04		
Plakophilin-3	PKP3	87	No	No	No	No			8.724E-05	
Plasma serine protease inhibitor	SERPINA5	46	Yes	Yes	Yes	No		1.590E-03		
Plastin-1	PLS1	70	No	Yes	No	No		1.449E-04		
Plastin-2	LCP1	70	Yes	Yes	No	No			1.067E-03	2.849E-04
Platelet-derived growth factor C	PDGFC	39	No	Yes	Yes	No	3.432E-04			1.501E-04
Platelet-derived growth factor D	PDGFD	43	No	No	Yes	No			4.860E-04	
Plexin domain-containing protein 2	PLXDC2	60	No	No	Yes	Yes		4.401E-04		
Plexin-B1	PLXNB1	232	No	No	Yes	Yes			7.645E-05	
Plexin-D1	PLXND1	212	No	No	Yes	Yes		1.704E-05		
Poliovirus receptor-related protein 2	PVRL2	51	No	No	Yes	Yes		9.598E-05	8.655E-05	

Poliovirus receptor-related protein 4	PVRL4	55	No	No	Yes	Yes		6.129E-04	9.928E-04	
Poly [ADP-ribose] polymerase 1	PARP1	113	No	No	No	No			6.346E-05	1.421E-04
Polyadenylate-binding protein 2	PABPN1	33	No	No	No	No		6.138E-05		
Polymerase delta-interacting protein 2	POLDIP2	42	No	No	Non-Cl	No				7.871E-05
Polymeric immunoglobulin receptor	PIGR	83	No	Yes	Yes	Yes			9.826E-05	
Polypeptide N-acetylgalactosaminyltransferase 1	GALNT1	64	No	No	No	Yes	1.657E-04		1.340E-04	
Polypeptide N-acetylgalactosaminyltransferase 18	GALNT18	70	No	No	No	Yes			9.578E-05	
Polypeptide N-acetylgalactosaminyltransferase 6	GALNT6	71	No	No	Yes	Yes		2.731E-04	1.692E-04	
Porphobilinogen deaminase	HMBS	39	Yes	No	No	No		9.315E-05		
Pre-mRNA-processing factor 19	PRPF19	55	No	No	No	No		1.217E-04	5.816E-05	
Pre-mRNA-processing-splicing factor 8	PRPF8	274	No	No	No	No				3.251E-05
Probable aminopeptidase NPEPL1	NPEPL1	56	No	No	No	No		1.096E-04		
Probable cytosolic iron-sulfur protein assembly protein CIAO1	CIAO1	38	No	No	Non-Cl	No			1.028E-04	8.519E-05
Probable serine carboxypeptidase CPVL	CPVL	54	No	Yes	Yes	No			4.749E-04	7.339E-05
Programmed cell death protein 10	PDCD10	25	Yes	Yes	Non-Cl	No		2.221E-04	1.360E-04	
Prolow-density lipoprotein receptor-related protein 1	LRP1	505	Yes	Yes	Yes	Yes				1.365E-04
Prolyl 3-hydroxylase 1	LEPRE1	83	No	Yes	Yes	No	4.637E-05			1.175E-04
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	61	No	No	Yes	No		1.255E-04	2.177E-04	
Prolyl 4-hydroxylase subunit alpha-2	P4HA2	61	No	No	Yes	No		3.582E-05		
Prominin-2	PROM2	92	No	Yes	Yes	Yes		1.378E-04	2.445E-04	
Proprotein convertase subtilisin/kexin type 6	PCSK6	106	Yes	No	No	No		7.839E-05		
Proprotein convertase subtilisin/kexin type 9	PCSK9	74	No	Yes	Yes	No	2.543E-03			3.081E-04
Prostaglandin reductase 1	PTGR1	36	No	Yes	No	No			7.130E-05	1.219E-04
Prostatic acid phosphatase	ACPP	45	Yes	Yes	Yes	Yes		8.612E-05		
Proteasomal ubiquitin receptor ADRM1	ADRM1	42	No	No	No	No		7.011E-05		7.117E-05
Proteasome activator complex subunit 4	PSME4	211	Yes	No	No	No			1.273E-05	
Proteasome assembly chaperone 1	PSMG1	33	No	No	No	No			1.007E-04	1.801E-04
Proteasome subunit beta type-8	PSMB8	30	Yes	Yes	Non-Cl	No			8.055E-04	4.997E-04
Protein arginine N-methyltransferase 5	PRMT5	73	No	No	No	No			4.552E-05	
Protein C10	C12orf57	13	No	No	Non-Cl	No			2.327E-04	
Protein CutA	CUTA	19	No	Yes	No	Yes		8.707E-04		
Protein DDI1 homolog 2	DDI2	45	No	No	Non-Cl	No			7.224E-05	
Protein enabled homolog	ENAH	67	Yes	No	No	No			4.930E-05	
Protein ERGIC-53	LMAN1	58	No	Yes	Yes	Yes				1.819E-04
Protein flightless-1 homolog	FLII	145	Yes	No	No	No			4.173E-05	
Protein ITFG3	ITFG3	60	No	Yes	No	Yes		1.720E-04		
Protein kinase C and casein kinase substrate in neurons protein 3	PACSIN3	48	Yes	No	No	No		1.911E-04		
Protein lifeguard 3	TMBIM1	35	No	Yes	No	Yes			1.859E-04	
Protein MEMO 1	MEMO1	34	No	No	Non-Cl	No		1.125E-04	1.163E-04	
Protein MON2 homolog	MON2	190	No	No	No	No			2.039E-05	
Protein Niban	FAM129A	103	No	Yes	No	No				9.383E-05
Protein OS-9	OS9	76	No	Yes	Yes	No		1.782E-04		
Protein phosphatase 1 regulatory subunit 12A	PPP1R12A	115	No	No	No	No		2.302E-05		5.582E-05
Protein phosphatase 1 regulatory subunit 7	PPP1R7	42	No	Yes	Non-Cl	No		2.368E-04		
Protein phosphatase 1G	PPM1G	59	Yes	No	No	No		2.172E-04	1.059E-04	
Protein phosphatase inhibitor 2	PPP1R2	23	No	No	Non-Cl	No		1.622E-04		
Protein phosphatase methylesterase 1	PPME1	42	No	Yes	Non-Cl	No		1.107E-04		1.203E-04
Protein S100-A13	S100A13	11	Yes	Yes	No	No		3.392E-04	4.637E-04	
Protein S100-A4	S100A4	12	Yes	Yes	Non-Cl	No			1.040E-03	
Protein S100-A6	S100A6	10	Yes	Yes	Non-Cl	No			5.899E-04	
Protein S100-A8	S100A8	11	Yes	Yes	No	No	1.314E-03		2.185E-03	
Protein S100-P	S100P	10	No	Yes	No	No			1.711E-03	
Protein SEC13 homolog	SEC13	36	No	Yes	Non-Cl	No		1.623E-04		1.082E-04
Protein sidekick-1	SDK1	242	No	No	No	Yes		2.343E-04		
Protein transport protein Sec24A	SEC24A	120	No	No	No	Yes		3.978E-05		
Protein transport protein Sec31A	SEC31A	133	No	No	No	No		9.273E-05		
Protein tweety homolog 3	TTYH3	58	No	Yes	No	Yes		9.978E-05		
Protein-arginine deiminase type-2	PADI2	76	Yes	Yes	Non-Cl	No			2.016E-04	
Protein-glutamine gamma-glutamyltransferase 2	TGM2	77	Yes	Yes	Non-Cl	No				4.118E-04
Proteoglycan 4	PRG4	151	No	Yes	Yes	No			3.259E-05	
Protocadherin-1	PCDH1	115	Yes	No	Yes	Yes		1.605E-04	2.033E-04	
Protocadherin-20	PCDH20	105	No	No	No	Yes			5.464E-05	
Protocadherin-7	PCDH7	116	No	No	Yes	Yes		3.948E-04		
Pseudouridine-5'-monophosphatase	HDHD1	25	No	No	No	No				1.767E-04
Putative deoxyribose-phosphate aldolase	DERA	35	No	Yes	No	No		5.906E-05		
Putative gamma-glutamyltranspeptidase 3	GGT3P	62	No	No	Yes	Yes			5.192E-05	
helicase DHX15	DHX15	91	No	No	No	No		2.982E-05		
Putative RNA-binding protein 3	RBM3	17	No	Yes	Non-Cl	No				2.225E-04
Pyridoxal phosphate phosphatase	PDXP	32	No	No	No	No				1.358E-04
Pyridoxine-5'-phosphate oxidase	PNPO	30	No	No	Non-Cl	No		2.362E-04	1.354E-04	
Pyrroline-5-carboxylate reductase 3	PYCRL	29	No	No	Non-Cl	No		2.070E-04		
Quinone oxidoreductase	CRYZ	35	No	Yes	Non-Cl	No			3.208E-04	1.587E-04
Quinone oxidoreductase PIG3	TP53I3	36	No	No	Non-Cl	No	2.142E-04	1.143E-04		
Rac GTPase-activating protein 1	RACGAP1	71	No	Yes	No	No		4.541E-05		
Ran-binding protein 3	RANBP3	60	No	No	No	No			8.188E-05	6.100E-05
Ran-specific GTPase-activating protein	RANBP1	23	No	No	Non-Cl	No				1.751E-04
Rap1 GTPase-GDP dissociation stimulator 1	RAP1GDS1	66	No	Yes	Non-Cl	No			6.736E-05	

Ras suppressor protein 1	RSU1	32	Yes	Yes	Non-Cl	No			1.052E-04	
Ras-related protein Rab-13	RAB13	23	No	Yes	Non-Cl	No			3.434E-04	
Ras-related protein Rab-25	RAB25	23	No	Yes	No	No			1.353E-04	
Ras-related protein Rab-32	RAB32	25	Yes	No	Non-Cl	No				1.034E-04
Ras-related protein Rab-5B	RAB5B	24	No	Yes	Non-Cl	No			8.690E-04	4.846E-04
Ras-related protein R-Ras	RRAS	23	No	Yes	Non-Cl	No		2.188E-04		1.586E-04
Ras-related protein R-Ras2	RRAS2	23	Yes	Yes	Non-Cl	No				1.420E-04
Receptor tyrosine-protein kinase erbB-2	ERBB2	138	Yes	Yes	Yes	Yes			1.551E-03	
Receptor tyrosine-protein kinase erbB-4	ERBB4	146	Yes	No	Yes	Yes		1.453E-05		
Receptor-type tyrosine-protein phosphatase alpha	PTPRA	91	No	Yes	Yes	Yes			7.271E-05	
Receptor-type tyrosine-protein phosphatase eta	PTPRJ	146	Yes	Yes	Yes	Yes		1.064E-04		
Receptor-type tyrosine-protein phosphatase gamma	PTPRG	162	Yes	No	No	Yes		1.574E-04		8.826E-05
Receptor-type tyrosine-protein phosphatase U	PTPRU	162	Yes	No	Yes	Yes		1.640E-05	6.848E-05	
Regulation of nuclear pre-mRNA domain-containing protein 1B	RPRD1B	37	No	No	No	No			3.223E-04	3.538E-04
Regulator of chromosome condensation	RCC1	45	Yes	No	No	No		6.764E-05		
Regulator of microtubule dynamics protein 1	RMDN1	36	No	No	No	No			2.018E-04	
Renin receptor	ATP6AP2	39	No	Yes	Yes	Yes	8.852E-04	6.357E-04		
Replication protein A 70 kDa DNA-binding subunit	RPA1	68	No	No	No	No			4.679E-05	4.646E-05
Reticulon-4 receptor-like 1	RTN4RL1	49	No	No	Yes	No		1.520E-04		
Retinoblastoma-binding protein 5	RBBP5	59	No	No	No	No		3.562E-05		
Rho GDP-dissociation inhibitor 2	ARHGDIB	23	Yes	Yes	No	No				3.192E-04
Rho-associated protein kinase 1	ROCK1	158	Yes	No	No	No		3.490E-05		
Rho-related GTP-binding protein RhoB	RHOB	22	No	Yes	No	No			2.642E-04	
Ribonuclease UK114	HRSP12	14	No	Yes	Non-Cl	No		3.478E-04		
Ribonucleoside-diphosphate reductase large subunit	RRM1	90	No	No	No	No		2.994E-05	1.238E-04	
Ribonucleoside-diphosphate reductase subunit M2	RRM2	45	Yes	No	No	No		2.198E-04	1.047E-04	
Ribonucleoside-diphosphate reductase subunit M2 B	RRM2B	41	Yes	No	No	No		3.258E-04		
Ribosomal RNA small subunit methyltransferase NEP1	EMG1	27	No	No	No	No				1.417E-04
Ribosome biogenesis protein WDR12	WDR12	48	No	No	No	No			6.930E-05	5.499E-05
Ribosome maturation protein SBDS	SBDS	29	No	No	Non-Cl	No				1.633E-04
Ribosome-binding protein 1	RRBP1	152	Yes	No	No	Yes		5.022E-05	6.192E-05	
RNA polymerase II subunit A C-terminal domain phosphatase SSU72	SSU72	23	No	No	Non-Cl	No				1.186E-04
RNA-binding protein 8A	RBM8A	20	No	No	No	No				2.351E-04
Scavenger receptor class B member 1	SCARB1	61	No	Yes	No	Yes		7.709E-05	1.897E-04	
Sec1 family domain-containing protein 1	SCFD1	72	No	No	No	Yes		8.189E-05		
Secernin-2	SCRN2	47	No	Yes	No	No		8.981E-05		
Secretory carrier-associated membrane protein 3	SCAMP3	38	No	Yes	No	Yes			8.448E-05	
Sedoheptulokinase	SHPK	51	No	No	Non-Cl	No		1.307E-04	7.292E-05	
Selenide, water dikinase 1	SEPHS1	36	Yes	No	Non-Cl	No		2.431E-04	8.972E-05	
Selenium-binding protein 1	SELENBP1	52	Yes	Yes	No	No		8.160E-04	3.681E-03	
Semaphorin-3B	SEMA3B	83	Yes	No	Yes	No		1.009E-04	3.105E-04	
Semaphorin-3C	SEMA3C	85	Yes	No	Yes	No		5.559E-04	6.873E-05	
Semaphorin-3E	SEMA3E	89	No	No	Yes	No			1.432E-04	
Semaphorin-4C	SEMA4C	93	No	No	Yes	Yes		3.530E-04		
Semaphorin-4D	SEMA4D	96	Yes	No	Yes	Yes		7.242E-05	4.702E-05	
Semaphorin-6B	P15121	95	No	No	Yes	Yes		1.018E-04		
Septin-2	SEPT2	41	No	No	No	No		1.694E-04		1.926E-04
Septin-7	SEPT7	51	No	No	No	No		1.095E-04	9.316E-05	
Septin-8	SEPT8	56	No	No	No	No		6.916E-05		
Septin-9	SEPT9	65	No	No	Non-Cl	No		1.700E-04		
Sequestosome-1	SQSTM1	48	No	Yes	Non-Cl	No				2.362E-04
Serglycin	SRGN	18	No	Yes	Yes	No				2.923E-03
Serine hydroxymethyltransferase, cytosolic	SHMT1	53	No	Yes	No	No		1.466E-04	1.332E-04	
Serine hydroxymethyltransferase, mitochondrial	SHMT2	56	No	Yes	No	No				1.359E-04
Serine incorporator 5	SERINC5	47	No	Yes	No	Yes		2.464E-04		
Serine protease HTRA1	HTRA1	51	Yes	Yes	Yes	No	5.567E-03	2.768E-04		
Serine/arginine-rich splicing factor 1	SRSF1	28	No	No	No	No		3.445E-04		4.442E-04
Serine/arginine-rich splicing factor 3	SRSF3	19	No	No	No	No		2.027E-04		
Serine/arginine-rich splicing factor 7	SRSF7	27	No	No	No	No		2.593E-04		
Serine/threonine-protein kinase DCLK1	DCLK1	82	No	No	No	No		2.116E-04		
regulatory subunit B alpha isoform	PPP2R2A	52	No	No	No	No		8.503E-05	8.927E-05	
regulatory subunit delta isoform	PPP2R5D	70	No	No	No	No		4.730E-05		
Serine/threonine-protein phosphatase 4 catalytic subunit	PPP4C	35	Yes	No	No	No			1.135E-04	
Serine/threonine-protein phosphatase 6 catalytic subunit	PPP6C	35	No	No	No	No				1.514E-04
Serpin B8	SERPINB8	43	No	No	No	No				2.474E-04
Serpin B9	SERPINB9	42	No	Yes	Non-Cl	No				2.304E-04
SET and MYND domain-containing protein 5	SMYD5	47	No	No	No	No				5.565E-05
S-formylglutathione hydrolase	ESD	31	No	Yes	Non-Cl	No	2.139E-04	1.513E-04		
SH3 domain-binding glutamic acid-rich-like protein	SH3BGRL	13	Yes	Yes	Non-Cl	No		2.369E-03		7.137E-04
Shootin-1	KIAA1598	72	No	No	No	No		7.550E-05		
Sialidase-1	NEU1	45	No	Yes	No	Yes			8.356E-05	
Signal recognition particle 19 kDa protein	SRP19	16	No	No	Non-Cl	No		2.304E-04		2.005E-04
Signal transducer and activator of transcription 1-alpha/beta	STAT1	83	Yes	Yes	No	No		6.988E-05		
Signal transducer and activator of transcription 3	STAT3	88	No	No	Non-Cl	No			1.511E-04	
Small nuclear ribonucleoprotein F	SNRPF	10	No	No	No	No				3.338E-04
Small nuclear ribonucleoprotein G-like protein	SNRPGP15	9	No	No	Non-Cl	No		3.725E-04		

Small nuclear ribonucleoprotein Sm D1	SNRPD1	13	Yes	Yes	No	No		1.145E-03		
Small nuclear ribonucleoprotein Sm D2	SNRPD2	14	Yes	Yes	Non-Cl	No		6.089E-04		5.869E-04
S-methyl-5'-thioadenosine phosphorylase	MTAP	31	No	Yes	No	No			8.116E-05	
Sodium/potassium-transporting ATPase subunit beta-1	ATP1B1	35	No	Yes	No	Yes		2.997E-04	6.731E-04	
Sodium-dependent multivitamin transporter	SLC5A6	69	Yes	Yes	No	Yes			1.852E-04	
Solute carrier family 12 member 2	SLC12A2	131	No	Yes	No	Yes		1.975E-05		
Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	54	Yes	Yes	No	Yes			3.069E-04	
Sorting nexin-1	SNX1	59	Yes	No	No	Yes		9.053E-05		
Sorting nexin-2	SNX2	58	No	No	Non-Cl	No		1.192E-04		6.681E-05
Sorting nexin-3	SNX3	19	No	Yes	No	No				3.544E-04
Sorting nexin-5	SNX5	47	No	No	No	No		1.879E-04		
SPARC	SPARC	35	Yes	Yes	Yes	No	9.125E-03			4.209E-04
SPARC-related modular calcium-binding protein 1	SMOC1	48	No	No	Yes	No	1.222E-03		5.365E-05	
Spermine synthase	SMS	41	Yes	No	No	No		1.442E-04		2.828E-04
Sphingomyelin phosphodiesterase	SMPD1	70	Yes	No	No	Yes	5.282E-05		4.610E-05	
Splicing factor 1	SF1	68	No	No	No	No				4.533E-05
Splicing factor 3A subunit 3	SF3A3	59	Yes	No	No	No		1.034E-04	1.154E-04	
Splicing factor 3B subunit 1	SF3B1	146	No	No	No	No		1.602E-04		7.142E-05
Splicing factor 3B subunit 2	SF3B2	100	No	No	No	No		2.141E-05		
Splicing factor 3B subunit 5	SF3B5	10	No	No	Non-Cl	No		3.865E-04		6.022E-04
Splicing factor U2AF 35 kDa subunit	U2AF1	28	Yes	No	No	No		1.203E-04		1.434E-04
Splicing factor, proline- and glutamine-rich	SFPQ	76	No	No	No	No		1.205E-04	6.497E-05	
Src substrate cortactin	CTTN	62	Yes	Yes	No	No		5.103E-04	3.595E-04	
Stanniocalcin-2	STC2	33	Yes	Yes	Yes	No		1.041E-03		
Structural maintenance of chromosomes protein 1A	SMC1A	143	No	Yes	No	No		1.541E-05		
Structural maintenance of chromosomes protein 3	SMC3	142	No	Yes	No	No				2.352E-05
Structural maintenance of chromosomes protein 4	SMC4	147	No	No	No	No				5.834E-05
Sulfatase-modifying factor 1	SUMF1	41	No	No	Yes	No			9.319E-05	
Sulfatase-modifying factor 2	SUMF2	34	No	No	Yes	No		9.514E-05	1.310E-03	
Sulfotransferase 1A1	SULT1A1	34	No	No	Non-Cl	No			9.830E-05	
SUMO-activating enzyme subunit 2	UBA2	71	Yes	No	No	No		7.384E-05		
Superkiller viralicidic activity 2-like 2	SKIV2L2	118	No	No	No	No		2.733E-05		2.207E-05
Superoxide dismutase [Mn], mitochondrial	SOD2	25	Yes	No	No	No				2.882E-04
Suppressor of tumorigenicity 14 protein	ST14	95	Yes	No	No	Yes		1.448E-04	1.034E-03	
Sushi domain-containing protein 2	SUSD2	90	No	Yes	Yes	Yes			1.328E-03	
Synaptogyrin-2	SYNGR2	25	No	Yes	No	Yes		1.279E-04	2.581E-04	
Syndecan-1	SDC1	32	Yes	Yes	Yes	Yes			1.865E-04	
Syntaxin-3	STX3	33	No	Yes	No	Yes			7.947E-05	
Syntaxin-7	STX7	30	No	Yes	No	Yes			3.107E-04	2.882E-04
Syntaxin-binding protein 2	STXBP2	66	No	Yes	No	No		8.828E-05	6.895E-05	
Syntaxin-binding protein 3	STXBP3	68	No	Yes	No	No		6.383E-05	5.888E-05	
TBC1 domain family member 4	TBC1D4	147	No	No	No	No				1.792E-05
Tenascin	TNC	241	No	Yes	Yes	No	4.379E-04			1.050E-04
Tetraspanin-15	TSPAN15	33	No	Yes	No	Yes			1.983E-04	
TGF-beta receptor type-1	TGFBR1	56	No	No	Yes	Yes			8.128E-05	
Thioredoxin domain-containing protein 12	TXNDC12	19	No	No	Yes	No				2.026E-04
THO complex subunit 4	ALYREF	27	No	No	No	No			3.187E-04	2.678E-04
Thymidine phosphorylase	TYMP	50	No	No	Non-Cl	No			3.129E-04	
Thymosin beta-10	TMSB10	5	Yes	No	Non-Cl	No			1.073E-03	7.860E-04
Tissue factor pathway inhibitor	TFPI	35	Yes	Yes	Yes	No		1.412E-04		2.452E-04
Tissue-type plasminogen activator	PLAT	63	Yes	Yes	Yes	No	1.806E-04			7.286E-04
TOM1-like protein 2	TOM1L2	56	No	Yes	No	No		4.720E-05		
Transcription elongation factor A protein 1	TCEA1	34	No	No	Non-Cl	No		1.424E-04		6.343E-04
Transcription elongation factor SPT6	SUPT6H	199	Yes	No	No	No			3.642E-05	
Transcription intermediary factor 1-beta	TRIM28	89	No	No	No	No		9.667E-05	9.654E-05	
Transcriptional activator protein Pur-beta	PURB	33	No	No	No	No			1.670E-04	
Transforming acidic coiled-coil-containing protein 2	TACC2	309	Yes	No	No	No			1.364E-05	
Transforming growth factor beta-1	TGFB1	44	Yes	Yes	Yes	No		8.622E-05		1.173E-04
Transforming growth factor beta-2	TGFB2	48	No	Yes	Yes	No				2.101E-03
Transgelin	TAGLN	23	No	Yes	Non-Cl	No	1.126E-03			6.638E-04
Trans-L-3-hydroxyproline dehydratase	L3HYPDH	38	No	No	Non-Cl	No				9.844E-05
Translation initiation factor eIF-2B subunit alpha	EIF2B1	34	No	No	No	No		1.259E-04		3.215E-04
Translational activator GCN1	GCN1L1	293	No	Yes	No	No		1.435E-05		
Translin-associated protein X	TSNAX	33	No	Yes	No	No		9.819E-05		
Transmembrane glycoprotein NMB	GPNMB	64	No	No	Yes	Yes			7.062E-05	
Transportin-3	TNPO3	104	No	Yes	Non-Cl	No				2.492E-05
Trefoil factor 1	TFF1	9	Yes	No	Yes	No		6.742E-03		
Trefoil factor 3	TFF3	9	Yes	No	Yes	No		2.787E-03		
Tripeptidyl-peptidase 2	TPP2	138	No	No	No	No			1.385E-04	1.381E-04
tRNA (cytosine(34)-C(5))-methyltransferase	NSUN2	86	No	No	No	No				5.253E-05
tRNA (guanine-N(7)-)-methyltransferase	METTL1	31	No	No	Non-Cl	No		8.671E-05	1.694E-04	
tRNA methyltransferase 112 homolog	TRMT112	14	No	No	Non-Cl	No				4.178E-04
tRNA-splicing ligase Rtc8 homolog	RTCB	55	No	No	Non-Cl	No		5.639E-05		
Trophoblast glycoprotein	TPBG	46	Yes	Yes	Yes	Yes		1.493E-04		
Tropomodulin-3	TMOD3	40	No	No	Non-Cl	No		1.472E-04		
Tubulin-specific chaperone D	TBCD	133	No	No	No	No				5.361E-05
Tubulin--tyrosine ligase-like protein 12	TTLL12	74	No	No	Non-Cl	No			1.171E-04	

Tumor necrosis factor receptor superfamily member 1A	TNFRSF1A	50	Yes	Yes	Yes	Yes	2.660E-04	8.353E-05		
Tumor susceptibility gene 101 protein	TSG101	44	Yes	Yes	Non-Cl	No		1.577E-04		
Tyrosine-protein kinase receptor UFO	AXL	98	No	Yes	Yes	Yes	1.335E-04			3.863E-04
Tyrosine-protein phosphatase non-receptor type 1	PTPN1	50	Yes	No	No	Yes		8.812E-05		
Tyrosine--tRNA ligase, cytoplasmic	YARS	59	No	No	No	No		1.431E-04		
U1 small nuclear ribonucleoprotein 70 kDa	SNRNP70	52	No	No	No	No		9.774E-05		
U1 small nuclear ribonucleoprotein A	SNRPA	31	No	No	Non-Cl	No		8.429E-05		
U2 small nuclear ribonucleoprotein A'	SNRPA1	28	Yes	No	Non-Cl	No		1.883E-04		
U4/U6.U5 tri-snRNP-associated protein 2	USP39	65	No	No	Non-Cl	No		4.197E-05		
Ubiquitin carboxyl-terminal hydrolase 15	USP15	112	No	No	No	No		1.937E-05		
Ubiquitin conjugation factor E4 A	UBE4A	123	No	No	Non-Cl	No		4.028E-05		
Ubiquitin domain-containing protein UBFD1	UBFD1	33	No	No	Non-Cl	No		1.241E-04		
Ubiquitin thioesterase OTUB1	OTUB1	31	Yes	Yes	No	No		9.316E-04	5.735E-04	
Ubiquitin-conjugating enzyme E2 H	UBE2H	21	No	No	No	No			3.853E-04	
Ubiquitin-fold modifier-conjugating enzyme 1	UFC1	19	No	Yes	No	No			1.745E-04	1.735E-04
Ubiquitin-like modifier-activating enzyme 6	UBA6	118	No	No	No	No		1.822E-05	6.673E-05	
Ubiquitin-like protein ISG15	ISG15	18	No	No	No	No		2.589E-04		
Ubiquitin-protein ligase E3A	UBE3A	101	Yes	No	No	No			3.294E-05	2.628E-05
UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2	B3GNT2	46	No	No	No	Yes	1.736E-04	2.539E-04		
UDP-glucose 6-dehydrogenase	UGDH	55	Yes	Yes	No	No		1.544E-04		8.156E-05
UDP-N-acetylhexosamine pyrophosphorylase	UAP1	59	Yes	No	Non-Cl	No	2.875E-04			1.979E-04
Unconventional myosin-Ib	MYO1B	132	No	Yes	No	No		1.380E-04		
Unconventional myosin-Id	MYO1D	116	No	Yes	No	No			5.814E-05	
Unconventional myosin-VI	MYO6	150	Yes	Yes	No	No		5.900E-05		
Unconventional myosin-XVIIIa	MYO18A	233	No	No	No	No			1.721E-05	
UPF0553 protein C9orf64	C9orf64	39	No	No	No	No		3.214E-04		3.550E-04
UPF0556 protein C19orf10	C19orf10	19	No	No	Yes	No		2.732E-04	7.078E-04	
UPF0687 protein C20orf27	C20orf27	19	No	No	No	No		3.560E-04	1.685E-04	
UPF0696 protein C11orf68	C11orf68	27	No	No	Non-Cl	No				1.378E-04
Urokinase plasminogen activator surface receptor	PLAUR	37	Yes	No	Yes	No				2.070E-04
Urokinase-type plasminogen activator	PLAU	49	Yes	Yes	Yes	No	8.652E-03			1.193E-03
Uroporphyrinogen decarboxylase	UROD	41	No	No	Non-Cl	No				2.818E-04
Utrophin	UTRN	394	No	Yes	No	No		2.485E-05		
Vacuolar protein sorting-associated protein 28 homolog	VPS28	25	No	Yes	Non-Cl	No	2.200E-04	2.354E-04		
Vacuolar protein sorting-associated protein 45	VPS45	65	No	No	No	No		1.415E-04		
Vacuolar protein sorting-associated protein VTA1 homolog	VTA1	34	No	Yes	No	No		1.545E-04	1.517E-04	
Vacuolar protein-sorting-associated protein 25	VPS25	21	No	Yes	No	No		2.458E-04		2.628E-04
Vacuolar-sorting protein SNF8	SNF8	29	No	Yes	No	No		1.119E-04	8.902E-05	
Vascular endothelial growth factor receptor 1	FLT1	151	Yes	Yes	Yes	Yes		1.420E-05		
Vesicle-associated membrane protein 3	VAMP3	11	No	Yes	No	Yes				2.300E-04
Vesicle-associated membrane protein 8	VAMP8	11	No	Yes	No	Yes			2.931E-04	
Vesicle-associated membrane protein-associated protein B/C	VAPB	27	No	No	No	Yes		4.142E-04	5.007E-04	
Vesicle-trafficking protein SEC22b	SEC22B	25	No	No	No	Yes		1.757E-04		1.359E-04
Vigilin	HDLBP	141	No	No	No	No		3.722E-05		
Vitamin K-dependent protein S	PROS1	75	Yes	Yes	Yes	No	9.151E-05			7.687E-05
V-set and transmembrane domain-containing protein 2-like protein	VSTM2L	22	No	No	Yes	No			3.984E-04	
V-type proton ATPase catalytic subunit A	ATP6V1A	68	No	Yes	No	No		4.642E-05	8.473E-05	
V-type proton ATPase subunit B, brain isoform	ATP6V1B2	57	No	Yes	Non-Cl	No			4.591E-05	
V-type proton ATPase subunit G 1	ATP6V1G1	14	No	Yes	No	No			5.506E-04	5.408E-04
WD repeat-containing protein 5	WDR5	37	Yes	No	Non-Cl	No		1.284E-04		8.646E-05
Xaa-Pro dipeptidase	PEPD	55	No	Yes	No	No			1.890E-04	2.206E-04
Xylosyltransferase 1	XYLT1	108	No	No	Yes	No		1.139E-04	3.161E-04	
Xylulose kinase	XYLB	58	No	No	Non-Cl	No			1.088E-04	
YTH domain family protein 1	YTHDF1	61	No	No	No	No				7.177E-05
Zinc finger protein 622	ZNF622	54	No	No	No	No				6.073E-05
Zinc transporter ZIP6	SLC39A6	85	Yes	No	Yes	Yes		1.002E-04		
Zinc-alpha-2-glycoprotein	AZGP1	34	Yes	Yes	Yes	No		1.598E-04	4.716E-04	
Zymogen granule membrane protein 16	ZG16	18	No	No	Yes	No			1.152E-03	

APPENDIX 2: List of identified membrane proteins

MW = molecular weight; BC = breast cancer; NSAF = normalized spectral abundance factor; Nil expression indicated by grey-shaded box

Identified membrane proteins present in all four breast epithelial cell lines							
		BC	MW	Average NSAF			
Identified Proteins	Gene	database	(kDa)	HMEC	MCF7	SKBR3	MDA231
26S protease regulatory subunit 7	PSMC2	No	49	2.798E-04	2.654E-04	1.895E-04	1.951E-04
26S protease regulatory subunit 8	PSMC5	No	46	3.219E-04	1.088E-04	2.600E-04	8.442E-05
26S proteasome non-ATPase regulatory subunit 2	PSMD2	Yes	100	2.156E-04	1.406E-04	1.202E-04	2.979E-04
26S proteasome non-ATPase regulatory subunit 3	PSMD3	No	61	5.757E-04	2.750E-04	1.533E-04	2.172E-04
28S ribosomal protein S29, mitochondrial	DAP3	No	46	1.813E-04	2.849E-04	2.404E-04	1.972E-04
40S ribosomal protein S11	RPS11	No	18	1.961E-03	7.259E-04	5.398E-04	6.164E-04
40S ribosomal protein S13	RPS13	No	17	1.993E-03	1.733E-03	1.702E-03	1.162E-03
40S ribosomal protein S14	RPS14	Yes	16	3.140E-03	1.013E-03	3.468E-03	8.986E-04
40S ribosomal protein S15a	RPS15A	No	15	9.889E-04	2.973E-03	1.786E-03	2.218E-03
40S ribosomal protein S16	RPS16	Yes	16	1.449E-03	1.821E-03	1.406E-03	2.151E-03
40S ribosomal protein S18	RPS18	Yes	18	2.542E-03	2.008E-03	3.471E-03	3.949E-03
40S ribosomal protein S2	RPS2	No	31	8.209E-04	2.622E-03	9.525E-04	2.460E-03
40S ribosomal protein S20	RPS20	No	13	9.408E-04	7.118E-04	1.209E-03	5.028E-04
40S ribosomal protein S24	RPS24	No	15	2.058E-03	5.161E-03	1.876E-03	6.059E-03
40S ribosomal protein S3	RPS3	No	27	2.850E-03	1.202E-03	3.771E-03	1.035E-03
40S ribosomal protein S3a	RPS3A	No	30	3.868E-04	9.381E-04	2.212E-03	8.038E-04
40S ribosomal protein S5	RPS5	No	23	5.797E-03	2.435E-03	3.565E-03	2.751E-03
40S ribosomal protein S6	RPS6	No	29	1.975E-03	1.734E-03	1.573E-03	3.359E-03
40S ribosomal protein S8	RPS8	No	24	3.692E-03	5.714E-03	4.719E-03	7.431E-03
40S ribosomal protein SA	RPSA	No	33	1.554E-03	3.658E-04	7.366E-04	5.645E-04
5'-nucleotidase	NT5E	No	63	1.943E-03	4.257E-05	2.723E-04	3.291E-03
60S ribosomal protein L10	RPL10	No	25	4.475E-04	1.469E-03	9.183E-04	1.664E-03
60S ribosomal protein L12	RPL12	No	18	3.117E-03	8.236E-04	4.757E-03	8.239E-04
60S ribosomal protein L13	RPL13	No	24	7.300E-04	1.005E-03	1.750E-03	2.262E-03
60S ribosomal protein L13a	RPL13A	No	24	1.881E-03	3.997E-03	4.825E-04	3.912E-03
60S ribosomal protein L14	RPL14	No	23	8.833E-04	5.938E-03	4.487E-04	5.768E-03
60S ribosomal protein L18	RPL18	No	22	3.116E-03	9.122E-03	3.002E-03	1.168E-02
60S ribosomal protein L18a	RPL18A	No	21	1.563E-03	7.279E-04	1.174E-03	2.502E-03
60S ribosomal protein L23a	RPL23A	Yes	18	1.231E-03	1.064E-03	1.462E-03	1.213E-03
60S ribosomal protein L28	RPL28	No	16	6.685E-04	1.600E-03	1.952E-03	2.504E-03
60S ribosomal protein L3	RPL3	No	46	6.196E-04	9.345E-04	1.319E-03	1.666E-03
60S ribosomal protein L32	RPL32	No	16	6.675E-04	5.338E-04	1.746E-03	1.012E-03
60S ribosomal protein L4	RPL4	No	48	1.211E-03	8.137E-04	4.549E-04	1.175E-03
60S ribosomal protein L6	RPL6	No	33	1.410E-03	3.886E-03	2.181E-03	5.187E-03
60S ribosomal protein L7	RPL7	No	29	8.491E-04	1.117E-03	5.793E-04	2.311E-03
60S ribosomal protein L7a	RPL7A	No	30	1.034E-03	1.214E-03	4.517E-04	1.518E-03
6-phosphofructokinase type C	PFKP	No	86	3.639E-04	2.043E-04	1.667E-04	3.520E-04
78 kDa glucose-regulated protein	HSPA5	No	72	6.058E-04	9.191E-04	4.207E-04	1.354E-03
Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	No	69	1.088E-04	2.462E-04	2.499E-04	8.113E-05
Adenosine 3'-phospho 5'-phosphosulfate transporter 1	SLC35B2	No	48	9.262E-04	1.035E-03	7.218E-04	3.474E-04
Adipocyte plasma membrane-associated protein	APMAP	No	46	6.255E-04	5.815E-03	1.129E-03	7.873E-04
ADP-dependent glucokinase	ADPGK	No	54	2.551E-04	3.102E-04	1.925E-04	5.920E-05
All-trans-retinol 13,14-reductase	RETSAT	No	67	3.065E-04	2.450E-04	5.317E-04	2.987E-04
Alpha-enolase	ENO1	Yes	47	5.466E-04	8.962E-04	2.201E-04	7.719E-04
Alpha-mannosidase 2	MAN2A1	No	131	1.151E-04	2.706E-04	2.714E-04	1.662E-04
Aminoacyl tRNA synthase complex-interacting multifunctional protein 1	AIMP1	No	34	3.126E-04	1.551E-04	4.262E-04	9.378E-05
Amyloid beta A4 protein	APP	No	87	5.903E-05	2.152E-05	2.645E-04	1.280E-04
Annexin A2	ANXA2	No	39	2.722E-03	1.294E-03	9.609E-04	1.170E-02
Antigen peptide transporter 1	TAP1	Yes	87	2.723E-04	3.093E-04	6.251E-04	4.650E-04
Apolipoprotein O	APOO	No	22	5.498E-04	2.445E-04	4.259E-04	4.230E-04
Apoptosis regulator BAX	BAX	Yes	21	4.054E-04	5.135E-04	5.566E-04	5.072E-04
Apoptosis-inducing factor 1, mitochondrial	AIFM1	No	67	1.809E-04	6.313E-04	4.151E-04	2.519E-04
Arginine--tRNA ligase, cytoplasmic	RARS	No	75	3.526E-04	2.213E-04	5.636E-04	2.071E-04
Aspartate--tRNA ligase, cytoplasmic	DARS	No	57	2.466E-04	2.531E-04	9.679E-05	1.814E-04
Atlastin-2	ATL2	No	66	3.698E-04	2.870E-04	3.939E-04	9.403E-05
Atlastin-3	ATL3	No	61	1.588E-04	3.065E-03	1.184E-03	5.725E-04
ATP synthase F(0) complex subunit B1, mitochondrial	ATP5F1	No	29	9.258E-04	5.594E-03	1.101E-03	6.286E-04
ATP synthase subunit alpha, mitochondrial	ATP5A1	No	60	8.974E-04	4.446E-03	1.334E-03	3.678E-04
ATP synthase subunit beta, mitochondrial	ATP5B	No	57	1.020E-03	6.359E-03	1.490E-03	2.876E-04
ATP synthase subunit g, mitochondrial	ATP5L	No	11	3.785E-03	2.644E-03	3.482E-03	1.896E-03
ATP-binding cassette sub-family D member 3	ABCD3	Yes	75	3.018E-04	3.363E-04	9.759E-04	7.779E-04
ATP-dependent RNA helicase A	DHX9	Yes	141	2.127E-04	2.056E-04	8.032E-04	1.996E-04
ATP-dependent RNA helicase DDX3X	DDX3X	No	73	9.256E-04	2.961E-04	4.107E-04	2.888E-04
Basal cell adhesion molecule	BCAM	No	67	1.000E-03	4.444E-04	1.525E-03	2.466E-04
B-cell receptor-associated protein 31	BCAP31	No	28	6.058E-04	9.614E-03	7.382E-03	6.232E-03
Bifunctional glutamate/proline--tRNA ligase	EPRS	Yes	171	3.637E-04	2.904E-04	6.994E-04	2.943E-04

Brain acid soluble protein 1	BASP1	No	23	5.573E-04	8.006E-04	2.141E-03	2.976E-04
CAD protein	CAD	Yes	243	3.478E-04	2.680E-04	5.423E-05	3.290E-04
Calcium-binding mitochondrial carrier protein Aralar2	SLC25A13	No	74	3.270E-04	4.956E-04	5.114E-04	3.737E-04
Calcium-binding mitochondrial carrier protein SCaMC-1	SLC25A24	No	53	3.462E-04	7.425E-04	2.832E-04	7.152E-04
Calcium-transporting ATPase type 2C member 1	ATP2C1	Yes	101	2.303E-04	2.105E-04	6.931E-04	2.386E-04
Calmodulin	CALM1	No	17	5.839E-04	2.581E-04	1.810E-03	3.652E-04
Calnexin	CANX	Yes	68	4.180E-04	1.172E-03	1.278E-03	7.342E-04
cAMP-dependent protein kinase type II-alpha regulatory subunit	PRKAR2A	No	46	2.755E-04	1.118E-04	1.819E-04	7.242E-05
Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	No	88	2.023E-04	1.472E-03	2.105E-03	3.037E-04
Carnitine O-palmitoyltransferase 2, mitochondrial	CPT2	No	74	4.389E-04	1.352E-04	4.680E-04	3.010E-05
Catenin delta-1	CTNND1	No	108	5.089E-04	2.433E-04	1.114E-04	1.791E-04
Cathepsin D	CTSD	Yes	45	1.162E-04	6.082E-04	3.251E-04	1.574E-04
CD59 glycoprotein	CD59	No	14	4.921E-03	8.406E-04	7.337E-03	1.176E-03
CD63 antigen	CD63	Yes	26	1.029E-03	9.538E-04	7.768E-04	4.769E-04
CD9 antigen	CD9	No	25	1.585E-03	5.466E-04	6.323E-04	4.232E-04
CDGSH iron-sulfur domain-containing protein 2	CISD2	No	15	1.330E-03	1.986E-03	2.982E-03	2.184E-03
CDK5 regulatory subunit-associated protein 3	CDK5RAP3	No	57	2.666E-04	2.782E-04	4.688E-04	1.600E-04
CDP-diacylglycerol--inositol 3-phosphatidyltransferase	CDIPT	No	24	2.977E-04	3.153E-03	9.578E-04	8.671E-04
Charged multivesicular body protein 6	CHMP6	No	23	5.939E-04	1.975E-04	1.303E-03	3.602E-04
Choline transporter-like protein 1	SLC44A1	No	73	1.522E-04	1.972E-04	1.681E-04	4.340E-05
Cleft lip and palate transmembrane protein 1	CLPTM1	No	76	1.526E-04	4.784E-04	6.848E-04	2.344E-04
Cluster of 14-3-3 protein theta	YWHAQ	Yes	28	6.254E-04	5.886E-04	1.496E-03	1.320E-03
Cluster of 40S ribosomal protein S27	RPS27	No	9	2.290E-03	1.168E-03	5.786E-03	2.211E-03
Cluster of 60S ribosomal protein L11	RPL11	No	20	1.830E-03	4.525E-04	1.918E-03	5.495E-04
Cluster of Actin, cytoplasmic 1	ACTB	Yes	42	2.559E-02	7.425E-03	1.832E-02	7.812E-03
Cluster of ADP/ATP translocase 2	SLC25A5	No	33	1.400E-03	2.399E-02	8.749E-04	5.838E-03
Cluster of Alpha-actinin-4	ACTN4	No	105	4.794E-04	4.486E-04	1.175E-03	7.431E-04
Cluster of AP-2 complex subunit beta	AP2B1	Yes	105	7.630E-05	1.566E-04	3.627E-04	1.990E-04
Cluster of ATP synthase subunit f, mitochondrial	ATP5J2	No	11	1.296E-03	2.690E-03	1.230E-03	8.332E-04
Cluster of ATPase family AAA domain-containing protein 3B	ATAD3B	No	73	2.560E-04	3.306E-04	2.602E-04	4.172E-04
Cluster of Cell division control protein 42 homolog	CDC42	No	21	1.947E-03	7.962E-03	1.769E-03	1.917E-03
Cluster of Clathrin heavy chain 1	CLTC	Yes	192	1.367E-04	8.466E-04	1.234E-03	8.758E-04
Cluster of EH domain-containing protein 1	EHD1	No	61	1.741E-04	4.576E-05	1.558E-04	2.244E-04
Cluster of Eukaryotic initiation factor 4A-I	EIF4A1	Yes	46	4.110E-04	5.145E-04	2.105E-04	6.607E-04
Cluster of Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	Yes	36	2.196E-03	2.154E-03	1.754E-03	1.634E-03
Cluster of GTPase NRas	NRAS	Yes	21	2.542E-03	1.752E-03	1.779E-03	2.040E-03
Cluster of Guanine nucleotide-binding protein G(i) subunit alpha-2	GNAI2	No	40	1.887E-03	4.356E-04	1.109E-03	3.111E-03
Cluster of Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	GNB1	No	37	4.975E-03	1.270E-03	2.118E-03	2.932E-03
Cluster of Guanine nucleotide-binding protein subunit alpha-13	GNA13	No	44	3.492E-04	3.398E-04	8.306E-04	4.140E-04
Cluster of Heat shock cognate 71 kDa protein	HSPA8	No	71	1.012E-03	1.610E-03	1.599E-03	2.357E-03
Cluster of Heat shock protein HSP 90-alpha	HSP90AA1	No	85	8.487E-04	1.411E-03	6.329E-04	1.400E-03
Cluster of Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	No	39	3.967E-04	1.455E-03	4.056E-03	1.938E-04
Cluster of Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	No	49	1.631E-04	3.528E-04	6.623E-04	9.609E-05
Cluster of Heterogeneous nuclear ribonucleoprotein U	HNRNPU	No	91	5.461E-04	1.587E-03	1.719E-03	1.305E-03
Cluster of Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	Yes	34	6.652E-04	3.589E-03	1.978E-03	4.041E-03
Cluster of Histone H2B type F-S	H2BFS	No	14	2.054E-03	9.999E-03	7.207E-03	1.437E-02
Cluster of Importin subunit alpha-5	KPNA1	No	60	2.219E-04	6.104E-05	1.106E-04	8.265E-05
Cluster of Isoform 2 of Extended synaptotagmin-2	ESYT2	No	99	2.007E-04	4.825E-04	6.745E-04	6.158E-04
Cluster of Isoform 2 of Reticulon-4	RTN4	Yes	40	1.717E-04	2.508E-04	1.521E-04	7.216E-04
Cluster of Isoform 2 of Unconventional myosin-Ic	MYO1C	No	118	1.759E-03	3.485E-04	1.980E-03	1.963E-03
Cluster of Isoform Gnas-2 of Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	GNAS	Yes	44	2.565E-04	7.639E-04	7.063E-04	5.304E-04
Cluster of Keratin, type I cytoskeletal 14	KRT14	Yes	52	6.353E-02	7.618E-03	8.517E-03	9.455E-03
Cluster of Lamina-associated polypeptide 2, isoform alpha	TMPO	No	75	1.841E-04	4.933E-04	3.741E-04	9.276E-04
Cluster of Myosin regulatory light chain 12B	MYL12B	No	20	6.668E-03	2.828E-04	4.164E-03	4.630E-04
Cluster of Myosin-9	MYH9	Yes	227	8.236E-03	6.413E-04	9.851E-03	1.498E-03
Cluster of Nuclear mitotic apparatus protein 1	NUMA1	Yes	238	5.160E-05	5.433E-04	4.611E-04	5.357E-04
Cluster of Plectin	PLEC	No	532	2.443E-03	4.584E-04	8.056E-04	1.184E-03
Cluster of Polyadenylate-binding protein 1	PABPC1	No	71	5.713E-04	6.466E-04	2.103E-04	3.708E-04
Cluster of Protein lin-7 homolog C	LIN7C	No	22	3.218E-04	2.418E-04	4.280E-04	1.222E-04
Cluster of Protein transport protein Sec61 subunit alpha isoform 1	SEC61A1	No	52	8.723E-04	1.295E-03	1.236E-03	9.327E-04
Cluster of Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	No	91	7.042E-05	1.620E-04	4.535E-04	7.392E-05
Cluster of Ras-related C3 botulinum toxin substrate 1	RAC1	No	21	2.244E-03	1.036E-03	1.066E-03	3.086E-03
Cluster of Ras-related protein Rab-10	RAB10	No	23	1.132E-02	1.986E-02	2.739E-02	1.625E-02
Cluster of Ras-related protein Rab-2A	RAB2A	No	24	2.779E-03	1.126E-02	8.137E-03	9.637E-03
Cluster of Ras-related protein Rab-6A	RAB6A	No	24	3.831E-03	4.135E-03	3.876E-03	3.569E-03
Cluster of Ras-related protein Rap-1A	RAP1A	No	21	2.541E-03	8.024E-03	3.798E-03	3.685E-03
Cluster of RNA-binding motif protein, X chromosome	RBMX	No	42	4.465E-04	1.357E-03	2.778E-03	8.181E-04
Cluster of Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	PPP2R1A	No	65	1.878E-04	1.046E-04	8.204E-05	5.973E-05

Cluster of Serine/threonine-protein phosphatase PP1-beta catalytic subunit	PPP1CB	No	37	3.799E-04	1.994E-04	1.478E-04	3.385E-04
Cluster of Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	No	113	3.075E-03	1.253E-03	6.174E-03	1.076E-03
Cluster of Transforming protein RhoA	RHOA	Yes	22	3.704E-04	1.355E-03	2.368E-03	3.172E-03
Cluster of Transportin-1	TNPO1	No	102	3.330E-04	5.914E-05	1.855E-04	1.313E-04
Cluster of Tubulin alpha-4A chain	TUBA4A	No	50	9.302E-03	1.222E-02	2.099E-03	7.950E-03
Cluster of Tubulin beta chain	TUBB	No	50	1.475E-02	7.284E-03	2.269E-03	7.615E-03
Cluster of Tyrosine-protein kinase Fyn	FYN	No	61	7.335E-04	1.362E-04	3.811E-04	1.255E-04
Cluster of Unconventional myosin-Ib	MYO1B	No	132	2.545E-03	3.157E-04	2.514E-04	4.068E-04
Cluster of Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	No	70	3.565E-04	3.468E-04	7.341E-04	1.852E-04
Coatomer subunit alpha	COPA	No	138	2.642E-04	4.020E-04	1.759E-04	2.873E-04
Coatomer subunit beta	COPB1	No	107	2.786E-04	3.906E-04	2.420E-04	1.854E-04
Coatomer subunit beta'	COPB2	Yes	102	9.872E-05	2.471E-04	4.902E-04	1.041E-04
Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial	CHCHD3	No	26	3.321E-04	1.109E-03	6.886E-04	7.277E-04
Cullin-associated NEDD8-dissociated protein 1	CAND1	No	136	1.759E-04	2.852E-04	6.962E-05	1.456E-04
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	No	53	1.945E-04	8.939E-04	6.209E-04	1.416E-04
Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	No	48	5.309E-04	1.406E-03	1.457E-03	5.695E-04
Cytochrome c oxidase protein 20 homolog	COX20	No	13	8.116E-04	8.224E-04	1.916E-03	7.926E-04
Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	COX4I1	Yes	20	5.858E-04	2.582E-03	1.372E-03	2.107E-03
Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	No	532	2.867E-04	3.236E-04	2.336E-04	3.833E-04
Cytoplasmic dynein 1 light intermediate chain 1	DYNC1LI1	No	57	1.509E-04	1.001E-04	1.402E-04	1.317E-04
Cytoplasmic FMR1-interacting protein 1	CYFIP1	Yes	145	1.473E-04	2.171E-04	1.257E-04	1.764E-04
Cytoskeleton-associated protein 4	CKAP4	No	66	1.311E-04	4.675E-04	2.182E-04	1.149E-03
DDRGK domain-containing protein 1	DDRGK1	No	36	4.107E-04	5.324E-04	6.873E-04	7.098E-04
Dedicator of cytokinesis protein 7	DOCK7	No	243	5.925E-05	9.581E-06	2.266E-05	2.969E-05
Dehydrogenase/reductase SDR family member 7B	DHRS7B	No	35	2.698E-04	4.753E-04	6.342E-04	1.068E-03
Desmoplakin	DSP	No	332	8.394E-04	1.415E-04	2.368E-04	3.802E-05
Disintegrin and metalloproteinase domain-containing protein 10	ADAM10	No	84	6.637E-05	2.669E-05	4.309E-04	9.856E-05
DNA-dependent protein kinase catalytic subunit	PRKDC	No	469	1.202E-03	3.619E-04	1.100E-03	6.923E-04
DnaJ homolog subfamily A member 1	DNAJA1	No	45	3.262E-04	1.922E-04	1.499E-04	1.768E-04
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	DDOST	No	51	8.325E-04	6.456E-03	2.513E-03	1.778E-03
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	Yes	69	2.230E-03	5.574E-03	4.147E-03	3.423E-03
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	Yes	69	9.365E-04	3.154E-03	1.130E-03	9.334E-04
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	DAD1	No	12	5.745E-03	3.442E-03	2.276E-03	2.936E-03
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	STT3A	No	81	6.605E-04	8.648E-04	6.079E-04	6.084E-04
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	STT3B	No	94	1.470E-04	4.817E-04	5.483E-04	3.548E-04
Dolichyl-phosphate beta-glucosyltransferase	ALG5	No	37	1.653E-04	5.626E-04	2.602E-04	4.386E-04
E3 ubiquitin-protein ligase UBR4	UBR4	No	574	4.669E-05	5.356E-05	5.009E-05	8.890E-05
E3 UFM1-protein ligase 1	UFL1	No	90	1.308E-04	3.902E-04	1.532E-04	4.112E-04
EH domain-containing protein 4	EHD4	No	61	2.017E-04	3.063E-05	1.538E-04	1.562E-04
Elongation factor 1-delta	EEF1D	Yes	31	1.906E-04	1.020E-04	2.920E-04	1.898E-04
Elongation factor 1-gamma	EEF1G	No	50	1.320E-03	6.169E-04	1.482E-03	5.074E-04
Elongation factor Tu, mitochondrial	TUFM	No	50	1.528E-04	1.141E-04	8.778E-04	7.491E-05
Elongation of very long chain fatty acids protein 1	ELOVL1	No	33	9.187E-04	1.538E-03	2.637E-03	2.093E-03
Emerin	EMD	No	29	4.923E-04	9.644E-04	1.079E-03	1.401E-03
Endonuclease domain-containing 1 protein	ENDOD1	No	55	1.088E-03	3.971E-04	3.305E-04	1.779E-04
Endoplasmic reticulum metallopeptidase 1	ERMP1	No	100	4.571E-04	5.725E-04	1.242E-03	4.282E-05
Endoplasmic reticulum-Golgi intermediate compartment protein 1	ERGIC1	No	33	4.093E-04	1.539E-03	1.880E-03	4.600E-04
Endoplasmin	HSP90B1	No	92	2.699E-04	3.254E-04	1.050E-04	3.121E-04
Epoxide hydrolase 1	EPHX1	No	53	1.533E-04	2.465E-04	1.127E-03	1.205E-04
ER membrane protein complex subunit 1	EMC1	No	112	1.295E-04	5.494E-04	8.509E-05	3.694E-04
ER membrane protein complex subunit 4	EMC4	No	20	9.811E-04	4.599E-04	1.099E-03	7.567E-04
Erlin-2	ERLIN2	No	38	1.455E-03	8.433E-04	5.712E-04	5.076E-04
Erythrocyte band 7 integral membrane protein	STOM	No	32	1.009E-03	4.622E-04	1.722E-03	3.202E-04
Estradiol 17-beta-dehydrogenase 12	HSD17B12	Yes	34	4.289E-04	9.943E-04	6.242E-04	3.019E-03
Eukaryotic initiation factor 4A-III	EIF4A3	No	47	1.884E-04	5.405E-04	1.521E-03	2.618E-04
Exportin-1	XPO1	No	123	8.187E-05	7.240E-05	4.528E-05	8.733E-05
Exportin-2	CSE1L	Yes	110	1.731E-04	2.282E-04	1.971E-04	1.887E-04
Extended synaptotagmin-1	ESYT1	No	123	4.709E-04	6.720E-04	2.942E-04	6.149E-04
F-actin-capping protein subunit alpha-1	CAPZA1	No	33	6.056E-04	1.830E-04	1.290E-03	1.555E-04
Far upstream element-binding protein 3	FUBP3	No	62	2.389E-04	2.249E-04	3.541E-04	5.891E-05
Fatty acid synthase	FASN	Yes	273	1.851E-04	7.040E-04	1.592E-03	1.274E-03
Fatty aldehyde dehydrogenase	ALDH3A2	Yes	55	4.939E-04	2.994E-04	2.992E-04	2.459E-04

Filamin-A	FLNA	Yes	281	1.184E-04	3.485E-04	2.911E-04	4.433E-04
Flotillin-1	FLOT1	No	47	2.515E-04	7.310E-04	6.915E-04	5.967E-04
Flotillin-2	FLOT2	Yes	47	3.076E-04	9.037E-04	5.853E-04	7.093E-04
Fodrin alpha chain	SPTAN1	No	285	2.892E-05	9.610E-05	5.949E-04	9.802E-05
Fructose-bisphosphate aldolase A	ALDOA	No	39	3.636E-04	1.762E-04	3.202E-04	1.210E-04
Glutamine--tRNA ligase	QARS	No	88	1.455E-04	1.218E-04	5.444E-04	1.133E-04
Glutathione S-transferase kappa 1	GSTK1	No	25	3.802E-04	1.770E-04	1.323E-03	1.295E-04
Glycerol-3-phosphate dehydrogenase, mitochondrial	GPD2	No	81	4.656E-04	1.389E-03	3.451E-04	5.031E-04
Glycosylphosphatidylinositol anchor attachment 1 protein	GPAA1	No	68	1.123E-04	1.597E-04	1.355E-04	9.423E-05
Golgi apparatus protein 1	GLG1	No	135	3.562E-04	3.916E-04	3.907E-04	3.771E-05
Golgin subfamily A member 7	GOLGA7	No	16	8.272E-04	2.076E-04	6.072E-04	5.255E-04
GPI transamidase component PIG-T	PIGT	Yes	66	1.262E-04	1.311E-04	2.289E-04	1.884E-04
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	GNG12	No	8	6.735E-03	1.125E-03	7.274E-03	5.420E-03
Guanine nucleotide-binding protein G(q) subunit alpha	GNAQ	No	42	4.366E-04	9.962E-05	5.435E-04	1.532E-04
Guanine nucleotide-binding protein subunit alpha-11	GNA11	Yes	42	6.542E-04	7.837E-05	5.077E-04	3.436E-04
HEAT repeat-containing protein 1	HEATR1	No	242	1.497E-04	9.325E-05	6.655E-05	4.129E-05
Heat shock 70 kDa protein 1A/1B	HSPA1A	Yes	70	6.278E-04	1.058E-03	9.696E-04	9.538E-04
Heat shock protein beta-1	HSPB1	Yes	23	3.338E-03	2.759E-03	1.883E-03	1.025E-03
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	No	78	7.818E-04	2.043E-03	1.258E-03	1.398E-03
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	No	71	2.956E-04	3.151E-04	5.464E-04	4.739E-05
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	No	37	4.788E-04	1.778E-03	3.563E-03	6.822E-05
Histone H4	HIST1H4A	No	11	3.510E-02	6.310E-02	9.969E-03	9.718E-02
HLA class I histocompatibility antigen, B-41 alpha chain	HLA-B	No	41	4.307E-03	7.201E-04	3.019E-04	1.826E-02
Hornerin	HRNR	No	282	7.061E-05	7.537E-05	7.760E-05	5.624E-05
Importin subunit alpha-1	KPNA2	No	58	2.211E-04	1.686E-04	2.483E-04	1.914E-04
Importin subunit beta-1	KPNB1	No	97	5.348E-04	5.068E-04	1.929E-04	7.885E-04
Importin-7	IPO7	No	120	2.843E-04	7.946E-05	1.040E-04	2.718E-04
Inositol 1,4,5-trisphosphate receptor type 3	ITPR3	Yes	304	7.687E-05	1.845E-04	1.255E-04	1.905E-04
Inositol monophosphatase 3	IMPAD1	No	39	4.178E-04	5.675E-04	1.028E-03	2.429E-04
Integrin alpha-V	ITGAV	No	116	5.669E-04	2.409E-04	2.058E-04	2.779E-04
Integrin beta-1	ITGB1	No	88	9.908E-04	7.679E-05	1.057E-04	5.960E-04
Interleukin enhancer-binding factor 2	ILF2	Yes	43	2.854E-04	2.330E-04	6.166E-04	1.016E-04
Interleukin enhancer-binding factor 3	ILF3	Yes	95	6.655E-05	1.810E-04	9.980E-04	3.898E-05
Isoform 2 of 4F2 cell-surface antigen heavy chain	SLC3A2	Yes	58	5.540E-03	2.978E-03	1.997E-03	3.378E-03
Isoform 2 of Basigin	BSG	No	29	2.601E-03	7.654E-04	1.157E-03	2.060E-03
Isoform 2 of Neutral alpha-glucosidase AB	GANAB	No	109	1.828E-04	2.022E-03	2.328E-04	6.088E-04
Isoform 2 of Protein flightless-1 homolog	FLII	Yes	138	2.719E-04	3.560E-05	9.199E-05	7.195E-05
Isoform 8 of Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	No	176	3.105E-05	4.738E-05	7.443E-05	3.778E-05
Isoform 8 of Filamin-B	FLNB	Yes	282	1.492E-04	9.056E-05	7.368E-04	3.499E-04
Isoform B of Phosphate carrier protein, mitochondrial	SLC25A3	No	40	1.687E-03	3.489E-03	2.816E-03	6.302E-03
Isoleucine--tRNA ligase, cytoplasmic	IARS	Yes	145	2.978E-04	2.588E-04	1.998E-04	2.068E-04
Junction plakoglobin	JUP	No	82	1.814E-03	4.636E-04	1.329E-03	1.279E-04
Keratin, type I cytoskeletal 18	KRT18	Yes	48	1.122E-03	1.822E-02	1.552E-02	6.145E-03
Keratin, type I cytoskeletal 9	KRT9	No	62	1.071E-02	1.817E-03	2.723E-03	2.594E-03
Keratin, type II cytoskeletal 8	KRT8	Yes	54	3.399E-03	1.712E-02	8.771E-03	8.574E-03
Lamin-B1	LMNB1	No	66	2.933E-04	1.527E-03	1.041E-03	1.318E-03
Lamin-B2	LMNB2	No	68	3.822E-04	4.830E-04	4.778E-04	9.142E-04
Lanosterol 14-alpha demethylase	CYP51A1	Yes	57	5.495E-04	2.596E-04	2.618E-04	3.495E-04
Lanosterol synthase	LSS	No	83	1.315E-03	8.772E-04	5.950E-04	1.196E-03
Large neutral amino acids transporter small subunit 1	SLC7A5	Yes	55	4.282E-03	1.890E-03	3.967E-03	1.810E-03
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	No	158	2.388E-04	2.097E-04	1.729E-04	6.403E-05
Leucine--tRNA ligase, cytoplasmic	LARS	No	134	2.078E-04	6.771E-05	1.307E-04	3.732E-05
Leucyl-cystinyl aminopeptidase	LNPEP	No	117	6.738E-05	1.103E-04	8.337E-05	1.006E-04
Long-chain fatty acid transport protein 4	SLC27A4	No	72	1.079E-04	4.847E-04	2.237E-04	3.633E-04
Long-chain-fatty-acid--CoA ligase 1	ACSL1	No	78	2.810E-04	1.256E-04	2.244E-04	1.321E-04
Long-chain-fatty-acid--CoA ligase 3	ACSL3	No	80	7.695E-04	4.678E-04	1.977E-03	1.595E-03
Lysophosphatidylcholine acyltransferase 1	LPCAT1	No	59	1.417E-04	1.172E-03	9.762E-04	4.268E-04
Lysophospholipid acyltransferase 7	MBOAT7	No	53	3.695E-04	8.962E-04	8.177E-04	1.165E-03
Lysosome membrane protein 2	SCARB2	No	54	4.648E-04	3.148E-04	5.282E-04	9.268E-05
Lysosome-associated membrane glycoprotein 1	LAMP1	No	45	8.329E-04	1.061E-04	6.918E-04	1.864E-04
Malectin	MLEC	No	32	3.391E-04	6.098E-04	3.749E-04	7.054E-04
Mannosyl-oligosaccharide glucosidase	MOGS	No	92	1.523E-04	3.637E-04	8.602E-05	4.438E-04
Matrin-3	MATR3	No	95	1.730E-04	4.089E-04	9.166E-04	3.121E-04
Membrane-associated progesterone receptor component 2	PGRMC2	No	24	8.826E-04	6.087E-04	1.486E-03	8.012E-04
Metaxin-1	MTX1	Yes	51	3.395E-04	3.344E-04	2.826E-04	2.286E-04
Methionine--tRNA ligase, cytoplasmic	MARS	No	101	4.663E-04	2.148E-04	5.110E-04	1.331E-04
Microsomal glutathione S-transferase 1	MGST1	Yes	18	4.369E-03	4.325E-03	5.350E-03	1.722E-03
Mimitin, mitochondrial	NDUFAF2	No	20	7.486E-04	5.259E-04	8.510E-04	2.894E-04
Minor histocompatibility antigen H13	HM13	No	41	9.911E-04	1.422E-03	3.145E-03	1.738E-03
Mitochondrial 2-oxoglutarate/malate carrier protein	SLC25A11	No	34	6.138E-04	2.929E-03	1.754E-03	1.878E-03
Mitochondrial antiviral-signaling protein	MAVS	No	57	3.595E-04	3.747E-04	5.741E-04	3.339E-04
Mitochondrial import receptor subunit TOM40 homolog	TOMM40	No	38	2.672E-04	4.984E-03	8.350E-04	9.523E-04

Mitochondrial inner membrane protein	IMMT	No	84	3.137E-04	1.515E-03	1.074E-03	1.124E-03
Multidrug resistance-associated protein 1	ABCC1	No	172	1.438E-04	1.013E-04	2.733E-04	1.056E-04
Myb-binding protein 1A	MYBBP1A	No	149	1.921E-04	6.198E-04	1.361E-04	2.982E-04
Myelin protein zero-like protein 1	MPZL1	Yes	29	3.525E-04	1.055E-04	4.051E-04	1.989E-04
Myoferlin	MYOF	Yes	235	1.385E-03	6.931E-04	2.666E-04	2.321E-03
Myosin light polypeptide 6	MYL6	Yes	17	5.321E-03	5.345E-04	6.053E-03	1.332E-03
Myristoylated alanine-rich C-kinase substrate	MARCKS	No	32	1.267E-03	2.210E-04	1.190E-03	7.684E-04
N-acetylgalactosaminyltransferase 7	GALNT7	No	75	7.824E-05	2.027E-04	3.097E-04	2.065E-04
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	NDUFB10	No	21	6.141E-04	1.239E-03	1.256E-03	8.137E-04
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	NDUFS3	No	30	3.694E-04	1.001E-03	3.151E-04	5.856E-04
NADH-cytochrome b5 reductase 3	CYB5R3	No	34	1.303E-03	4.842E-04	3.254E-04	1.295E-03
NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	NDUFS1	Yes	79	2.728E-04	3.445E-04	1.983E-04	1.760E-04
NADPH--cytochrome P450 reductase	POR	No	77	6.505E-04	2.419E-03	2.431E-03	5.886E-04
Neuroblast differentiation-associated protein AHNAK	AHNAK	No	629	1.573E-05	4.805E-05	6.648E-04	2.724E-04
Neutral amino acid transporter B(0)	SLC1A5	No	57	1.983E-03	2.021E-03	1.323E-03	2.572E-03
Neutral cholesterol ester hydrolase 1	NCEH1	No	46	1.051E-03	5.696E-04	1.137E-03	1.050E-03
Nicalin	NCLN	No	63	1.376E-04	4.630E-04	3.402E-04	5.826E-04
Nicastrin	NCSTN	No	78	4.985E-04	1.640E-04	1.192E-04	2.474E-04
Niemann-Pick C1 protein	NPC1	No	142	4.518E-05	3.088E-05	5.634E-05	5.112E-05
Nodal modulator 3	NOMO3	No	134	8.246E-05	1.412E-04	5.985E-05	3.009E-04
Nuclear pore complex protein Nup155	NUP155	Yes	155	8.884E-05	4.439E-05	1.373E-04	3.835E-05
Nuclear pore complex protein Nup160	NUP160	Yes	162	2.771E-05	5.910E-05	1.170E-04	3.755E-05
Nuclear pore complex protein Nup205	NUP205	Yes	228	3.641E-05	5.019E-05	1.136E-04	3.407E-05
Nucleolar GTP-binding protein 1	GTPBP4	Yes	74	6.554E-05	3.246E-04	1.136E-04	4.260E-04
Nucleolar protein 56	NOP56	No	66	4.792E-04	2.369E-04	5.828E-04	3.705E-04
Nucleolar protein 58	NOP58	No	60	2.277E-04	1.228E-04	2.057E-04	1.849E-04
Nucleolar RNA helicase 2	DDX21	Yes	87	1.070E-04	2.848E-04	8.249E-04	5.514E-04
Nucleolin	NCL	Yes	77	1.980E-04	1.809E-04	4.750E-04	7.698E-05
OCIA domain-containing protein 1	OCIAD1	No	28	3.738E-04	6.094E-04	1.972E-04	4.408E-04
Pericentriolar material 1 protein	PCM1	No	229	2.940E-05	3.039E-05	1.586E-04	2.948E-05
Plasma membrane calcium-transporting ATPase 1	ATP2B1	No	139	5.285E-04	6.289E-04	2.783E-04	5.015E-04
Plasma membrane calcium-transporting ATPase 4	ATP2B4	No	138	7.145E-04	3.697E-04	2.821E-04	4.747E-04
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	No	45	1.554E-04	5.958E-04	5.508E-04	2.578E-04
Plexin-B2	PLXNB2	No	205	3.141E-05	1.774E-05	1.107E-04	5.107E-05
Poly(rC)-binding protein 1	PCBP1	Yes	37	4.025E-04	5.032E-04	6.823E-04	5.823E-04
Poly(rC)-binding protein 2	PCBP2	Yes	39	6.146E-04	4.635E-04	5.975E-04	4.429E-04
Polypeptide N-acetylgalactosaminyltransferase 1	GALNT1	No	64	2.199E-04	7.294E-05	1.511E-04	1.560E-04
Polypeptide N-acetylgalactosaminyltransferase 2	GALNT2	Yes	65	3.860E-04	3.635E-04	2.800E-04	5.514E-04
Prelamin-A/C	LMNA	Yes	74	8.996E-04	7.462E-04	2.547E-03	1.960E-03
Prenylcysteine oxidase 1	PCYOX1	No	57	4.133E-04	1.173E-03	7.770E-04	6.137E-04
Probable ATP-dependent RNA helicase DDX17	DDX17	No	80	6.093E-04	2.069E-04	8.403E-04	2.250E-04
Probable ATP-dependent RNA helicase DDX5	DDX5	Yes	69	8.037E-04	4.121E-04	8.480E-04	3.013E-04
Probable cation-transporting ATPase 13A1	ATP13A1	No	133	1.006E-04	3.147E-04	1.103E-04	1.782E-04
Probable rRNA-processing protein EBP2	EBNA1BP2	No	35	1.300E-04	2.690E-04	1.982E-04	3.352E-04
Profilin-1	PFN1	Yes	15	6.841E-04	5.126E-04	1.099E-03	1.594E-03
Prohibitin	PHB	No	30	1.868E-03	6.274E-03	1.775E-03	5.594E-03
Prohibitin-2	PHB2	No	33	2.017E-03	7.605E-03	8.103E-04	6.066E-03
Prostaglandin E synthase	PTGES	No	17	1.401E-03	4.807E-04	9.462E-04	1.047E-03
Protein disulfide-isomerase A3	PDIA3	No	57	3.027E-04	7.361E-04	6.194E-04	1.475E-04
Protein disulfide-isomerase A6	PDIA6	No	48	2.800E-04	4.223E-04	1.916E-04	1.767E-04
Protein ERGIC-53	LMAN1	No	58	9.734E-05	3.400E-04	2.337E-04	4.433E-04
Protein FAM3C	FAM3C	No	25	1.252E-03	6.720E-04	7.899E-03	1.165E-03
Protein lunapark	LNP	No	48	5.859E-04	2.084E-04	5.330E-04	5.108E-04
Protein scribble homolog	SCRIB	No	175	5.744E-05	1.441E-04	5.779E-05	9.691E-05
Protein transport protein Sec16A	SEC16A	No	234	6.011E-05	8.850E-05	1.300E-04	1.803E-05
Protein transport protein Sec61 subunit beta	SEC61B	Yes	10	4.128E-03	2.271E-03	5.214E-03	3.553E-03
Protein XRP2	RP2	No	40	4.024E-04	1.390E-04	2.691E-04	3.487E-04
Putative ribosomal RNA methyltransferase NOP2	NOP2	No	89	1.293E-04	1.244E-04	4.168E-04	7.278E-05
Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	PDHA1	Yes	43	3.223E-04	7.273E-05	5.563E-04	6.407E-05
Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	PDHB	No	39	3.483E-04	4.401E-05	1.317E-03	6.547E-05
Pyruvate kinase PKM	PKM	No	58	9.770E-04	5.304E-04	2.901E-04	9.913E-04
Rab3 GTPase-activating protein non-catalytic subunit	RAB3GAP2	No	156	6.953E-05	5.832E-05	1.114E-04	6.703E-05
Ragulator complex protein LAMTOR1	LAMTOR1	No	18	1.085E-03	7.420E-04	2.228E-03	1.194E-03
Ras GTPase-activating-like protein IQGAP1	IQGAP1	No	189	2.791E-04	2.560E-04	4.935E-04	3.049E-04
Ras-related protein Rab-11B	RAB11B	No	24	4.194E-03	1.169E-02	1.742E-03	7.883E-03
Ras-related protein Rab-21	RAB21	No	24	1.215E-03	1.443E-03	1.712E-03	1.899E-03
Ras-related protein Rab-5A	RAB5A	No	24	8.827E-04	8.357E-04	7.218E-04	1.442E-03
Ras-related protein Rab-5B	RAB5B	No	24	1.019E-03	9.030E-04	1.582E-03	1.498E-03
Ras-related protein Rab-5C	RAB5C	No	23	2.612E-03	1.855E-03	7.989E-03	4.691E-03
Ras-related protein Rab-7a	RAB7A	No	23	3.645E-03	9.868E-03	5.129E-03	4.567E-03

Ras-related protein Rab-9A	RAB9A	Yes	23	5.852E-04	4.985E-04	1.075E-03	4.895E-04
Ras-related protein R-Ras	RRAS	No	23	7.140E-04	2.751E-04	5.543E-04	8.791E-04
Ras-related protein R-Ras2	RRAS2	Yes	23	5.957E-04	1.780E-04	7.003E-04	1.124E-03
Redox-regulatory protein FAM213A	FAM213A	No	26	1.043E-03	7.741E-04	8.936E-04	9.407E-04
Retinol dehydrogenase 11	RDH11	No	35	9.361E-04	1.094E-03	1.358E-03	1.275E-03
Ribonuclease inhibitor	RNH1	No	50	1.680E-04	4.331E-05	2.051E-04	1.273E-04
Ribosomal L1 domain-containing protein 1	RSL1D1	No	55	1.129E-04	3.171E-04	3.948E-04	2.704E-04
Ribosome-binding protein 1	RRBP1	Yes	152	7.022E-05	3.806E-04	2.146E-04	3.868E-04
RNA-binding protein 14	RBM14	No	69	7.581E-04	1.133E-04	7.555E-04	1.096E-04
RRP12-like protein	RRP12	No	144	4.888E-05	2.183E-04	9.163E-05	1.400E-04
RuvB-like 1	RUVBL1	No	50	7.063E-04	2.657E-04	1.280E-04	4.077E-04
RuvB-like 2	RUVBL2	No	51	3.969E-04	2.965E-04	1.817E-04	3.279E-04
Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	Yes	115	2.471E-03	2.367E-03	3.009E-03	2.595E-03
Sec1 family domain-containing protein 1	SCFD1	No	72	3.786E-04	1.364E-04	1.836E-04	1.134E-04
Secretory carrier-associated membrane protein 1	SCAMP1	Yes	38	2.638E-04	1.087E-03	1.266E-03	5.717E-04
Secretory carrier-associated membrane protein 3	SCAMP3	No	38	1.236E-03	1.727E-03	5.074E-03	2.023E-03
Serine palmitoyltransferase 1	SPTLC1	No	53	1.891E-04	2.904E-04	4.988E-04	1.128E-04
Serine/arginine repetitive matrix protein 2	SRRM2	No	300	7.861E-05	1.938E-04	1.028E-04	5.578E-05
Serine/arginine-rich splicing factor 1	SRSF1	No	28	6.375E-04	8.009E-04	1.161E-03	1.203E-04
Sideroflexin-1	SFXN1	No	36	1.373E-03	4.076E-03	1.610E-03	2.033E-03
Signal recognition particle receptor subunit alpha	SRPR	Yes	70	9.937E-05	4.094E-04	3.190E-04	1.008E-03
Signal recognition particle receptor subunit beta	SRPRB	Yes	30	3.678E-04	3.413E-03	8.425E-04	2.578E-03
Signal recognition particle subunit SRP68	SRP68	No	71	1.011E-04	1.540E-04	1.759E-04	1.813E-04
Sn1-specific diacylglycerol lipase beta	DAGLB	No	74	6.183E-05	1.070E-04	8.855E-05	9.319E-05
Sodium/potassium-transporting ATPase subunit beta-1	ATP1B1	No	35	6.042E-04	6.128E-04	2.089E-03	6.662E-04
Sodium/potassium-transporting ATPase subunit beta-3	ATP1B3	Yes	32	3.257E-03	1.375E-03	1.072E-03	1.289E-03
Sortilin	SORT1	No	92	9.577E-05	1.114E-04	6.908E-04	3.521E-05
Spectrin beta chain, non-erythrocytic 1	SPTBN1	No	275	6.197E-05	7.056E-05	2.638E-04	1.216E-04
Splicing factor 3B subunit 1	SF3B1	No	146	1.002E-04	4.170E-04	7.102E-04	1.863E-04
Splicing factor 3B subunit 3	SF3B3	No	136	4.545E-05	6.391E-05	3.572E-04	2.721E-05
Splicing factor U2AF 35 kDa subunit	U2AF1	Yes	28	1.995E-04	4.079E-04	3.000E-04	3.178E-04
Stomatin-like protein 2, mitochondrial	STOML2	No	39	1.154E-03	1.649E-03	1.685E-03	1.822E-03
Stress-70 protein, mitochondrial	HSPA9	No	74	1.778E-04	4.418E-04	6.324E-04	3.793E-04
Sulfide:quinone oxidoreductase, mitochondrial	SQRDL	No	50	1.318E-03	1.105E-03	6.126E-04	3.655E-04
Surfeit locus protein 4	SURF4	No	30	1.137E-03	4.483E-03	2.979E-03	2.138E-03
Synaptic vesicle membrane protein VAT-1 homolog	VAT1	No	42	9.917E-04	1.767E-04	3.024E-04	2.727E-04
Synaptogyrin-2	SYNGR2	No	25	1.035E-03	1.099E-03	1.824E-03	7.004E-04
Synaptosomal-associated protein 23	SNAP23	No	23	5.191E-04	3.898E-04	8.186E-04	5.874E-04
Syntaxin-12	STX12	No	32	7.668E-04	2.793E-04	3.960E-04	3.314E-04
Syntaxin-4	STX4	No	34	2.157E-04	2.577E-04	2.004E-04	1.652E-04
Syntaxin-7	STX7	No	30	1.383E-03	2.812E-04	1.203E-03	1.235E-03
Tapasin	TAPBP	No	48	9.275E-05	8.929E-05	2.136E-04	1.009E-04
TAR DNA-binding protein 43	TARDBP	Yes	45	1.786E-04	3.878E-04	6.338E-04	8.458E-05
T-complex protein 1 subunit alpha	TCP1	Yes	60	1.658E-03	2.490E-04	6.244E-04	3.487E-04
T-complex protein 1 subunit beta	CCT2	Yes	57	4.942E-04	4.090E-04	1.022E-03	1.403E-04
T-complex protein 1 subunit delta	CCT4	Yes	58	1.500E-03	8.872E-04	9.537E-04	4.238E-04
T-complex protein 1 subunit epsilon	CCT5	Yes	60	4.224E-04	3.988E-04	6.933E-04	2.780E-04
T-complex protein 1 subunit eta	CCT7	No	59	1.003E-03	2.893E-04	6.382E-04	2.071E-04
T-complex protein 1 subunit gamma	CCT3	No	61	2.349E-03	4.811E-04	1.114E-03	7.137E-04
T-complex protein 1 subunit theta	CCT8	No	60	1.419E-03	4.377E-04	9.818E-04	3.781E-04
Thioredoxin-related transmembrane protein 1	TMX1	No	32	8.054E-04	1.236E-03	4.344E-04	2.046E-03
Torsin-1A-interacting protein 1	TOR1AIP1	No	66	6.419E-04	5.155E-04	4.495E-04	7.053E-04
Transferrin receptor protein 1	TFRC	Yes	85	7.989E-04	2.417E-03	2.961E-03	1.440E-03
Transitional endoplasmic reticulum ATPase	VCP	No	89	6.644E-05	1.481E-04	2.555E-04	3.517E-04
Translational activator GCN1	GCN1L1	No	293	3.371E-04	3.846E-04	1.111E-04	1.347E-04
Translocation protein SEC63 homolog	SEC63	No	88	1.622E-04	4.479E-04	1.895E-04	3.534E-04
Translocon-associated protein subunit delta	SSR4	No	19	1.832E-03	1.543E-03	2.682E-03	3.477E-03
Transmembrane 9 superfamily member 2	TM9SF2	No	76	3.579E-04	9.109E-04	1.671E-03	6.958E-04
Transmembrane 9 superfamily member 4	TM9SF4	No	75	2.622E-04	5.413E-04	6.706E-04	4.030E-04
Transmembrane and TPR repeat-containing protein 3	TMTC3	No	104	1.039E-04	2.046E-04	1.822E-04	1.134E-04
Transmembrane emp24 domain-containing protein 10	TMED10	No	25	1.807E-03	2.763E-03	1.248E-03	1.680E-03
Transmembrane emp24 domain-containing protein 3	TMED3	No	25	3.668E-04	1.047E-03	1.568E-03	1.341E-04
Transmembrane emp24 domain-containing protein 7	TMED7	No	25	6.762E-04	1.190E-03	6.971E-04	8.233E-04
Transmembrane protein 214	TMEM214	No	77	2.126E-04	1.408E-04	8.802E-05	1.578E-04
Transmembrane protein 245	TMEM245	No	101	1.704E-04	2.033E-04	2.853E-04	3.275E-04
Transmembrane protein 43	TMEM43	No	45	1.102E-03	1.081E-03	5.518E-04	1.472E-03
Tumor-associated calcium signal transducer 2	TACSTD2	Yes	36	7.631E-04	5.672E-04	2.894E-03	1.071E-04
Tyrosine-protein phosphatase non-receptor type 1	PTPN1	Yes	50	8.677E-04	1.890E-03	1.066E-03	1.504E-03
UBX domain-containing protein 4	UBXN4	No	57	3.114E-04	3.101E-04	1.446E-04	2.535E-04
Ufm1-specific protease 2	UFSP2	No	53	1.697E-04	1.206E-04	1.801E-04	1.062E-04
Unconventional myosin-VI	MYO6	Yes	150	4.216E-04	4.911E-04	8.026E-04	8.448E-05
Up-regulated during skeletal muscle growth protein 5	USMG5	No	6	2.775E-03	2.526E-03	2.270E-03	7.567E-04
Valine--tRNA ligase	VARS	Yes	140	4.114E-04	5.409E-05	5.749E-04	9.259E-05

Vesicle-associated membrane protein 3	VAMP3	No	11	4.012E-03	6.814E-04	2.956E-03	2.100E-03
Vesicle-associated membrane protein-associated protein A	VAPA	No	28	1.349E-03	3.233E-03	5.534E-03	3.729E-03
Vesicle-associated membrane protein-associated protein B/C	VAPB	No	27	1.512E-03	2.221E-03	3.607E-03	1.675E-03
Vesicle-fusing ATPase	NSF	Yes	83	9.373E-05	1.714E-04	9.876E-05	6.573E-05
Vesicle-trafficking protein SEC22b	SEC22B	No	25	1.620E-03	3.051E-03	3.404E-03	3.191E-03
Vesicular integral-membrane protein VIP36	LMAN2	No	40	5.346E-04	2.298E-03	1.839E-03	2.321E-03
Voltage-dependent anion-selective channel protein 1	VDAC1	No	31	5.704E-03	5.536E-03	1.105E-03	3.210E-03
Voltage-dependent anion-selective channel protein 2	VDAC2	No	32	4.086E-03	2.076E-03	2.860E-03	1.531E-03
X-ray repair cross-complementing protein 5	XRCC5	No	83	2.136E-04	1.538E-04	3.562E-04	7.972E-05
X-ray repair cross-complementing protein 6	XRCC6	No	70	7.234E-04	1.923E-04	4.343E-04	1.068E-04
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	No	101	1.319E-04	5.888E-05	1.469E-04	1.142E-04
Identified membrane proteins present in breast cancer cell lines only							
		BC	MW	Average NSAF			
Identified Proteins	**Gene**	**database**	**(kDa)**	**HMEC**	**MCF7**	**SKBR3**	**MDA231**
116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	No	109		1.548E-04	1.490E-04	1.671E-04
14-3-3 protein gamma	YWHAG	No	28		2.259E-04	4.326E-04	4.349E-04
14-3-3 protein zeta/delta	YWHAZ	Yes	28		6.166E-04	1.364E-03	1.517E-03
39S ribosomal protein L49, mitochondrial	MRPL49	No	19		3.849E-04	8.743E-04	5.921E-04
40S ribosomal protein S15	RPS15	No	17		5.826E-04	1.937E-03	1.036E-03
40S ribosomal protein S17	RPS17	No	16		5.966E-04	1.803E-03	1.880E-03
60S ribosomal protein L15	RPL15	No	24		1.650E-03	1.424E-03	1.612E-03
60S ribosomal protein L23	RPL23	No	15		5.450E-04	7.795E-04	5.806E-04
60S ribosomal protein L27	RPL27	No	16		5.299E-04	8.108E-04	1.408E-03
60S ribosomal protein L35	RPL35	No	15		7.595E-04	9.768E-04	2.343E-03
60S ribosomal protein L8	RPL8	No	28		5.027E-04	1.107E-03	1.662E-03
6-phosphofructokinase, liver type	PFKL	Yes	85		8.636E-05	1.035E-04	2.518E-05
Acetyl-coenzyme A transporter 1	SLC33A1	No	61		1.484E-04	1.716E-04	9.791E-05
Actin-related protein 2	ACTR2	No	45		1.348E-04	5.198E-04	8.740E-05
Actin-related protein 2/3 complex subunit 3	ARPC3	Yes	21		2.940E-04	1.401E-03	2.498E-04
Actin-related protein 3	ACTR3	No	47		1.707E-04	4.546E-04	9.438E-05
Acyl-CoA desaturase	SCD	Yes	42		1.654E-03	7.574E-04	8.311E-05
Acyl-CoA-binding domain-containing protein 5	ACBD5	No	60		1.217E-04	2.675E-04	1.752E-04
ADP-ribosylation factor 3	ARF3	No	21		1.037E-03	3.341E-03	1.323E-03
AFG3-like protein 2	AFG3L2	No	89		1.770E-04	1.371E-04	1.481E-04
Alpha-1,2-mannosyltransferase ALG9	ALG9	No	70		1.390E-04	1.773E-04	1.074E-04
Ancient ubiquitous protein 1	AUP1	No	53		1.912E-04	3.798E-04	3.137E-04
AP-2 complex subunit alpha-1	AP2A1	No	108		1.183E-04	1.978E-04	1.447E-04
Apolipoprotein L2	APOL2	No	37		1.084E-04	6.702E-04	8.853E-05
Aspartyl/asparaginyl beta-hydroxylase	ASPH	No	86		7.454E-05	6.397E-05	1.762E-04
ATP-binding cassette sub-family D member 1	ABCD1	No	83		1.137E-04	3.685E-04	1.126E-04
ATP-binding cassette sub-family E member 1	ABCE1	No	67		4.260E-05	2.011E-04	4.981E-05
ATP-citrate synthase	ACLY	No	121		2.625E-05	1.828E-04	5.429E-04
Autophagy-related protein 9A	ATG9A	No	94		7.708E-05	8.737E-05	1.599E-04
CAAX prenyl protease 1 homolog	ZMPSTE24	No	55		5.579E-04	1.809E-03	6.250E-04
Calcium uniporter protein, mitochondrial	MCU	No	40		1.544E-04	2.721E-04	6.860E-05
Calcium-binding mitochondrial carrier protein Aralar1	SLC25A12	No	75		3.389E-04	2.638E-04	2.691E-04
Cation-dependent mannose-6-phosphate receptor	M6PR	Yes	31		4.132E-04	1.648E-03	5.068E-04
Cation-independent mannose-6-phosphate receptor	IGF2R	No	274		2.343E-05	1.189E-04	4.141E-05
CCR4-NOT transcription complex subunit 11	CNOT1	No	267		5.979E-05	2.041E-05	3.103E-05
Charged multivesicular body protein 4b	CHMP4B	No	25		1.655E-04	5.871E-04	2.374E-04
Chitobiosyldiphosphodolichol beta-mannosyltransferase	ALG1	No	53		2.075E-04	4.386E-04	9.334E-05
Chloride channel CLIC-like protein 1	CLCC1	No	62		2.766E-04	3.449E-04	9.710E-05
Clathrin interactor 1	CLINT1	No	68		3.467E-04	1.728E-04	1.991E-04
Cluster of 2-oxoglutarate dehydrogenase, mitochondrial	OGDH	No	116		1.273E-04	2.126E-04	2.426E-05
Cluster of 60S ribosomal protein L26	RPL26	No	17		1.418E-03	1.770E-03	1.041E-03
Cluster of 60S ribosomal protein L36a-like	RPL36AL	No	12		3.374E-04	1.380E-03	4.962E-04
Cluster of ADP-ribosylation factor-like protein 8B	ARL8B	No	22		7.579E-04	7.812E-04	4.450E-04
Cluster of Afadin	MLLT4	Yes	207		4.657E-05	7.210E-05	4.317E-05
Cluster of E3 SUMO-protein ligase RanBP2	RANBP2	No	358		6.125E-05	8.816E-05	2.950E-05
Cluster of ER lumen protein retaining receptor 2	KDELR2	No	24		1.641E-03	1.596E-03	5.478E-04
Cluster of Hexokinase-1	HK1	Yes	102		9.529E-04	2.502E-04	8.098E-04
Cluster of Isoform B of Protein SON	SON	Yes	250		7.644E-05	2.004E-04	5.241E-05
Cluster of Kinesin-1 heavy chain	KIF5B	Yes	110		1.652E-04	5.035E-05	2.102E-04
Cluster of Mitochondrial glutamate carrier 1	SLC25A22	No	34		1.613E-04	1.848E-04	1.221E-04
Cluster of Nucleoside diphosphate kinase A	NME1	Yes	17		7.023E-04	1.013E-03	3.304E-04
Cluster of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	SMARCA5	No	122		1.324E-04	1.376E-04	1.883E-05
Cluster of Talin-1	TLN1	No	270		3.191E-05	7.039E-05	2.580E-04
Cluster of Transcription activator BRG1	SMARCA4	No	185		9.889E-05	1.262E-04	5.973E-05
Cluster of Ubiquitin-40S ribosomal protein S27a	RPS27A	No	18		3.783E-03	1.705E-02	2.967E-02
Coiled-coil domain-containing protein 47	CCDC47	No	56		6.832E-04	2.958E-04	1.683E-04
Coiled-coil-helix-coiled-coil-helix domain-containing protein 6, mitochondrial	CHCHD6	No	26		3.572E-04	4.666E-04	2.245E-04

Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	No	122		2.329E-04	3.369E-04	3.066E-05
Cytochrome b5	CYB5A	No	15		9.163E-04	2.550E-03	7.599E-04
Cytochrome b-c1 complex subunit 8	UQCRQ	No	10		2.727E-03	2.804E-03	3.097E-03
Cytochrome c oxidase assembly protein 3 homolog, mitochondrial	COA3	No	12		3.869E-04	5.737E-04	6.009E-04
Cytochrome c oxidase subunit 5A, mitochondrial	COX5A	No	17		6.886E-04	1.133E-03	4.285E-04
Cytochrome c1, heme protein, mitochondrial	CYC1	No	35		2.051E-03	1.302E-03	8.775E-04
Cytochrome c-type heme lyase	HCCS	Yes	31		2.669E-04	7.231E-04	4.591E-04
Cytosolic non-specific dipeptidase	CNDP2	No	53		6.199E-05	1.987E-04	8.496E-05
Desmoglein-2	DSG2	Yes	122		5.081E-05	2.037E-04	1.095E-04
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	DLAT	No	69		5.039E-05	2.794E-04	1.800E-04
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	No	49		8.355E-05	2.083E-04	4.209E-05
Dihydroorotate dehydrogenase (quinone), mitochondrial	DHODH	No	43		3.487E-04	3.612E-04	1.837E-04
Disks large homolog 1	DLG1	Yes	100		4.971E-05	1.084E-04	7.092E-05
DNA replication licensing factor MCM7	MCM7	Yes	81		6.433E-05	6.744E-05	6.859E-05
DnaJ homolog subfamily B member 12	DNAJB12	No	42		2.230E-04	4.173E-04	9.139E-05
DnaJ homolog subfamily C member 13	DNAJC13	No	254		2.834E-04	1.083E-04	2.091E-04
E3 ubiquitin-protein ligase RNF213	RNF213	No	591		1.961E-05	1.855E-04	8.230E-05
Eukaryotic translation initiation factor 2 subunit 1	EIF2S1	No	36		6.509E-05	3.073E-04	6.053E-05
Eukaryotic translation initiation factor 3 subunit I	EIF3I	No	37		1.214E-04	1.452E-04	1.350E-04
Exportin-5	XPO5	No	136		1.965E-05	4.956E-05	1.632E-05
Fatty acyl-CoA reductase 1	FAR1	No	59		4.408E-04	3.907E-04	9.127E-04
Fragile X mental retardation syndrome-related protein 2	FXR2	Yes	74		6.723E-05	9.036E-05	1.511E-04
Golgi SNAP receptor complex member 1	GOSR1	Yes	29		3.577E-04	1.129E-03	1.992E-04
GPI transamidase component PIG-S	PIGS	No	62		3.234E-04	3.220E-04	1.727E-04
GPI-anchor transamidase	PIGK	No	45		5.413E-04	4.331E-04	2.776E-04
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	No	31		3.460E-04	6.343E-04	9.836E-05
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	No	46		2.422E-04	6.455E-04	4.773E-05
Heterogeneous nuclear ribonucleoprotein U-like protein 1	HNRNPUL1	No	96		1.229E-04	2.688E-04	1.757E-04
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	No	85		3.543E-04	1.443E-04	8.527E-05
Histone H1.0	H1F0	No	21		4.275E-04	3.653E-04	6.338E-04
Histone H3.1t	HIST3H3	No	16		2.261E-03	7.093E-04	5.801E-03
Homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 1 protein	HERPUD1	No	44		6.388E-05	1.493E-04	2.472E-04
Importin-5	IPO5	No	124		1.193E-04	6.579E-05	2.144E-04
Integral membrane protein GPR180	GPR180	No	49		1.123E-04	4.831E-04	1.225E-04
Inverted formin-2	INF2	No	136		2.089E-04	4.246E-04	9.555E-05
Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1	No	47		1.272E-04	2.585E-04	4.605E-05
Isoform 2 of Dynamin-like 120 kDa protein, mitochondrial	OPA1	No	116		4.596E-04	3.604E-04	1.812E-04
Isoform 3 of Heterogeneous nuclear ribonucleoprotein K	HNRNPK	No	49		1.019E-03	1.029E-03	4.702E-04
Isoform 9 of SUN domain-containing protein 1	SUN1	No	102		1.496E-04	1.466E-04	2.740E-04
Isoform A of Syntaxin-16	STX16	No	35		3.804E-04	4.092E-04	1.219E-04
Kinectin	KTN1	No	156		8.679E-05	1.782E-04	2.051E-04
Laminin subunit alpha-5	LAMA5	Yes	400		1.389E-05	5.876E-05	2.527E-05
LEM domain-containing protein 2	LEMD2	No	57		2.087E-04	3.088E-04	5.899E-05
LETM1 and EF-hand domain-containing protein 1, mitochondrial	LETM1	No	83		2.481E-04	1.630E-04	9.976E-05
Major facilitator superfamily domain-containing protein 1	MFSD1	No	51		1.147E-04	4.621E-04	2.409E-04
Membrane-associated progesterone receptor component 1	PGRMC1	No	22		7.364E-04	1.975E-03	1.132E-03
Metallo-beta-lactamase domain-containing protein 2	MBLAC2	No	31		1.622E-04	4.400E-04	2.151E-04
Microsomal glutathione S-transferase 3	MGST3	No	17		1.046E-03	1.486E-03	5.667E-04
Mitochondrial dicarboxylate carrier	SLC25A10	No	31		4.281E-04	1.045E-03	8.447E-05
Mitochondrial import inner membrane translocase subunit TIM50	TIMM50	No	40		6.813E-04	6.054E-04	5.744E-04
Mitochondrial import receptor subunit TOM22 homolog	TOMM22	No	16		1.869E-03	3.532E-03	2.714E-03
Mitochondrial import receptor subunit TOM70	TOMM70A	Yes	67		2.376E-04	1.408E-04	1.536E-04
MKI67 FHA domain-interacting nucleolar phosphoprotein	NIFK	No	34		2.928E-04	2.878E-04	1.433E-04
Monoacylglycerol lipase ABHD12	ABHD12	No	45		4.168E-04	2.114E-04	2.573E-04
N-acetyltransferase 10	NAT10	No	116		4.129E-05	2.237E-04	1.258E-04
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	NDUFA10	No	41		2.349E-04	2.370E-04	1.523E-04
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	NDUFA13	No	17		1.758E-03	1.163E-03	1.152E-03
Nck-associated protein 1	NCKAP1	No	129		4.360E-05	6.166E-05	4.714E-05
Neurogenic locus notch homolog protein 2	NOTCH2	No	265		6.394E-06	2.454E-05	2.002E-05
Nuclear pore complex protein Nup93	NUP93	No	93		7.350E-05	2.644E-04	8.641E-05
Nuclear pore membrane glycoprotein 210	NUP210	No	205		4.155E-04	1.138E-04	2.139E-05
Nuclease-sensitive element-binding protein 1	YBX1	No	36		1.234E-03	1.003E-03	1.996E-04
Peptidyl-prolyl cis-trans isomerase A	PPIA	Yes	18		3.092E-04	1.431E-03	5.982E-04
Peroxisomal membrane protein 11B	PEX11B	No	28		3.818E-04	1.011E-03	2.802E-04
Peroxisomal membrane protein PEX13	PEX13	Yes	44		1.182E-04	1.758E-04	1.309E-04
Peroxisomal membrane protein PEX16	PEX16	No	39		1.120E-04	2.177E-04	1.284E-04
Phosphatidylinositide phosphatase SAC1	SACM1L	No	67		1.775E-04	3.657E-04	3.156E-04

Pinin	PNN	No	82		1.580E-04	4.546E-04	5.502E-05
Polyadenylate-binding protein 2	PABPN1	No	33		1.645E-04	2.390E-04	6.473E-05
Pre-mRNA-processing factor 19	PRPF19	No	55		1.138E-04	2.651E-04	7.451E-05
Pre-mRNA-processing-splicing factor 8	PRPF8	No	274		2.058E-04	1.653E-04	1.002E-04
Probable ATP-dependent RNA helicase DDX27	DDX27	No	90		1.173E-04	1.835E-04	1.117E-04
Programmed cell death 6-interacting protein	PDCD6IP	No	96		6.172E-05	3.324E-04	1.579E-04
Prolactin regulatory element-binding protein	PREB	No	45		4.251E-04	2.881E-04	7.016E-05
Protein disulfide-isomerase	P4HB	Yes	57		2.361E-04	2.346E-04	3.867E-05
Protein FAM49B	FAM49B	Yes	37		1.137E-04	1.877E-04	2.864E-04
Protein LYRIC	MTDH	No	64		4.233E-04	1.254E-04	6.253E-04
Protein S100-A11	S100A11	Yes	12		2.804E-04	5.808E-04	7.163E-04
Protein sel-1 homolog 1	SEL1L	No	89		5.330E-05	7.516E-05	5.475E-05
Protein THEM6	THEM6	No	24		3.035E-04	6.511E-04	1.897E-04
Protein transport protein Sec23A	SEC23A	No	86		4.262E-05	1.107E-04	1.284E-04
Protein transport protein Sec23B	SEC23B	No	86		7.025E-05	1.588E-04	1.092E-04
Protein YIF1A	YIF1A	No	32		3.717E-04	5.684E-04	1.453E-04
Ras-related protein Rab-13	RAB13	No	23		2.384E-04	9.457E-04	3.377E-04
Ras-related protein Rab-22A	RAB22A	Yes	22		6.667E-04	8.040E-04	2.987E-04
Ras-related protein Ral-A	RALA	No	24		7.214E-04	6.439E-04	9.200E-04
Ras-related protein Ral-B	RALB	Yes	23		6.039E-04	1.357E-03	6.974E-04
Receptor expression-enhancing protein 5	REEP5	No	21		4.258E-04	2.645E-03	2.487E-04
Replication factor C subunit 3	RFC3	Yes	41		6.798E-05	1.676E-04	9.465E-05
Reticulon-3	RTN3	No	113		4.743E-05	1.742E-04	1.264E-04
Ribosomal biogenesis protein LAS1L	LAS1L	No	83		6.091E-05	1.944E-04	8.577E-05
RNA-binding protein EWS	EWSR1	No	68		1.713E-04	1.857E-04	2.482E-04
Secretory carrier-associated membrane protein 2	SCAMP2	Yes	37		4.805E-04	2.558E-04	5.276E-04
Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	No	32		9.268E-04	4.197E-04	2.383E-04
Small nuclear ribonucleoprotein Sm D2	SNRPD2	Yes	14		1.738E-04	2.145E-03	2.842E-04
Solute carrier family 12 member 9	SLC12A9	No	96		1.214E-04	4.283E-04	2.239E-04
Solute carrier family 22 member 18	SLC22A18	Yes	45		2.291E-04	2.578E-04	1.516E-04
Solute carrier family 35 member E1	SLC35E1	No	45		6.396E-04	2.302E-04	4.320E-04
Sphingosine-1-phosphate lyase 1	SGPL1	No	64		2.417E-04	2.755E-04	2.856E-04
Splicing factor, proline- and glutamine-rich	SFPQ	No	76		5.800E-05	1.699E-04	6.417E-05
SPRY domain-containing protein 7	SPRYD7	No	22		5.502E-04	6.851E-04	3.694E-04
Staphylococcal nuclease domain-containing protein 1	SND1	No	102		1.204E-04	1.866E-04	9.854E-05
Sulfatase-modifying factor 2	SUMF2	No	34		1.905E-04	7.181E-04	1.286E-04
SURP and G-patch domain-containing protein 2	SUGP2	No	120		2.258E-05	1.016E-04	4.110E-05
Synaptogyrin-1	SYNGR1	Yes	25		2.134E-04	6.180E-04	1.686E-04
Testis-expressed sequence 264 protein	TEX264	No	34		7.732E-05	1.943E-04	2.673E-04
Thioredoxin-related transmembrane protein 2	TMX2	No	34		9.794E-04	3.226E-04	8.797E-04
Transducin beta-like protein 2	TBL2	No	50		5.771E-04	1.815E-03	2.797E-04
Transducin beta-like protein 3	TBL3	No	89		6.086E-05	8.932E-05	4.791E-05
Transgelin-2	TAGLN2	No	22		3.277E-04	5.980E-04	3.965E-04
Trans-Golgi network integral membrane protein 2	TGOLN2	No	51		1.665E-04	9.919E-04	1.829E-04
Translocase of inner mitochondrial membrane domain-containing protein 1	TIMMDC1	No	32		1.152E-04	2.088E-04	8.247E-05
Translocating chain-associated membrane protein 1	TRAM1	No	43		1.976E-04	3.850E-04	3.828E-04
Transmembrane 9 superfamily member 3	TM9SF3	No	68		5.914E-04	4.490E-04	4.940E-04
Transmembrane and coiled-coil domain-containing protein 1	TMCO1	No	21		2.582E-03	2.306E-03	3.693E-03
Transmembrane and ubiquitin-like domain-containing protein 1	TMUB1	No	26		1.830E-04	3.328E-04	1.002E-04
Transmembrane emp24 domain-containing protein 2	TMED2	No	23		1.287E-02	1.731E-03	1.410E-03
Transmembrane emp24 domain-containing protein 9	TMED9	No	27		8.747E-04	4.499E-04	9.522E-04
Transmembrane protein 109	TMEM109	No	26		2.858E-04	7.706E-04	1.312E-03
Transmembrane protein 165	TMEM165	No	35		3.614E-04	1.131E-03	1.687E-04
Transmembrane protein 205	TMEM205	No	21		4.448E-03	2.279E-03	6.284E-04
Transmembrane protein 65	TMEM65	No	25		4.727E-04	2.360E-03	1.188E-04
Transmembrane protein 70, mitochondrial	TMEM70	No	29		2.754E-04	2.269E-03	2.563E-04
Triosephosphate isomerase	TPI1	Yes	31		2.280E-04	2.563E-04	2.216E-04
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	No	245		2.003E-04	1.633E-04	4.325E-05
Ubiquitin carboxyl-terminal hydrolase 7	USP7	No	128		5.204E-05	4.385E-05	2.707E-05
Ubiquitin-conjugating enzyme E2 J1	UBE2J1	No	35		1.522E-04	4.253E-04	2.318E-04
Ubiquitin-like modifier-activating enzyme 1	UBA1	No	118		6.310E-05	6.929E-05	5.461E-05
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	No	177		8.818E-05	4.619E-05	3.525E-05
UPF0420 protein C16orf58	C16orf58	No	51		1.868E-04	2.016E-04	5.145E-05
Utrophin	UTRN	No	394		2.705E-05	2.097E-05	9.163E-05
Vasodilator-stimulated phosphoprotein	VASP	No	40		9.554E-05	2.587E-04	1.023E-04
Very-long-chain enoyl-CoA reductase	TECR	No	36		1.119E-03	1.569E-04	1.072E-03
Vesicle transport protein SEC20	BNIP1	No	26		2.492E-04	2.617E-04	1.730E-04
Vesicle-associated membrane protein 7	VAMP7	No	25		1.258E-04	3.833E-04	1.322E-04
Vesicle-associated membrane protein 8	VAMP8	No	11		7.183E-04	4.036E-03	2.771E-03
V-type proton ATPase 116 kDa subunit a isoform 1	ATP6V0A1	No	96		2.305E-04	1.290E-04	4.605E-05
V-type proton ATPase catalytic subunit A	ATP6V1A	No	68		1.091E-04	2.539E-04	5.435E-05
Wolframin	WFS1	Yes	100		7.912E-04	2.288E-04	3.644E-04

Zinc transporter 7	SLC30A7	No	42		1.406E-04	3.226E-04	2.098E-04
Identified membrane proteins present in HMEC only							
		BC	MW	Average NSAF			
Identified Proteins	**Gene**	**database**	**(kDa)**	**HMEC**	**MCF7**	**SKBR3**	**MDA231**
40S ribosomal protein S28	RPS28	No	8	3.063E-03			
60S ribosomal protein L22	RPL22	No	15	1.488E-03			
Aldehyde dehydrogenase family 1 member A3	ALDH1A3	Yes	56	4.054E-04			
Aminopeptidase N	ANPEP	No	110	8.636E-04			
Anaphase-promoting complex subunit 7	ANAPC7	No	67	1.982E-04			
AP-3 complex subunit mu-1	AP3M1	No	47	1.517E-04			
Apolipoprotein A-I	APOA1	Yes	31	2.162E-04			
Asparagine--tRNA ligase, cytoplasmic	NARS	No	63	1.080E-04			
Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	Yes	469	1.628E-05			
Bystin	BYSL	Yes	50	9.508E-05			
Calcium-activated chloride channel regulator 2	CLCA2	No	104	1.423E-04			
Calmodulin-like protein 3	CALML3	No	17	1.355E-03			
Caveolin-2	CAV2	Yes	18	1.091E-03			
Cell division cycle protein 23 homolog	CDC23	No	69	6.666E-05			
CLIP-associating protein 1	CLASP1	No	169	5.676E-05			
Cluster of Isoform 4 of Sodium- and chloride-dependent creatine transporter 1	SLC6A8	No	58	1.126E-04			
Cluster of Keratin, type I cytoskeletal 15	KRT15	No	49	1.867E-02			
Cluster of Serine protease HTRA1	HTRA1	Yes	51	4.592E-04			
Collagen alpha-1(XVII) chain	COL17A1	Yes	150	5.594E-04			
Contactin-1	CNTN1	No	113	3.544E-04			
Corneodesmosin	CDSN	No	52	8.593E-05			
CTD small phosphatase-like protein	CTDSPL	Yes	31	2.297E-04			
C-type mannose receptor 2	MRC2	No	167	4.834E-05			
CUB domain-containing protein 1	CDCP1	No	93	7.875E-05			
CXADR-like membrane protein	CLMP	No	41	1.917E-04			
Dermcidin	DCD	No	11	2.629E-03			
Desmocollin-3	DSC3	Yes	100	9.786E-05			
Desmoglein-3	DSG3	Yes	108	8.617E-04			
Dihydroxyacetone phosphate acyltransferase	GNPAT	No	77	1.165E-04			
E3 ubiquitin/ISG15 ligase TRIM25	TRIM25	No	71	1.666E-04			
E3 ubiquitin-protein ligase TRIM32	TRIM32	No	72	7.603E-05			
EH domain-containing protein 2	EHD2	No	61	1.168E-04			
Endothelial protein C receptor	PROCR	Yes	27	7.040E-04			
Ephrin-B1	EFNB1	Yes	38	5.497E-04			
Exocyst complex component 5	EXOC5	No	82	1.343E-04			
F-box only protein 2	FBXO2	No	33	2.475E-04			
Fibroblast growth factor-binding protein 1	FGFBP1	Yes	26	6.064E-04			
Fragile X mental retardation syndrome-related protein 1	FXR1	No	70	1.338E-04			
Gamma-tubulin complex component 2	TUBGCP2	No	103	1.164E-04			
Gigaxonin	GAN	No	68	9.918E-05			
Glypican-1	GPC1	No	62	1.426E-04			
Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	No	66	3.245E-04			
Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	No	62	1.579E-04			
Isoform 1 of Laminin subunit alpha-3	LAMA3	Yes	189	3.331E-04			
Isoform 2 of Alpha-aminoadipic semialdehyde dehydrogenase	ALDH7A1	No	55	1.509E-04			
Isoform 3 of Dystonin	DST	Yes	307	5.979E-05			
Lactadherin	MFGE8	Yes	43	3.294E-04			
Laminin subunit gamma-2	LAMC2	Yes	131	1.197E-03			
Leucine zipper protein 1	LUZP1	No	120	9.564E-05			
Leucine-rich repeat transmembrane protein FLRT3	FLRT3	No	73	6.045E-04			
Lysophosphatidylcholine acyltransferase 2	LPCAT2	No	60	1.061E-04			
Mediator of RNA polymerase II transcription subunit 23	MED23	No	156	8.549E-05			
Metalloreductase STEAP3	STEAP3	No	55	1.501E-04			
Moesin	MSN	Yes	68	1.611E-04			
Myosin phosphatase Rho-interacting protein	MPRIP	No	117	1.818E-04			
Nucleolar complex protein 2 homolog	NOC2L	No	85	1.345E-04			
Nucleolar pre-ribosomal-associated protein 1	URB1	No	254	2.792E-05			
Ornithine aminotransferase, mitochondrial	OAT	No	49	4.354E-04			
Phospholipid scramblase 1	PLSCR1	Yes	35	1.994E-04			
Pleckstrin homology domain-containing family A member 5	PLEKHA5	No	127	6.564E-05			
Pleckstrin homology-like domain family B member 2	PHLDB2	No	142	5.846E-05			
Poly [ADP-ribose] polymerase 4	PARP4	No	193	5.392E-05			
PRA1 family protein 2	PRAF2	No	19	9.003E-04			
Probable cation-transporting ATPase 13A3	ATP13A3	No	138	3.246E-05			
Probable phospholipid-transporting ATPase IF	ATP11B	Yes	134	6.762E-05			
Proliferation-associated protein 2G4	PA2G4	No	44	3.030E-04			

Proteasome activator complex subunit 4	PSME4	Yes	211	2.519E-04			
Protein FAM83F	FAM83F	No	55	1.465E-04			
Protein FAM98A	FAM98A	No	55	1.601E-04			
Protein furry homolog-like	FRYL	No	340	7.038E-05			
Protein S100-A8	S100A8	Yes	11	8.747E-04			
RAF proto-oncogene serine/threonine-protein kinase	RAF1	No	73	9.182E-05			
Replication factor C subunit 2	RFC2	No	39	1.681E-04			
Rho-related GTP-binding protein RhoD	RHOD	No	23	3.404E-04			
Ribosomal RNA small subunit methyltransferase NEP1	EMG1	No	27	2.294E-04			
Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon isoform	PPP2R5E	No	55	2.036E-04			
Serpin B5	SERPINB5	Yes	42	2.424E-04			
Signal-induced proliferation-associated 1-like protein 1	SIPA1L1	No	200	2.752E-05			
Sodium-coupled neutral amino acid transporter 2	SLC38A2	No	56	1.530E-04			
Sodium-dependent neutral amino acid transporter B(0)AT2	SLC6A15	No	82	8.150E-05			
SPATS2-like protein	SPATS2L	No	62	1.136E-04			
Synaptotagmin-like protein 1	SYTL1	No	62	1.027E-04			
Tripartite motif-containing protein 29	TRIM29	Yes	66	5.236E-04			
Tripartite motif-containing protein 4	TRIM4	No	57	1.627E-04			
Tyrosine-protein phosphatase non-receptor type substrate 1	SIRPA	No	55	2.320E-04			
Unconventional myosin-Ie	MYO1E	No	127	2.351E-04			
Unhealthy ribosome biogenesis protein 2 homolog	URB2	No	171	1.178E-04			
Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	No	125	3.608E-05			
Identified membrane proteins present in MCF7 only							
		BC	**MW**	**Average NSAF**			
Identified Proteins	**Gene**	**database**	**(kDa)**	**HMEC**	**MCF7**	**SKBR3**	**MDA231**
1-acyl-sn-glycerol-3-phosphate acyltransferase gamma	AGPAT3	No	43		8.610E-05		
28S ribosomal protein S18b, mitochondrial	MRPS18B	No	29		7.947E-05		
28S ribosomal protein S21, mitochondrial	MRPS21	No	11		4.571E-04		
28S ribosomal protein S22, mitochondrial	MRPS22	No	41		2.038E-04		
28S ribosomal protein S26, mitochondrial	MRPS26	No	24		1.410E-04		
28S ribosomal protein S27, mitochondrial	MRPS27	Yes	48		9.905E-05		
28S ribosomal protein S34, mitochondrial	MRPS34	No	26		1.485E-04		
28S ribosomal protein S35, mitochondrial	MRPS35	No	37		7.326E-05		
28S ribosomal protein S5, mitochondrial	MRPS5	No	48		1.090E-04		
28S ribosomal protein S7, mitochondrial	MRPS7	No	28		1.217E-04		
28S ribosomal protein S9, mitochondrial	MRPS9	No	46		1.228E-04		
39S ribosomal protein L12, mitochondrial	MRPL12	No	21		1.250E-04		
39S ribosomal protein L16, mitochondrial	MRPL16	No	28		1.446E-04		
39S ribosomal protein L17, mitochondrial	MRPL17	No	20		3.542E-04		
39S ribosomal protein L18, mitochondrial	MRPL18	No	21		1.109E-04		
39S ribosomal protein L19, mitochondrial	MRPL19	No	34		9.900E-05		
39S ribosomal protein L3, mitochondrial	MRPL3	No	39		4.540E-05		
39S ribosomal protein L45, mitochondrial	MRPL45	No	35		1.040E-04		
40S ribosomal protein S7	RPS7	No	22		4.723E-04		
5'-3' exoribonuclease 2	XRN2	No	109		2.548E-05		
5'-AMP-activated protein kinase catalytic subunit alpha-1	PRKAA1	No	64		2.826E-05		
Abhydrolase domain-containing protein 16A	ABHD16A	No	63		1.645E-04		
Actin-like protein 6A	ACTL6A	No	47		5.768E-05		
Acyl carrier protein, mitochondrial	NDUFAB1	No	17		3.844E-04		
Acylglycerol kinase, mitochondrial	AGK	No	47		6.850E-05		
Acyl-protein thioesterase 1	LYPLA1	Yes	25		1.407E-04		
Adenine phosphoribosyltransferase	APRT	No	20		8.055E-04		
Adenylate cyclase type 9	ADCY9	Yes	151		1.555E-05		
Alpha/beta hydrolase domain-containing protein 17C	ABHD17C	No	36		8.551E-05		
Alpha-mannosidase 2x	MAN2A2	No	131		5.308E-05		
Amine oxidase [flavin-containing] A	MAOA	Yes	60		3.918E-04		
Amine oxidase [flavin-containing] B	MAOB	No	59		1.479E-03		
Amphoterin-induced protein 2	AMIGO2	No	58		6.955E-05		
Anaphase-promoting complex subunit 1	ANAPC1	No	217		2.046E-05		
Anterior gradient protein 2 homolog	AGR2	Yes	20		2.074E-04		
Argininosuccinate synthase	ASS1	No	47		4.977E-05		
Armadillo repeat protein deleted in velo-cardio-facial syndrome	ARVCF	Yes	105		3.445E-05		
Armadillo repeat-containing X-linked protein 3	ARMCX3	No	43		1.269E-04		
Arylsulfatase D	ARSD	No	65		4.705E-05		
Asparagine synthetase [glutamine-hydrolyzing]	ASNS	Yes	64		7.406E-05		
ATP synthase subunit d, mitochondrial	ATP5H	No	18		5.113E-04		
ATP synthase subunit gamma, mitochondrial	ATP5C1	Yes	33		3.877E-04		
ATPase ASNA1	ASNA1	No	39		8.992E-05		
ATPase family AAA domain-containing protein 1	ATAD1	No	41		4.591E-05		
ATP-binding cassette sub-family A member 12	ABCA12	No	293		6.344E-05		
ATP-binding cassette sub-family A member 2	ABCA2	Yes	270		1.500E-05		
ATP-binding cassette sub-family B member 6, mitochondrial	ABCB6	No	94		2.663E-04		

ATP-binding cassette sub-family B member 7, mitochondrial	ABCB7	No	83		8.753E-05		
ATP-binding cassette sub-family G member 1	ABCG1	Yes	76		8.686E-05		
Baculoviral IAP repeat-containing protein 6	BIRC6	No	530		2.433E-05		
Battenin	CLN3	Yes	48		4.505E-05		
B-cell receptor-associated protein 29	BCAP29	No	28		6.556E-05		
Bcl-2-like protein 11	BCL2L11	No	22		1.250E-04		
Beta-galactosidase-1-like protein 2	GLB1L2	No	72		3.720E-05		
Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2	PAPSS2	No	70		6.001E-05		
Bifunctional purine biosynthesis protein PURH	ATIC	No	65		7.722E-05		
Bombesin receptor-activated protein C6orf89	C6orf89	No	40		1.382E-04		
Bromodomain and WD repeat-containing protein 1	BRWD1	No	263		3.346E-05		
Calcineurin B homologous protein 3	TESC	No	25		1.167E-04		
Calcium and integrin-binding protein 1	CIB1	Yes	22		4.254E-04		
Calcium signal-modulating cyclophilin ligand	CAMLG	Yes	33		1.867E-04		
Calmodulin-regulated spectrin-associated protein 3	CAMSAP3	No	135		1.938E-05		
Calreticulin	CALR	Yes	48		2.214E-04		
Caprin-1	CAPRIN1	No	78		2.337E-05		
Carbonic anhydrase 12	CA12	Yes	39		8.166E-05		
Carboxypeptidase D	CPD	No	153		4.921E-05		
Catenin delta-2	CTNND2	No	133		3.240E-05		
CCAAT/enhancer-binding protein zeta	CEBPZ	Yes	121		3.964E-05		
CCR4-NOT transcription complex subunit 11	CNOT11	No	55		3.144E-05		
Cell cycle control protein 50A	TMEM30A	No	41		1.211E-04		
Cell cycle control protein 50B	TMEM30B	No	39		9.509E-05		
Cell differentiation protein RCD1 homolog	RQCD1	No	34		6.599E-05		
Ceramide synthase 2	CERS2	No	45		4.749E-04		
Ceramide synthase 6	CERS6	No	45		1.034E-04		
Ceroid-lipofuscinosis neuronal protein 6	CLN6	No	36		1.549E-04		
Choline transporter-like protein 2	SLC44A2	No	80		1.006E-04		
Choline-phosphate cytidylyltransferase A	PCYT1A	No	42		1.682E-04		
Chromodomain-helicase-DNA-binding protein 3	CHD3	No	227		2.050E-05		
Cirhin	CIRH1A	No	77		2.416E-05		
Citrate synthase, mitochondrial	CS	No	52		4.400E-05		
Cleavage and polyadenylation specificity factor subunit 1	CPSF1	No	161		1.095E-05		
Cluster of [Pyruvate dehydrogenase (acetyl-transferring)] kinase isozyme 3, mitochondrial	PDK3	Yes	47		4.082E-05		
Cluster of Ankyrin repeat and KH domain-containing protein 1	ANKHD1	No	269		6.520E-06		
Cluster of ATP-dependent RNA helicase DDX39A	DDX39A	No	49		1.697E-04		
Cluster of ATP-dependent RNA helicase DDX54	DDX54	No	99		3.701E-05		
Cluster of Chromobox protein homolog 3	CBX3	Yes	21		3.635E-04		
Cluster of Creatine kinase U-type, mitochondrial	CKMT1A	No	47		6.747E-05		
Cluster of Ephrin type-B receptor 4	EPHB4	Yes	108		2.929E-05		
Cluster of Isoform 4 of Fragile X mental retardation protein 1	FMR1	Yes	68		1.137E-04		
Cluster of Lipopolysaccharide-responsive and beige-like anchor protein	LRBA	Yes	319		1.029E-05		
Cluster of Metal transporter CNNM4	CNNM4	Yes	87		2.138E-05		
Cluster of Nesprin-2	SYNE2	No	796		1.112E-05		
Cluster of Peptidyl-prolyl cis-trans isomerase FKBP8	FKBP8	No	45		2.155E-04		
Cluster of Serine/threonine-protein kinase MST4	MST4	No	47		3.984E-05		
Cluster of Ubiquilin-2	UBQLN2	No	66		3.286E-05		
Clustered mitochondria protein homolog	CLUH	No	147		2.250E-05		
Cohesin subunit SA-2	STAG2	No	141		2.010E-05		
Coiled-coil domain-containing protein 115	CCDC115	No	20		1.520E-04		
Coiled-coil domain-containing protein 51	CCDC51	No	46		6.977E-05		
Coiled-coil domain-containing protein 90B, mitochondrial	CCDC90B	No	30		1.596E-04		
Cold shock domain-containing protein E1	CSDE1	No	89		3.496E-05		
Collagen alpha-1(XVIII) chain	COL18A1	No	178		1.125E-05		
CTP synthase 2	CTPS2	No	66		3.591E-05		
Cullin-4B	CUL4B	No	104		3.141E-05		
Cyclin-dependent kinase 13	CDK13	No	165		1.601E-05		
Cytochrome c oxidase assembly protein COX15 homolog	COX15	No	46		1.290E-04		
Cytochrome c oxidase subunit 5B, mitochondrial	COX5B	No	14		3.095E-04		
Cytochrome P450 4F22	CYP4F22	No	62		6.388E-05		
Cytosol aminopeptidase	LAP3	No	56		3.847E-05		
D-3-phosphoglycerate dehydrogenase	PHGDH	Yes	57		1.664E-04		
DDB1- and CUL4-associated factor 7	DCAF7	No	39		5.769E-05		
Dedicator of cytokinesis protein 6	DOCK6	No	230		1.555E-05		
Delta(14)-sterol reductase	TM7SF2	No	46		2.021E-04		
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	ECH1	No	36		3.976E-04		
Dephospho-CoA kinase domain-containing protein	DCAKD	No	27		2.619E-04		
Diablo homolog, mitochondrial	DIABLO	No	27		3.754E-04		
Diacylglycerol kinase epsilon	DGKE	No	64		1.127E-04		
DNA mismatch repair protein Msh2	MSH2	Yes	105		3.574E-05		

DNA mismatch repair protein Msh6	MSH6	Yes	153		4.528E-05		
DNA replication licensing factor MCM3	MCM3	No	91		2.604E-05		
DNA replication licensing factor MCM6	MCM6	Yes	93		3.971E-05		
DnaJ homolog subfamily A member 3, mitochondrial	DNAJA3	No	52		1.277E-04		
DnaJ homolog subfamily C member 1	DNAJC1	No	64		2.489E-04		
DnaJ homolog subfamily C member 3	DNAJC3	No	58		4.068E-05		
Dolichol kinase	DOLK	No	59		3.667E-05		
Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase	ALG3	No	50		2.312E-04		
Double-stranded RNA-binding protein Staufen homolog 2	STAU2	No	63		9.912E-05		
E3 ubiquitin-protein ligase HECTD3	HECTD3	No	97		4.243E-05		
E3 ubiquitin-protein ligase HERC2	HERC2	No	527		3.653E-05		
E3 ubiquitin-protein ligase HUWE1	HUWE1	No	482		2.952E-05		
E3 ubiquitin-protein ligase listerin	LTN1	No	201		2.785E-05		
E3 ubiquitin-protein ligase synoviolin	SYVN1	No	68		6.646E-05		
E3 ubiquitin-protein ligase UBR5	UBR5	No	309		1.727E-05		
Electron transfer flavoprotein subunit alpha, mitochondrial	ETFA	No	35		3.143E-04		
Endoplasmic reticulum lectin 1	ERLEC1	No	55		8.282E-05		
Endoplasmic reticulum resident protein 29	ERP29	No	29		1.503E-04		
Endoplasmic reticulum-Golgi intermediate compartment protein 2	ERGIC2	No	43		1.061E-04		
Ensconsin	MAP7	No	84		9.334E-05		
Equilibrative nucleoside transporter 1	SLC29A1	No	50		8.874E-05		
Exonuclease 3'-5' domain-containing protein 2	EXD2	No	70		1.440E-04		
Fatty-acid amide hydrolase 1	FAAH	Yes	63		1.408E-04		
Flap endonuclease 1	FEN1	Yes	43		5.335E-05		
Flavin reductase (NADPH)	BLVRB	No	22		1.979E-04		
Focadhesin	FOCAD	No	200		1.549E-05		
FUN14 domain-containing protein 2	FUNDC2	No	21		3.675E-04		
Galactokinase	GALK1	Yes	42		1.235E-04		
Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3	B3GAT3	No	37		1.836E-04		
Ganglioside-induced differentiation-associated protein 1	GDAP1	No	41		2.020E-04		
GDNF family receptor alpha-1	GFRA1	Yes	51		1.929E-04		
General transcription factor 3C polypeptide 1	GTF3C1	No	239		1.970E-05		
General transcription factor 3C polypeptide 3	GTF3C3	No	101		3.619E-05		
General transcription factor II-I	GTF2I	Yes	112		2.503E-05		
Glucose-6-phosphate isomerase	GPI	Yes	63		8.952E-05		
Glucose-6-phosphate translocase	SLC37A4	No	46		1.459E-04		
Glycerol-3-phosphate dehydrogenase 1-like protein	GPD1L	No	38		1.289E-04		
Glycoprotein endo-alpha-1,2-mannosidase-like protein	MANEAL	No	51		1.567E-04		
Glycosaminoglycan xylosylkinase	FAM20B	No	46		1.104E-04		
Golgi integral membrane protein 4	GOLIM4	No	82		3.511E-05		
Golgi to ER traffic protein 4 homolog	GET4	No	37		2.124E-04		
GPI ethanolamine phosphate transferase 3	PIGO	No	119		1.613E-04		
GPI mannosyltransferase 1	PIGM	No	49		4.975E-05		
GTP-binding nuclear protein Ran	RAN	No	24		7.673E-05		
Guanine nucleotide-binding protein-like 3	GNL3	No	62		1.487E-04		
HEAT repeat-containing protein 6	HEATR6	No	129		1.865E-04		
Histone H1x	H1FX	No	22		7.715E-04		
Histone-lysine N-methyltransferase EHMT1	EHMT1	No	141		2.821E-05		
Huntingtin	HTT	No	348		2.318E-05		
ICOS ligand	ICOSLG	No	33		6.610E-05		
Immunoglobulin superfamily member 3	IGSF3	No	135		1.343E-05		
Inositol 1,4,5-trisphosphate receptor type 1	ITPR1	Yes	314		3.784E-05		
Insulin-like growth factor 1 receptor	IGF1R	Yes	155		2.289E-05		
Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	IDH3A	No	40		1.221E-04		
Isoform 2 of AP-1 complex subunit mu-2	AP1M2	Yes	48		8.692E-05		
Isoform 2 of Band 4.1-like protein 5	EPB41L5	No	58		1.600E-04		
Isoform 2 of Epimerase family protein SDR39U1	SDR39U1	No	31		2.153E-04		
Isoform 2 of NADH dehydrogenase [ubiquinone] flavoprotein 3, mitochondrial	NDUFV3	No	51		1.463E-04		
Isoform 2 of Phosphatidylinositol N-acetylglucosaminyltransferase subunit Q	PIGQ	No	65		2.110E-05		
Isoform 2 of Receptor-type tyrosine-protein phosphatase kappa	PTPRK	No	162		4.607E-05		
Isoform 2 of Ribonucleoprotein PTB-binding 1	RAVER1	No	78		3.294E-05		
Isoform 3 of Protein kinase C-binding protein 1	ZMYND8	No	110		2.463E-05		
Isoform 5 of Probable E3 ubiquitin-protein ligase HECTD4	HECTD4	No	145		8.024E-06		
Isoform Beta of E3 ubiquitin-protein ligase TRIM33	TRIM33	No	121		3.481E-05		
Keratinocyte-associated transmembrane protein 2	KCT2	No	29		1.390E-04		
Kynureninase	KYNU	Yes	52		5.255E-05		
L-2-hydroxyglutarate dehydrogenase, mitochondrial	L2HGDH	No	50		9.606E-05		
La-related protein 1	LARP1	No	124		8.496E-05		

LETM1 domain-containing protein 1	LETMD1	No	42		8.907E-05		
Liprin-alpha-1	PPFIA1	Yes	136		2.405E-05		
Long-chain fatty acid transport protein 3	SLC27A3	No	79		1.042E-04		
Low affinity cationic amino acid transporter 2	SLC7A2	No	72		6.805E-04		
Lysine-specific demethylase 5B	KDM5B	Yes	176		1.313E-05		
Lysosomal acid phosphatase	ACP2	No	48		5.777E-05		
Macoilin	TMEM57	No	76		5.502E-05		
Maestro heat-like repeat-containing protein family member 1	MROH1	No	181		1.010E-05		
Major centromere autoantigen B	CENPB	No	65		4.658E-05		
Major facilitator superfamily domain-containing protein 5	MFSD5	No	50		5.498E-05		
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	MAN1A1	No	73		3.742E-05		
Membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase	PKMYT1	Yes	55		6.271E-05		
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	No	61		4.395E-05		
Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	MCCC1	Yes	80		3.338E-05		
Methyltransferase-like protein 7B	METTL7B	No	28		1.832E-04		
MIP18 family protein FAM96A	FAM96A	No	18		1.513E-04		
Mitochondrial calcium uniporter regulator 1	MCUR1	No	40		7.837E-05		
Mitochondrial coenzyme A transporter SLC25A42	SLC25A42	No	35		7.538E-05		
Mitochondrial fission process protein 1	MTFP1	No	18		2.438E-04		
Mitochondrial import inner membrane translocase subunit TIM44	TIMM44	No	51		1.780E-04		
Mitochondrial inner membrane protein OXA1L	OXA1L	No	49		6.343E-05		
MMS19 nucleotide excision repair protein homolog	MMS19	No	113		1.991E-05		
Mono [ADP-ribose] polymerase PARP16	PARP16	No	36		1.757E-04		
N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	B3GNT1	No	47		4.885E-05		
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	NDUFA11	No	15		2.799E-04		
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial	NDUFB5	No	22		6.370E-04		
NADH dehydrogenase [ubiquinone] 1 subunit C2	NDUFC2	No	14		6.471E-04		
NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	NDUFV2	No	27		1.458E-04		
NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	NDUFS2	No	53		1.397E-04		
NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	NDUFS5	No	13		2.654E-04		
NADH-ubiquinone oxidoreductase chain 5	MT-ND5	No	67		1.414E-04		
NEDD4-binding protein 3	N4BP3	No	60		4.449E-05		
Negative elongation factor B	NELFB	No	66		2.857E-05		
Negative elongation factor C/D	NELFCD	No	66		3.475E-05		
Neprilysin	MME	Yes	86		1.198E-04		
Neural cell adhesion molecule 2	NCAM2	No	93		1.303E-04		
Neurosecretory protein VGF	VGF	No	67		6.067E-05		
Neutral amino acid transporter A	SLC1A4	No	56		2.270E-04		
Nuclear pore complex protein Nup214	NUP214	No	214		1.951E-05		
Nucleolar and coiled-body phosphoprotein 1	NOLC1	Yes	74		4.135E-05		
Nucleolar protein 16	NOP16	No	21		1.801E-04		
Nucleolar transcription factor 1	UBTF	No	89		2.754E-05		
Nucleosome-remodeling factor subunit BPTF	BPTF	No	338		6.909E-06		
Obg-like ATPase 1	OLA1	No	45		7.377E-05		
Oxysterol-binding protein 1	OSBP	Yes	89		2.054E-05		
P2X purinoceptor 4	P2RX4	Yes	43		9.496E-05		
Paired amphipathic helix protein Sin3a	SIN3A	No	145		2.004E-05		
Paraplegin	SPG7	Yes	88		2.550E-05		
PDZ domain-containing protein GIPC1	GIPC1	No	36		6.157E-05		
Pecanex-like protein 3	PCNXL3	No	222		2.005E-05		
Pentatricopeptide repeat domain-containing protein 3, mitochondrial	PTCD3	No	79		2.976E-05		
Peptide deformylase, mitochondrial	PDF	No	27		1.332E-04		
Peptidyl-tRNA hydrolase ICT1, mitochondrial	ICT1	No	24		1.239E-04		
Peripheral plasma membrane protein CASK	CASK	No	105		1.731E-05		
Peroxiredoxin-4	PRDX4	Yes	31		8.845E-05		
Peroxisomal 2,4-dienoyl-CoA reductase	DECR2	No	31		5.491E-05		
Peroxisomal acyl-coenzyme A oxidase 3	ACOX3	Yes	78		1.710E-04		
Peroxisomal multifunctional enzyme type 2	HSD17B4	No	80		2.813E-04		
Pescadillo homolog	PES1	Yes	68		8.232E-05		
Phosphatidylserine decarboxylase proenzyme	PISD	Yes	47		6.954E-05		
Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	PCK2	No	71		1.285E-04		
Plasminogen receptor (KT)	PLGRKT	No	17		2.485E-04		
Pogo transposable element with ZNF domain	POGZ	No	155		2.744E-05		
Pre-B-cell leukemia transcription factor-interacting protein 1	PBXIP1	No	81		7.265E-05		
Pre-mRNA-splicing factor SPF27	BCAS2	Yes	26		9.009E-05		
Presqualene diphosphate phosphatase	PPAPDC2	No	32		9.799E-05		
Probable ATP-dependent RNA helicase DDX31	DDX31	No	94		2.844E-05		
Probable ATP-dependent RNA helicase DHX36	DHX36	No	115		1.591E-05		

Probable E3 ubiquitin-protein ligase HERC1	HERC1	No	532		1.662E-05		
Probable helicase with zinc finger domain	HELZ	No	219		1.869E-05		
Probable palmitoyltransferase ZDHHC20	ZDHHC20	No	42		4.540E-05		
Probable phospholipid-transporting ATPase IC	ATP8B1	No	144		5.564E-05		
Probable phospholipid-transporting ATPase IG	ATP11C	No	129		4.692E-05		
Programmed cell death protein 6	PDCD6	No	22		3.140E-04		
Proline dehydrogenase 1, mitochondrial	PRODH	No	68		2.672E-05		
Proline-, glutamic acid- and leucine-rich protein 1	PELP1	No	120		1.029E-04		
Prostaglandin E synthase 2	PTGES2	No	42		3.108E-04		
Proteasome activator complex subunit 2	PSME2	Yes	27		1.387E-04		
Protein arginine N-methyltransferase 1	PRMT1	No	42		6.769E-05		
Protein canopy homolog 2	CNPY2	No	21		8.809E-05		
Protein CASP	CUX1	No	77		4.809E-05		
Protein FAM134A	FAM134A	No	58		3.733E-05		
Protein FAM162A	FAM162A	No	17		2.868E-04		
Protein FAM210A	FAM210A	No	31		1.683E-04		
Protein FAM84B	FAM84B	No	34		1.417E-04		
Protein kinase C delta type	PRKCD	Yes	78		4.903E-05		
Protein LSM14 homolog B	LSM14B	No	42		7.307E-05		
Protein MON2 homolog	MON2	No	190		2.158E-05		
Protein O-mannosyl-transferase 1	POMT1	No	85		3.312E-05		
Protein phosphatase 1G	PPM1G	Yes	59		4.433E-05		
Protein polybromo-1	PBRM1	No	193		2.090E-05		
Protein RFT1 homolog	RFT1	No	60		3.404E-04		
Protein SCO1 homolog, mitochondrial	SCO1	No	34		1.624E-04		
Protein transport protein Sec31A	SEC31A	No	133		1.636E-05		
Protein tyrosine phosphatase type IVA 1	PTP4A1	No	20		4.759E-04		
Protein Wiz	WIZ	No	179		9.711E-06		
Protein YIPF3	YIPF3	No	38		8.038E-05		
Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	No	25		1.460E-04		
Putative ATP-dependent RNA helicase DHX57	DHX57	No	156		2.052E-05		
Putative oxidoreductase GLYR1	GLYR1	No	61		1.323E-04		
Rab-like protein 3	RABL3	No	26		1.493E-04		
Ran GTPase-activating protein 1	RANGAP1	No	64		5.515E-05		
Ras GTPase-activating protein-binding protein 2	G3BP2	No	54		1.414E-04		
Ras-related protein Rab-17	RAB17	No	23		9.384E-04		
Ras-related protein Rab-25	RAB25	No	23		3.861E-04		
Receptor expression-enhancing protein 6	REEP6	No	21		3.531E-04		
Regulator of microtubule dynamics protein 3	RMDN3	No	52		8.561E-05		
Remodeling and spacing factor 1	RSF1	No	164		5.606E-05		
Required for meiotic nuclear division protein 1 homolog	RMND1	No	52		3.662E-04		
Reticulon-2	RTN2	No	59		7.623E-05		
Rhomboid domain-containing protein 2	RHBDD2	No	39		1.323E-04		
Ribosomal RNA processing protein 1 homolog B	RRP1B	No	84		3.640E-05		
Ribosome production factor 2 homolog	RPF2	No	36		1.421E-04		
RNA-binding protein Musashi homolog 2	MSI2	No	35		4.411E-04		
SAFB-like transcription modulator	SLTM	No	117		3.228E-05		
Sec1 family domain-containing protein 2	SCFD2	No	75		5.353E-05		
Serine beta-lactamase-like protein LACTB, mitochondrial	LACTB	No	61		1.024E-04		
Serine/arginine-rich splicing factor 10	SRSF10	No	31		2.048E-04		
Serine/arginine-rich splicing factor 5	SRSF5	No	31		1.026E-04		
Serine/arginine-rich splicing factor 9	SRSF9	No	26		1.081E-04		
Serine/threonine-protein kinase 11-interacting protein	STK11IP	No	121		1.915E-05		
Serine/threonine-protein kinase SMG1	SMG1	No	410		1.078E-05		
Sickle tail protein homolog	KIAA1217	Yes	214		8.251E-06		
SID1 transmembrane family member 1	SIDT1	No	94		7.028E-05		
Sister chromatid cohesion protein PDS5 homolog B	PDS5B	No	165		5.817E-05		
Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)	SLC6A14	Yes	72		1.588E-04		
Sodium/hydrogen exchanger 6	SLC9A6	No	74		3.065E-05		
Sodium-dependent multivitamin transporter	SLC5A6	Yes	69		1.387E-04		
Sodium-driven chloride bicarbonate exchanger	SLC4A10	No	126		4.281E-05		
Solute carrier family 12 member 2	SLC12A2	No	131		1.135E-04		
Solute carrier family 35 member B1	SLC35B1	Yes	36		2.753E-04		
Sorting nexin-2	SNX2	No	58		3.089E-05		
Structural maintenance of chromosomes protein 1A	SMC1A	No	143		3.326E-05		
Succinate dehydrogenase [ubiquinone] cytochrome b small subunit, mitochondrial	SDHD	No	17		1.522E-04		
Surfeit locus protein 1	SURF1	Yes	33		1.341E-04		
Surfeit locus protein 6	SURF6	No	41		6.490E-05		
Synapse-associated protein 1	SYAP1	No	40		9.263E-05		
Synaptosomal-associated protein 29	SNAP29	No	29		4.269E-04		
Syntaxin-17	STX17	No	33		1.105E-04		

Syntaxin-18	STX18	No	39		1.917E-04		
Syntaxin-3	STX3	No	33		1.110E-04		
Syntaxin-8	STX8	No	27		1.016E-04		
Testis-expressed sequence 10 protein	TEX10	No	106		7.501E-05		
Thioredoxin-related transmembrane protein 4	TMX4	No	39		1.213E-04		
TLC domain-containing protein 1	TLCD1	No	29		1.960E-04		
TOM1-like protein 2	TOM1L2	No	56		4.044E-05		
Torsin-1A	TOR1A	No	38		1.505E-04		
Trans-acting T-cell-specific transcription factor GATA-3	GATA3	Yes	48		6.473E-05		
Transcription intermediary factor 1-beta	TRIM28	No	89		5.354E-05		
Transcriptional regulator ATRX	ATRX	No	283		1.510E-05		
Transformer-2 protein homolog alpha	TRA2A	No	33		9.976E-05		
Transmembrane emp24 domain-containing protein 1	TMED1	No	25		4.010E-04		
Transmembrane protein 128	TMEM128	No	19		1.995E-04		
Transmembrane protein 199	TMEM199	No	23		1.745E-04		
Transmembrane protein 87A	TMEM87A	No	63		5.306E-05		
Transmembrane protein 97	TMEM97	No	21		2.737E-04		
Tumor protein D54	TPD52L2	No	22		1.175E-04		
Tumor protein p53-inducible protein 11	TP53I11	No	21		3.829E-04		
Tumor suppressor candidate 3	TUSC3	No	40		7.996E-05		
Ubiquinol-cytochrome c reductase complex chaperone CBP3 homolog	UQCC1	No	35		5.284E-05		
Ubiquitin carboxyl-terminal hydrolase 32	USP32	No	182		1.985E-05		
UDP-glucose 6-dehydrogenase	UGDH	Yes	55		8.098E-05		
UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	DPAGT1	Yes	46		6.820E-05		
Uncharacterized aarF domain-containing protein kinase 1	ADCK1	No	61		6.894E-05		
Uncharacterized protein C17orf62	C17orf62	No	21		1.307E-04		
Uncharacterized protein C1orf21	C1orf21	No	14		2.242E-04		
Uncharacterized protein C4orf32	C4orf32	No	15		5.261E-04		
Uncharacterized protein C6orf47	C6orf47	No	32		9.674E-05		
Uncharacterized protein C7orf50	C7orf50	No	22		2.265E-04		
Uncharacterized protein KIAA0195	KIAA0195	Yes	151		1.222E-05		
UPF0577 protein KIAA1324	KIAA1324	Yes	111		1.075E-04		
UPF0609 protein C4orf27	C4orf27	No	39		5.926E-05		
Vacuolar protein sorting-associated protein 13A	VPS13A	No	360		1.192E-05		
Vacuolar protein sorting-associated protein 13C	VPS13C	No	422		7.990E-06		
Vacuole membrane protein 1	VMP1	No	46		8.354E-05		
Vang-like protein 1	VANGL1	No	60		3.913E-05		
Very long-chain acyl-CoA synthetase	SLC27A2	Yes	70		1.129E-04		
Very-long-chain (3R)-3-hydroxyacyl-[acyl-carrier protein] dehydratase 2	PTPLB	No	28		1.438E-04		
Vesicle transport protein SFT2C	SFT2D3	No	22		1.126E-04		
Vitamin K epoxide reductase complex subunit 1	VKORC1	No	18		3.485E-04		
WD repeat-containing protein 36	WDR36	No	105		5.147E-05		
WD repeat-containing protein 43	WDR43	No	75		2.880E-05		
Wiskott-Aldrich syndrome protein family member 2	WASF2	No	54		5.603E-05		
Y-box-binding protein 2	YBX2	No	39		4.120E-04		
Zinc finger CCCH domain-containing protein 18	ZC3H18	No	106		3.834E-05		
Zinc finger ZZ-type and EF-hand domain-containing protein 1	ZZEF1	No	331		3.012E-05		
Zinc transporter ZIP6	SLC39A6	Yes	85		1.769E-04		
Identified membrane proteins present in SKBR3 only							
		BC	**MW**	**Average NSAF**			
Identified Proteins	**Gene**	**database**	**(kDa)**	**HMEC**	**MCF7**	**SKBR3**	**MDA231**
14-3-3 protein sigma	SFN	Yes	28			6.866E-04	
182 kDa tankyrase-1-binding protein	TNKS1BP1	No	182			3.507E-05	
2,4-dienoyl-CoA reductase, mitochondrial	DECR1	No	36			8.951E-04	
28S ribosomal protein S11, mitochondrial	MRPS11	No	21			8.643E-04	
39S ribosomal protein L48, mitochondrial	MRPL48	No	24			4.039E-04	
5-azacytidine-induced protein 1	AZI1	No	122			5.510E-05	
Adseverin	SCIN	No	80			3.472E-04	
A-kinase anchor protein 13	AKAP13	No	308			2.115E-05	
Alkaline phosphatase, placental type	ALPP	No	58			1.994E-04	
Alpha-2-HS-glycoprotein	AHSG	No	39			1.850E-03	
Annexin A7	ANXA7	No	53			1.485E-03	
Apolipoprotein D	APOD	Yes	21			2.496E-04	
ATP-dependent RNA helicase DDX1	DDX1	No	82			8.195E-05	
ATP-dependent RNA helicase DDX24	DDX24	No	96			9.796E-05	
Beta-galactoside alpha-2,6-sialyltransferase 1	ST6GAL1	Yes	47			6.450E-04	
Beta-secretase 2	BACE2	Yes	56			2.060E-04	
Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1	NDST1	No	101			1.787E-04	
Breast carcinoma-amplified sequence 1	BCAS1	Yes	62			4.059E-04	
Cell growth regulator with EF hand domain protein 1	CGREF1	No	32			4.363E-04	

Chaperone activity of bc1 complex-like, mitochondrial	ADCK3	No	72			1.980E-04	
Chloride channel protein 2	CLCN2	No	99			5.400E-05	
Cingulin	CGN	No	136			1.402E-04	
Clathrin light chain B	CLTB	No	25			2.118E-04	
Claudin-7	CLDN7	No	22			6.410E-04	
Cluster of Golgin subfamily A member 2	GOLGA2	No	113			2.008E-04	
Cluster of Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	No	38			9.560E-04	
Cluster of Isoform 3 of Leucine-rich repeat flightless-interacting protein 1	LRRFIP1	No	83			2.264E-04	
Cluster of KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	No	48			5.945E-04	
Cluster of Putative RNA-binding protein Luc7-like 2	LUC7L2	No	47			1.803E-04	
Cluster of Tetraspanin-13	TSPAN13	No	22			1.054E-03	
CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4	ST3GAL4	No	38			3.993E-04	
Collagen alpha-1(I) chain	COL1A1	Yes	139			3.223E-05	
Dehydrogenase/reductase SDR family member 2, mitochondrial	DHRS2	Yes	30			3.293E-02	
Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	ALDH4A1	Yes	62			1.279E-04	
Disintegrin and metalloproteinase domain-containing protein 15	ADAM15	Yes	93			9.902E-05	
EF-hand domain-containing protein D1	EFHD1	No	27			1.815E-03	
ELM2 and SANT domain-containing protein 1	ELMSAN1	No	115			5.819E-05	
Endophilin-B1	SH3GLB1	No	41			1.324E-04	
Ephrin type-B receptor 3	EPHB3	Yes	110			4.859E-05	
Eukaryotic translation initiation factor 5B	EIF5B	No	139			6.911E-05	
Exosome complex component MTR3	EXOSC6	No	28			3.052E-04	
Fatty acyl-CoA reductase 2	FAR2	No	59			3.473E-04	
F-box/LRR-repeat protein 20	FBXL20	No	48			2.469E-04	
Glucosylceramidase	GBA	No	60			1.806E-04	
Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1	C1GALT1	No	42			8.786E-04	
Golgi membrane protein 1	GOLM1	No	45			2.135E-04	
Golgi reassembly-stacking protein 2	GORASP2	No	47			2.386E-04	
Growth factor receptor-bound protein 7	GRB7	Yes	60			2.030E-04	
H/ACA ribonucleoprotein complex subunit 3	NOP10	No	8			1.143E-03	
Heterogeneous nuclear ribonucleoprotein L-like	HNRNPLL	No	60			2.014E-04	
Hexaprenyldihydroxybenzoate methyltransferase, mitochondrial	COQ3	No	41			3.164E-04	
Histone deacetylase complex subunit SAP18	SAP18	No	18			1.275E-03	
Immunoglobulin superfamily member 8	IGSF8	No	65			1.360E-04	
Involucrin	IVL	No	68			2.650E-04	
Isoform 2 of Melanophilin	MLPH	No	63			1.825E-04	
Isoform 2 of RUN and SH3 domain-containing protein 1	RUSC1	No	47			9.348E-05	
Isoform 3 of Natural resistance-associated macrophage protein 2	SLC11A2	Yes	65			8.307E-05	
Isoform 6 of Agrin	AGRN	No	215			2.110E-04	
Keratin, type II cytoskeletal 4	KRT4	No	57			3.371E-03	
Killer cell lectin-like receptor subfamily G member 2	KLRG2	No	43			2.339E-04	
Kinesin-like protein KIFC1	KIFC1	Yes	74			1.070E-04	
Leucine-rich repeat-containing protein 47	LRRC47	No	63			1.040E-04	
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	No	53			1.517E-04	
LMBR1 domain-containing protein 2	LMBRD2	No	81			2.413E-04	
Major facilitator superfamily domain-containing protein 6	MFSD6	No	88			2.581E-04	
Major facilitator superfamily domain-containing protein 9	MFSD9	No	51			1.019E-04	
Mitochondrial fission 1 protein	FIS1	No	17			1.437E-03	
Monocarboxylate transporter 8	SLC16A2	Yes	60			2.419E-04	
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4	NDUFA4	Yes	9			1.488E-03	
Non-histone chromosomal protein HMG-14	HMGN1	Yes	11			3.905E-03	
Nuclear receptor coactivator 5	NCOA5	No	66			1.047E-04	
Nuclear valosin-containing protein-like	NVL	No	95			1.828E-04	
ORM1-like protein 3	ORMDL3	No	17			2.321E-03	
Partner of Y14 and mago	WIBG	No	23			6.549E-04	
Periplakin	PPL	No	205			8.349E-05	
Peroxisome assembly factor 2	PEX6	No	104			1.347E-04	
Polymerase delta-interacting protein 3	POLDIP3	No	46			1.745E-04	
Polypeptide N-acetylgalactosaminyltransferase 10	GALNT10	Yes	69			1.194E-04	
Polyprenol reductase	SRD5A3	No	37			3.391E-04	
Polypyrimidine tract-binding protein 1	PTBP1	No	57			3.829E-04	
Pre-mRNA-splicing factor ISY1 homolog	ISY1	No	33			2.998E-04	
Presenilin-2	PSEN2	No	50			2.664E-04	
Probable ATP-dependent RNA helicase DDX28	DDX28	No	60			1.358E-04	
Probable phospholipid-transporting ATPase IA	ATP8A1	No	131			1.440E-04	
Protein BUD31 homolog	BUD31	No	17			1.016E-03	
Protein CASC4	CASC4	No	49			1.925E-04	
Protein HID1	HID1	No	89			9.158E-05	

Protein NDRG1	NDRG1	Yes	43			4.881E-04	
Protein NipSnap homolog 2	GBAS	No	34			3.822E-04	
Protein Red	IK	No	66			4.294E-04	
Proton-coupled folate transporter	SLC46A1	No	50			4.169E-04	
Putative RNA-binding protein 15	RBM15	No	107			1.615E-04	
Pyrroline-5-carboxylate reductase 1, mitochondrial	PYCR1	No	33			2.638E-04	
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	No	54			1.657E-04	
Ras-related GTP-binding protein C	RRAGC	No	44			2.426E-04	
Receptor tyrosine-protein kinase erbB-2	ERBB2	Yes	138			1.029E-03	
Regulator of chromosome condensation	RCC1	Yes	45			1.152E-04	
Replication initiator 1	REPIN1	No	64			2.119E-04	
Rho-related GTP-binding protein RhoB	RHOB	No	22			9.192E-04	
RNA-binding protein PNO1	PNO1	No	28			1.917E-04	
Scaffold attachment factor B1	SAFB	No	103			2.857E-04	
Signal recognition particle 14 kDa protein	SRP14	Yes	15			2.866E-03	
Solute carrier family 35 member F6	SLC35F6	No	40			1.627E-03	
Spliceosome-associated protein CWC15 homolog	CWC15	No	27			2.110E-04	
Splicing factor 3A subunit 1	SF3A1	No	89			6.093E-05	
Src substrate cortactin	CTTN	Yes	62			3.201E-04	
SWI/SNF complex subunit SMARCC2	SMARCC2	Yes	133			7.039E-05	
Target of Myb protein 1	TOM1	No	54			9.856E-05	
Testin	TES	Yes	48			4.865E-04	
Tetraspanin-14	TSPAN14	No	31			2.210E-04	
TGF-beta receptor type-1	TGFBR1	No	56			2.192E-04	
THO complex subunit 1	THOC1	No	76			2.342E-04	
THO complex subunit 2	THOC2	No	183			7.532E-05	
THO complex subunit 5 homolog	THOC5	No	79			2.106E-04	
THO complex subunit 6	THOC6	No	38			2.872E-04	
Tight junction protein ZO-3	TJP3	Yes	103			1.027E-04	
Transcriptional activator protein Pur-beta	PURB	No	33			3.461E-04	
Transmembrane protein 209	TMEM209	No	63			2.124E-04	
Transmembrane protein 63A	TMEM63A	No	92			2.383E-04	
Tripartite motif-containing protein 3	TRIM3	No	81			3.018E-04	
TRMT1-like protein	TRMT1L	No	82			1.502E-04	
U2 small nuclear ribonucleoprotein A'	SNRPA1	Yes	28			2.861E-04	
Ubiquinone biosynthesis monooxygenase COQ6	COQ6	No	51			1.830E-04	
Uncharacterized protein C20orf24	C20orf24	Yes	15			1.303E-03	
Uncharacterized protein C4orf3	C4orf3	No	8			1.430E-03	
UPF0378 protein KIAA0100	KIAA0100	Yes	254			4.273E-05	
Vasorin	VASN	No	72			9.061E-05	
V-set domain-containing T-cell activation inhibitor 1	VTCN1	No	31			3.916E-04	
V-type proton ATPase 116 kDa subunit a isoform 4	ATP6V0A4	No	96			1.583E-04	
WD repeat-containing protein 18	WDR18	No	47			2.755E-04	
WD40 repeat-containing protein SMU1	SMU1	No	58			2.826E-04	
YTH domain-containing protein 1	YTHDC1	No	85			8.185E-05	
Zinc finger CCCH domain-containing protein 11A	ZC3H11A	No	89			1.191E-04	
Zinc finger CCCH domain-containing protein 14	ZC3H14	No	83			1.146E-04	
Zinc finger protein RFP	TRIM27	No	58			2.323E-04	

Identified membrane proteins present in MDA231 only

Identified Proteins	Gene	BC database	MW (kDa)	Average NSAF			
				HMEC	MCF7	SKBR3	MDA231
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-3	PLCD3	No	89				3.125E-05
2',5'-phosphodiesterase 12	PDE12	No	67				3.954E-05
3'(2'),5'-bisphosphate nucleotidase 1	BPNT1	No	33				7.818E-05
39S ribosomal protein L47, mitochondrial	MRPL47	No	29				1.578E-04
60S ribosomal protein L29	RPL29	No	18				1.037E-03
6-phosphogluconate dehydrogenase, decarboxylating	PGD	No	53				6.092E-05
Actin-related protein 2/3 complex subunit 4	ARPC4	No	20				2.040E-04
ADP-ribosylation factor-like protein 6-interacting protein 6	ARL6IP6	No	25				2.367E-04
ADP-ribosylation factor-related protein 1	ARFRP1	No	23				1.456E-04
Alpha-parvin	PARVA	No	42				5.325E-05
Aminopeptidase B	RNPEP	Yes	73				3.793E-05
Ankyrin repeat and FYVE domain-containing protein 1	ANKFY1	No	128				6.356E-05
Annexin A5	ANXA5	No	36				2.609E-04
Anoctamin-6	ANO6	No	106				9.737E-05
Aurora kinase B	AURKB	Yes	39				2.143E-04
BAG family molecular chaperone regulator 2	BAG2	No	24				5.028E-04
Band 4.1-like protein 2	EPB41L2	No	113				2.486E-05
Beta-1-syntrophin	SNTB1	No	58				7.226E-05
Beta-2-microglobulin	B2M	No	14				6.816E-04
Bifunctional coenzyme A synthase	COASY	Yes	62				3.483E-05

Bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66	NO66	No	71				8.321E-05
Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase	GNE	No	79				3.438E-05
Butyrophilin subfamily 2 member A1	BTN2A1	No	60				6.394E-05
Calpain-2 catalytic subunit	CAPN2	No	80				4.262E-05
Calpain-5	CAPN5	No	73				4.597E-05
Calponin-3	CNN3	Yes	36				7.319E-05
Cellular tumor antigen p53	TP53	Yes	44				9.703E-05
Centromere protein V	CENPV	No	30				3.380E-04
Chondroitin sulfate proteoglycan 4	CSPG4	Yes	251				1.515E-05
Claudin domain-containing protein 1	CLDND1	No	29				9.876E-05
Cluster of Formin-like protein 3	FMNL3	No	117				1.750E-04
Cluster of Heat shock protein 105 kDa	HSPH1	Yes	97				1.732E-04
Cluster of Importin subunit alpha-4	KPNA3	No	58				7.683E-05
Cluster of Misshapen-like kinase 1	MINK1	Yes	150				3.270E-05
Cluster of Rho-associated protein kinase 2	ROCK2	No	161				1.374E-05
Cluster of Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA	No	36				1.277E-04
Cluster of Syntenin-1	SDCBP	No	32				3.119E-04
Collagen alpha-1(XIII) chain	COL13A1	No	70				1.100E-04
Cullin-2	CUL2	Yes	87				4.423E-05
Cyclin-dependent kinase 1	CDK1	No	34				1.472E-04
Cysteine--tRNA ligase, cytoplasmic	CARS	Yes	85				4.483E-05
Cytoplasmic aconitate hydratase	ACO1	No	98				2.210E-05
Cytoskeleton-associated protein 2	CKAP2	Yes	77				2.876E-05
Cytospin-B	SPECC1	No	119				4.163E-05
DCN1-like protein 3	DCUN1D3	No	34				9.435E-05
Dedicator of cytokinesis protein 4	DOCK4	No	225				1.254E-05
Dedicator of cytokinesis protein 9	DOCK9	No	236				1.943E-05
Delta-1-pyrroline-5-carboxylate synthase	ALDH18A1	No	87				2.398E-05
Deoxynucleotidyltransferase terminal-interacting protein 2	DNTTIP2	No	84				2.522E-05
DNA (cytosine-5)-methyltransferase 1	DNMT1	Yes	183				1.216E-05
DNA replication licensing factor MCM5	MCM5	Yes	82				3.382E-05
DnaJ homolog subfamily B member 6	DNAJB6	No	36				1.931E-04
Dolichyldiphosphatase 1	DOLPP1	No	27				1.229E-04
E3 ubiquitin-protein ligase Itchy homolog	ITCH	No	103				4.433E-05
E3 ubiquitin-protein ligase NEDD4-like	NEDD4L	Yes	112				1.012E-04
Early endosome antigen 1	EEA1	No	162				2.074E-05
EGF-like repeat and discoidin I-like domain-containing protein 3	EDIL3	No	54				5.975E-05
Endoglin	ENG	No	71				4.422E-05
Endothelial cell-selective adhesion molecule	ESAM	No	41				1.648E-04
Ephrin type-B receptor 2	EPHB2	No	117				7.829E-05
Estradiol 17-beta-dehydrogenase 11	HSD17B11	No	33				4.759E-04
Eukaryotic translation initiation factor 3 subunit C-like protein	EIF3CL	No	105				4.883E-05
Eukaryotic translation initiation factor 3 subunit K	EIF3K	No	25				1.580E-04
Eukaryotic translation initiation factor 3 subunit L	EIF3L	No	67				8.933E-05
Eukaryotic translation initiation factor 4 gamma 2	EIF4G2	No	102				3.651E-05
FAD-dependent oxidoreductase domain-containing protein 1	FOXRED1	No	54				4.076E-05
Far upstream element-binding protein 2	KHSRP	No	73				3.491E-05
Farnesyl pyrophosphate synthase	FDPS	Yes	48				6.983E-05
Fatty acid desaturase 2	FADS2	Yes	52				2.986E-04
FERM, RhoGEF and pleckstrin domain-containing protein 1	FARP1	Yes	119				7.991E-05
Fermitin family homolog 3	FERMT3	No	76				2.970E-05
Ferritin heavy chain	FTH1	No	21				1.042E-04
Fibronectin type III domain-containing protein 3B	FNDC3B	No	133				1.120E-04
Fibronectin type-III domain-containing protein 3A	FNDC3A	No	132				1.037E-04
Filamin-C	FLNC	No	291				2.372E-04
Formin-like protein 1	FMNL1	Yes	122				5.963E-05
Fukutin-related protein	FKRP	No	55				1.185E-04
Gamma-tubulin complex component 3	TUBGCP3	Yes	104				2.719E-05
General vesicular transport factor p115	USO1	No	108				5.160E-05
Glutathione S-transferase omega-1	GSTO1	No	28				7.911E-05
Golgi-associated plant pathogenesis-related protein 1	GLIPR2	No	17				4.424E-04
Growth/differentiation factor 15	GDF15	Yes	34				1.895E-04
GTP-binding protein Rheb	RHEB	Yes	20				4.078E-04
Guanine nucleotide-binding protein G(o) subunit alpha	GNAO1	No	40				1.121E-03
Heme-binding protein 1	HEBP1	Yes	21				1.048E-04
Ig alpha-1 chain C region	IGHA1	No	38				9.755E-05
Inhibitor of nuclear factor kappa-B kinase-interacting protein	IKBIP	No	39				3.750E-04
Inner centromere protein	INCENP	No	105				3.733E-05
Integrator complex subunit 6	INTS6	No	100				2.798E-05
Integrin alpha-1	ITGA1	No	131				2.105E-05

Integrin beta-3	ITGB3	No	87				6.227E-05
Integrin-linked protein kinase	ILK	No	51				6.673E-05
Intercellular adhesion molecule 1	ICAM1	Yes	58				1.292E-04
Isoform 2 of Galectin-8	LGALS8	No	40				9.282E-05
Isoform 2 of Mitochondrial fission factor	MFF	No	33				2.747E-04
Kin of IRRE-like protein 1	KIRREL	No	84				6.460E-05
Kinesin-like protein KIF14	KIF14	No	186				2.743E-05
Lamin-B receptor	LBR	Yes	71				6.264E-04
Laminin subunit gamma-1	LAMC1	No	178				2.370E-05
LanC-like protein 1	LANCL1	Yes	45				6.076E-05
Leucine-rich repeat-containing protein 8A	LRRC8A	No	94				7.772E-05
LIM and SH3 domain protein 1	LASP1	Yes	30				1.489E-04
Limb region 1 protein homolog	LMBR1	No	55				5.820E-05
L-lactate dehydrogenase B chain	LDHB	Yes	37				6.725E-04
Low-density lipoprotein receptor-related protein 10	LRP10	Yes	76				2.755E-05
Lysophosphatidic acid receptor 1	LPAR1	No	41				7.834E-05
Macrophage-capping protein	CAPG	Yes	38				3.318E-04
Metastasis-associated protein MTA2	MTA2	No	75				9.977E-05
Mitochondrial import receptor subunit TOM34	TOMM34	No	35				8.086E-05
Monocarboxylate transporter 2	SLC16A7	No	52				1.317E-04
Multidrug resistance-associated protein 4	ABCC4	No	150				7.103E-05
Nesprin-3	SYNE3	No	112				9.754E-05
Nestin	NES	No	177				1.534E-04
Neural cell adhesion molecule L1	L1CAM	No	140				3.785E-05
Neuropilin-1	NRP1	No	103				2.882E-04
Nucleolar MIF4G domain-containing protein 1	NOM1	No	96				1.089E-04
Oxysterol-binding protein-related protein 10	OSBPL10	No	84				2.496E-05
Partitioning defective 3 homolog	PARD3	Yes	151				1.449E-05
PDZ and LIM domain protein 1	PDLIM1	Yes	36				7.369E-05
Peptidyl-prolyl cis-trans isomerase FKBP1A	FKBP1A	No	12				2.709E-04
Peroxiredoxin-6	PRDX6	No	25				1.280E-04
Phosphatidylinositol-binding clathrin assembly protein	PICALM	Yes	71				7.525E-05
Phosphoglycerate mutase 1	PGAM1	No	29				2.327E-04
Phosphoprotein associated with glycosphingolipid-enriched microdomains 1	PAG1	No	47				5.708E-05
Plakophilin-4	PKP4	No	132				5.020E-05
Polymerase I and transcript release factor	PTRF	No	43				1.732E-04
Prefoldin subunit 2	PFDN2	Yes	17				1.564E-04
Probable glutathione peroxidase 8	GPX8	No	24				1.152E-04
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	PLOD1	Yes	84				1.297E-04
Profilin-2	PFN2	No	15				2.818E-04
Proliferating cell nuclear antigen	PCNA	Yes	29				9.447E-05
Prolyl 3-hydroxylase 1	LEPRE1	No	83				4.679E-05
Proteasome activator complex subunit 3	PSME3	No	30				1.379E-04
Protein CYR61	CYR61	Yes	42				3.851E-04
Protein FRG1	FRG1	No	29				1.844E-04
Protein kinase C delta-binding protein	PRKCDBP	No	28				3.385E-04
Protein Niban	FAM129A	No	103				8.582E-05
Protein phosphatase 1 regulatory subunit 12A	PPP1R12A	No	115				2.354E-05
Protein phosphatase 1 regulatory subunit 16A	PPP1R16A	No	58				3.751E-05
Protein PML	PML	Yes	98				2.246E-05
Protein sprouty homolog 2	SPRY2	No	35				2.054E-04
Protein sprouty homolog 4	SPRY4	No	33				1.487E-04
Protein unc-13 homolog D	UNC13D	No	123				6.953E-05
Protein-glutamine gamma-glutamyltransferase 2	TGM2	Yes	77				2.492E-04
Protein-methionine sulfoxide oxidase MICAL2	MICAL2	No	127				2.834E-04
Protocadherin Fat 1	FAT1	No	506				7.560E-06
Rab3 GTPase-activating protein catalytic subunit	RAB3GAP1	No	111				2.530E-05
Ras-related protein Rab-32	RAB32	Yes	25				8.880E-04
RelA-associated inhibitor	PPP1R13L	No	89				5.836E-05
Rho guanine nucleotide exchange factor 40	ARHGEF40	No	165				1.888E-05
Rho-related GTP-binding protein RhoG	RHOG	No	21				5.254E-04
Sacsin	SACS	Yes	521				1.738E-05
Secretory carrier-associated membrane protein 4	SCAMP4	No	26				6.690E-04
Segment polarity protein dishevelled homolog DVL-2	DVL2	No	79				2.669E-05
Selenoprotein K	SELK	No	11				2.562E-04
Serine incorporator 1	SERINC1	No	50				3.010E-04
Serine/threonine-protein kinase OSR1	OXSR1	No	58				6.394E-05
Serum deprivation-response protein	SDPR	No	47				9.108E-05
Small integral membrane protein 12	SMIM12	No	11				4.306E-04
Small integral membrane protein 7	SMIM7	No	9				3.802E-04
Solute carrier family 2, facilitated glucose transporter member 3	SLC2A3	Yes	54				2.814E-04
Solute carrier family 43 member 3	SLC43A3	No	55				2.000E-04

Spermatogenesis-associated protein 5	SPATA5	No	98				2.780E-05
Spermidine synthase	SRM	No	34				1.140E-04
Spermine synthase	SMS	Yes	41				1.345E-04
Sterol	SOAT1	Yes	65				2.940E-04
Structural maintenance of chromosomes protein 2	SMC2	No	136				3.268E-05
SUN domain-containing protein 2	SUN2	No	80				1.953E-04
Syndecan-4	SDC4	Yes	22				1.992E-04
Synemin	SYNM	Yes	173				3.482E-05
TBC1 domain family member 4	TBC1D4	No	147				1.912E-05
Tensin-3	TNS3	No	155				2.383E-05
Testis-specific Y-encoded-like protein 2	TSPYL2	No	79				2.835E-05
Thioredoxin-interacting protein	TXNIP	No	44				6.201E-05
Thioredoxin-like protein 1	TXNL1	No	32				1.012E-04
Threonylcarbamoyladenosine tRNA methylthiotransferase	CDKAL1	No	65				6.076E-05
Translin	TSN	Yes	26				8.687E-05
Translocon-associated protein subunit alpha	SSR1	Yes	32				1.108E-03
Transmembrane emp24 domain-containing protein 4	TMED4	No	26				1.296E-04
Transmembrane protein 11, mitochondrial	TMEM11	No	22				3.095E-04
Trifunctional purine biosynthetic protein adenosine-3	GART	Yes	108				2.970E-05
Tubulin-folding cofactor B	TBCB	No	27				9.632E-05
Tumor necrosis factor receptor superfamily member 10B	TNFRSF10B	Yes	48				5.510E-05
Tyrosine-protein kinase JAK1	JAK1	No	133				5.135E-05
Tyrosine-protein kinase receptor UFO	AXL	No	98				1.378E-04
Tyrosine-protein phosphatase non-receptor type 2	PTPN2	No	48				5.942E-05
Uncharacterized protein C15orf52	C15orf52	No	57				6.171E-05
Uncharacterized protein KIAA1109	KIAA1109	No	555				3.925E-06
Unconventional myosin-IXb	MYO9B	No	243				3.247E-05
Vesicle transport protein SFT2B	SFT2D2	Yes	18				2.096E-04
Vinculin	VCL	No	124				4.834E-05
V-type proton ATPase 116 kDa subunit a isoform 3	TCIRG1	No	93				2.367E-05
WD repeat-containing protein 1	WDR1	No	66				7.215E-05
Zinc transporter ZIP1	SLC39A1	No	34				2.412E-04

Identified membrane proteins present in at least two cell lines

Identified Proteins	Gene	BC database	MW (kDa)	Average NSAF			
				HMEC	MCF7	SKBR3	MDA231
14-3-3 protein epsilon	YWHAE	Yes	29		2.835E-04		5.495E-04
1-acyl-sn-glycerol-3-phosphate acyltransferase beta	AGPAT2	No	31		1.910E-04		1.047E-04
1-acyl-sn-glycerol-3-phosphate acyltransferase epsilon	AGPAT5	No	42		4.553E-05		5.442E-05
2',3'-cyclic-nucleotide 3'-phosphodiesterase	CNP	No	48			6.856E-04	1.860E-04
26S protease regulatory subunit 10B	PSMC6	No	44	3.738E-04	1.022E-04		7.712E-05
26S protease regulatory subunit 4	PSMC1	No	49	2.392E-04	1.123E-04		9.003E-05
26S protease regulatory subunit 6A	PSMC3	No	49	3.095E-04	2.430E-04		1.764E-04
26S protease regulatory subunit 6B	PSMC4	Yes	47		5.975E-05		9.694E-05
26S proteasome non-ATPase regulatory subunit 1	PSMD1	No	106	1.124E-04		1.525E-04	
26S proteasome non-ATPase regulatory subunit 11	PSMD11	No	47	4.239E-04	9.645E-05		1.492E-04
26S proteasome non-ATPase regulatory subunit 12	PSMD12	Yes	53	2.710E-04	6.339E-05		
26S proteasome non-ATPase regulatory subunit 13	PSMD13	No	43	2.117E-04	9.779E-05		1.344E-04
26S proteasome non-ATPase regulatory subunit 6	PSMD6	No	46	7.709E-04	4.268E-04		6.339E-05
26S proteasome non-ATPase regulatory subunit 7	PSMD7	Yes	37	1.728E-04	1.014E-04		1.372E-04
28S ribosomal protein S14, mitochondrial	MRPS14	No	15		2.505E-04	8.424E-04	
28S ribosomal protein S30, mitochondrial	MRPS30	Yes	50	1.629E-04	6.708E-05	4.351E-04	
28S ribosomal protein S36, mitochondrial	MRPS36	No	11		1.916E-04	5.921E-04	
39S ribosomal protein L11, mitochondrial	MRPL11	Yes	21		3.396E-04		1.686E-04
39S ribosomal protein L13, mitochondrial	MRPL13	No	21		2.425E-04		1.925E-04
39S ribosomal protein L15, mitochondrial	MRPL15	No	33		2.187E-04		2.590E-04
39S ribosomal protein L23, mitochondrial	MRPL23	No	18		2.081E-04		3.738E-04
39S ribosomal protein L24, mitochondrial	MRPL24	No	25		2.776E-04		2.462E-04
39S ribosomal protein L37, mitochondrial	MRPL37	Yes	48	2.073E-04	1.057E-04	4.726E-04	
39S ribosomal protein L4, mitochondrial	MRPL4	No	35		1.819E-04		2.019E-04
39S ribosomal protein L41, mitochondrial	MRPL41	No	15		1.440E-04		1.824E-04
39S ribosomal protein L44, mitochondrial	MRPL44	No	38		1.182E-04		1.166E-04
39S ribosomal protein L50, mitochondrial	MRPL50	No	18		2.298E-04		3.470E-04
39S ribosomal protein L9, mitochondrial	MRPL9	No	30		1.360E-04		1.998E-04
3-ketoacyl-CoA thiolase, peroxisomal	ACAA1	No	44		9.562E-05	4.214E-04	
3-ketodihydrosphingosine reductase	KDSR	No	36	6.046E-04	7.057E-05		1.300E-04
3-keto-steroid reductase	HSD17B7	No	38		1.942E-04		2.813E-04
40S ribosomal protein S10	RPS10	No	19	2.375E-03	7.684E-04	5.023E-03	
40S ribosomal protein S19	RPS19	No	16	1.204E-03	2.158E-04	3.570E-03	
40S ribosomal protein S23	RPS23	No	16	6.254E-04	7.233E-04		1.202E-03
40S ribosomal protein S25	RPS25	No	14	9.269E-04			1.373E-03
40S ribosomal protein S26	RPS26	No	13		7.299E-04		7.261E-04
40S ribosomal protein S4, X isoform	RPS4X	Yes	30	4.528E-04	1.648E-03		2.084E-03
40S ribosomal protein S9	RPS9	No	23	1.613E-03	1.004E-03		2.246E-03

60 kDa heat shock protein, mitochondrial	HSPD1	No	61	1.961E-04	4.411E-04		
60 kDa SS-A/Ro ribonucleoprotein	TROVE2	No	61	2.101E-04	6.204E-05		
60S acidic ribosomal protein P0	RPLP0	Yes	34	1.076E-03	2.767E-04		2.484E-04
60S acidic ribosomal protein P2	RPLP2	No	12	1.173E-03	2.426E-04		
60S ribosomal protein L10a	RPL10A	No	25		6.716E-04		2.117E-03
60S ribosomal protein L17	RPL17	No	21	1.097E-03	8.009E-04		5.686E-04
60S ribosomal protein L19	RPL19	No	23			3.549E-04	3.048E-04
60S ribosomal protein L21	RPL21	Yes	19		8.647E-04		1.473E-03
60S ribosomal protein L24	RPL24	No	18	1.406E-03	5.931E-04		1.193E-03
60S ribosomal protein L27a	RPL27A	No	17	1.563E-03	1.159E-03		1.889E-03
60S ribosomal protein L30	RPL30	No	13	1.365E-03	4.484E-04		
60S ribosomal protein L31	RPL31	Yes	14	2.014E-03	5.704E-04	1.697E-03	
60S ribosomal protein L34	RPL34	No	13		3.353E-04		4.924E-04
60S ribosomal protein L35a	RPL35A	No	13		2.508E-04		6.447E-04
60S ribosomal protein L36	RPL36	No	12	1.016E-03	1.036E-03		2.265E-03
60S ribosomal protein L38	RPL38	No	8	1.484E-03	2.443E-03		7.416E-03
60S ribosomal protein L5	RPL5	Yes	34		6.379E-04		6.412E-04
7-dehydrocholesterol reductase	DHCR7	Yes	54	8.980E-04	1.004E-03		7.451E-04
Absent in melanoma 1 protein	AIM1	Yes	189	3.211E-04	1.937E-05		
Acetolactate synthase-like protein	ILVBL	No	68		3.945E-04	6.751E-04	
Acetyl-CoA carboxylase 1	ACACA	No	266		5.954E-05		4.404E-05
Acid sphingomyelinase-like phosphodiesterase 3b	SMPDL3B	No	51	1.788E-04	1.243E-04	5.292E-04	
Actin-related protein 2/3 complex subunit 1B	ARPC1B	Yes	41			5.731E-04	6.672E-05
Activating signal cointegrator 1 complex subunit 3	ASCC3	No	251	2.879E-05	7.526E-06		
Activity-dependent neuroprotector homeobox protein	ADNP	Yes	124		1.418E-04		3.953E-05
Acyl-CoA:lysophosphatidylglycerol acyltransferase 1	LPGAT1	No	43		4.479E-05		1.366E-04
Adenosylhomocysteinase	AHCY	Yes	48		8.583E-05		8.105E-05
ADP-ribosylation factor 4	ARF4	No	21			5.369E-04	7.538E-04
ADP-ribosylation factor GTPase-activating protein 1	ARFGAP1	No	45	1.023E-04	2.216E-04		
ADP-ribosylation factor-like protein 1	ARL1	No	20		4.097E-04		1.576E-04
ADP-ribosylation factor-like protein 6-interacting protein 1	ARL6IP1	No	23		1.439E-04		2.389E-04
A-kinase anchor protein 1, mitochondrial	AKAP1	No	97		5.967E-05	5.370E-05	
Aladin	AAAS	No	60	1.386E-04	2.448E-04		2.920E-04
Alanine--tRNA ligase, cytoplasmic	AARS	Yes	107		1.738E-04		3.524E-05
Aldehyde dehydrogenase family 16 member A1	ALDH16A1	No	85		5.671E-05		2.594E-04
Alkyldihydroxyacetonephosphate synthase, peroxisomal	AGPS	No	73	3.703E-04			1.278E-04
Alpha-(1,6)-fucosyltransferase	FUT8	Yes	67		2.320E-04		9.947E-05
Alpha-1,3/1,6-mannosyltransferase ALG2	ALG2	No	47		1.284E-04		1.336E-04
Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	MGAT1	No	51	2.098E-04	1.710E-04	1.640E-04	
Alpha-soluble NSF attachment protein	NAPA	No	33	2.698E-04	2.059E-04		2.028E-04
Amyloid-like protein 2	APLP2	Yes	87			1.120E-04	5.752E-05
Annexin A1	ANXA1	Yes	39	1.384E-04			9.877E-04
Annexin A11	ANXA11	Yes	54			9.502E-04	1.097E-04
Annexin A6	ANXA6	No	76			6.646E-04	3.003E-04
Anoctamin-10	ANO10	No	76	1.233E-04	3.025E-05		9.851E-05
Antigen peptide transporter 2	TAP2	No	76			1.605E-04	8.446E-05
Apoptotic chromatin condensation inducer in the nucleus	ACIN1	No	152		4.845E-05	3.580E-04	
Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 1	ARAP1	No	162		3.306E-05		2.761E-05
Aspartyl aminopeptidase	DNPEP	No	52		1.090E-04	2.796E-04	
ATP synthase subunit e, mitochondrial	ATP5I	No	8	2.649E-03	1.429E-02		4.133E-04
ATP synthase subunit O, mitochondrial	ATP5O	No	23	6.365E-04	9.054E-04		2.060E-04
ATP-binding cassette sub-family B member 8, mitochondrial	ABCB8	No	80		2.181E-05	9.997E-05	
ATP-binding cassette sub-family F member 1	ABCF1	No	96		3.393E-05	1.597E-04	
ATP-dependent RNA helicase DDX50	DDX50	No	83		9.392E-05		2.903E-04
ATP-dependent RNA helicase DDX51	DDX51	No	72		5.567E-05		8.120E-05
ATP-dependent zinc metalloprotease YME1L1	YME1L1	Yes	86		1.251E-04		8.165E-05
Bcl2 antagonist of cell death	BAD	Yes	18	5.556E-04			4.275E-04
Bcl-2 homologous antagonist/killer	BAK1	No	23	4.507E-04	2.130E-04		
BCL2/adenovirus E1B 19 kDa protein-interacting protein 3	BNIP3	Yes	22		1.871E-04		4.418E-04
Bcl-2-like protein 13	BCL2L13	No	53	1.591E-04	7.485E-05		
Beta-1,4-galactosyltransferase 1	B4GALT1	No	44		1.323E-04	4.795E-04	
Beta-2-syntrophin	SNTB2	Yes	58		7.551E-05		7.443E-05
Beta-galactosidase	GLB1	No	76	2.305E-04	4.736E-05	2.093E-04	
Brefeldin A-inhibited guanine nucleotide-exchange protein 3	ARFGEF3	No	241		3.056E-05	6.056E-05	
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	Yes	102		2.008E-04		8.359E-05
CAAX prenyl protease 2	RCE1	No	36		1.701E-04		5.971E-05
Cadherin-1	CDH1	Yes	97	1.502E-04	1.267E-04		
Calcineurin B homologous protein 1	CHP1	No	22	6.627E-04	7.392E-04		5.547E-04
Calcium-transporting ATPase type 2C member 2	ATP2C2	No	103		9.705E-05	8.949E-05	
Calpain small subunit 1	CAPNS1	No	28		3.307E-04		5.084E-04
Calpain-1 catalytic subunit	CAPN1	No	82		1.049E-04		8.403E-05

Canalicular multispecific organic anion transporter 2	ABCC3	No	169	7.818E-05	7.117E-05		
Cancer-related nucleoside-triphosphatase	NTPCR	No	21		1.509E-04		1.306E-04
Catalase	CAT	No	60		6.991E-05	1.582E-04	
Catechol	COMTD1	No	29		9.238E-05	6.777E-04	
Catenin alpha-1	CTNNA1	Yes	100	4.086E-04	1.851E-04		6.395E-05
Catenin beta-1	CTNNB1	Yes	85	1.030E-02	3.556E-04		1.717E-04
Caveolin-1	CAV1	Yes	20	2.775E-03			2.473E-03
CD109 antigen	CD109	No	162	9.922E-04	2.480E-05		
CD151 antigen	CD151	No	28			5.153E-04	2.586E-04
CD276 antigen	CD276	No	57	2.309E-04	3.222E-04		
CD82 antigen	CD82	Yes	30	9.937E-04		5.442E-04	5.572E-04
CDGSH iron-sulfur domain-containing protein 1	CISD1	No	12		2.605E-04	1.091E-03	
Cell cycle and apoptosis regulator protein 2	CCAR2	No	103		2.706E-05	1.049E-04	
Cell division cycle 5-like protein	CDC5L	No	92		4.623E-05		4.919E-05
Cell surface glycoprotein MUC18	MCAM	No	72	1.625E-04			1.519E-04
Centromere/kinetochore protein zw10 homolog	ZW10	No	89	1.460E-04	1.311E-04		1.146E-04
Cholinephosphotransferase 1	CHPT1	Yes	45		8.809E-05		4.838E-05
Chromatin target of PRMT1 protein	CHTOP	No	26		3.574E-04		7.191E-04
Chromodomain-helicase-DNA-binding protein 4	CHD4	No	218		4.003E-05		2.111E-05
CKLF-like MARVEL transmembrane domain-containing protein 4	CMTM4	No	26		1.743E-04	2.972E-04	
Claudin-3	CLDN3	No	23		7.018E-04	9.258E-04	
Cleft lip and palate transmembrane protein 1-like protein	CLPTM1L	No	62		1.374E-04	1.314E-04	
CLIP-associating protein 2	CLASP2	No	141	6.422E-05			1.531E-05
Cluster of Brefeldin A-inhibited guanine nucleotide-exchange protein 2	ARFGEF2	No	202		2.371E-05	3.960E-05	
Cluster of Chloride intracellular channel protein 1	CLIC1	Yes	27		1.321E-04		7.676E-04
Cluster of Cytochrome b-c1 complex subunit Rieske, mitochondrial	UQCRFS1	Yes	30		3.809E-04		2.535E-04
Cluster of Double-stranded RNA-binding protein Staufen homolog 1	STAU1	No	63		3.583E-04	2.447E-04	
Cluster of Elongation factor 1-alpha 1	EEF1A1	No	50	1.108E-02	7.340E-03		5.239E-03
Cluster of Epithelial splicing regulatory protein 1	ESRP1	No	76	9.568E-05	5.874E-05		
Cluster of Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 1	GFPT1	No	79	1.361E-04	1.663E-03		3.588E-04
Cluster of GTP-binding protein SAR1a	SAR1A	No	22	4.231E-04	3.932E-04		3.499E-04
Cluster of Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	No	40		5.114E-04	8.378E-04	
Cluster of Histone deacetylase 1	HDAC1	Yes	55		4.045E-05	1.262E-04	
Cluster of Histone H1.4	HIST1H1E	Yes	22	1.972E-03	1.688E-03		2.758E-03
Cluster of Histone H2A type 1-B/E	HIST1H2AB	No	14		6.842E-04		2.800E-03
Cluster of Integrin alpha-6	ITGA6	No	127	1.614E-03	3.867E-05		1.783E-04
Cluster of Isoform 2 of Tropomyosin beta chain	TPM2	Yes	33	1.314E-03		2.141E-03	
Cluster of Keratin, type II cytoskeletal 7	KRT7	Yes	51	6.187E-03		6.489E-03	6.314E-03
Cluster of Kunitz-type protease inhibitor 1	SPINT1	No	58	3.669E-04	6.863E-05	4.319E-04	
Cluster of LIM domain and actin-binding protein 1	LIMA1	No	85	2.000E-04		3.258E-03	
Cluster of Membrane cofactor protein	CD46	No	44		1.947E-04	1.080E-03	
Cluster of Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5	MACF1	No	838	2.737E-05	3.171E-06		2.997E-05
Cluster of Polypeptide N-acetylgalactosaminyltransferase 4	GALNT4	No	67	8.893E-05	1.388E-04		
Cluster of Ras-related protein Rap-2c	RAP2C	No	21		8.351E-03		7.994E-04
Cluster of Receptor-type tyrosine-protein phosphatase F	PTPRF	Yes	213		9.737E-05	1.381E-04	
Cluster of Sodium/hydrogen exchanger 1	SLC9A1	Yes	91		1.199E-04	8.974E-05	
Cluster of Sphingomyelin phosphodiesterase 4	SMPD4	No	93		5.816E-05		7.036E-05
Cluster of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2	SMARCD2	No	59		1.507E-04	1.591E-04	
Cluster of Tropomyosin alpha-4 chain	TPM4	No	29	1.456E-03		1.890E-03	
Cluster of Tyrosine-protein kinase Lyn	LYN	Yes	59	3.307E-04			4.381E-04
Coatomer subunit delta	ARCN1	Yes	57	1.357E-04	1.272E-04		
Coatomer subunit gamma-1	COPG1	No	98	1.543E-04	3.290E-04		8.690E-05
Coatomer subunit zeta-1	COPZ1	No	20	3.582E-04	2.513E-04		
Complement decay-accelerating factor	CD55	No	41		1.809E-04		2.351E-04
Condensin complex subunit 1	NCAPD2	No	157		2.894E-05		2.141E-05
Conserved oligomeric Golgi complex subunit 3	COG3	No	94		3.938E-05		7.117E-05
Conserved oligomeric Golgi complex subunit 5	COG5	No	93		5.301E-05		4.810E-05
Core histone macro-H2A.1	H2AFY	No	40	2.792E-04	4.621E-04		1.463E-04
Crooked neck-like protein 1	CRNKL1	No	100		3.845E-05	1.560E-04	
cTAGE family member 5	CTAGE5	No	91			7.564E-05	4.815E-05
CTP synthase 1	CTPS1	No	67	1.639E-04			5.048E-05
Cysteine-rich with EGF-like domain protein 1	CRELD1	Yes	45	1.744E-04		1.981E-04	
Cytochrome b5 type B	CYB5B	No	16	1.487E-03	4.571E-03		3.120E-03
Cytochrome b-c1 complex subunit 7	UQCRB	No	14		1.711E-03	1.518E-03	
Cytochrome c oxidase subunit 2	COII	No	26	5.586E-04	1.610E-03		6.473E-04
Cytochrome c oxidase subunit 6C	COX6C	Yes	9		2.111E-03		9.654E-04
Cytochrome c oxidase subunit 7A-related protein, mitochondrial	COX7A2L	Yes	13		2.488E-04		2.992E-04

Cytochrome P450 20A1	CYP20A1	No	52		3.997E-04		1.695E-04
Cytochrome P450 4F11	CYP4F11	No	60	2.613E-04	1.109E-04		
Cytoplasmic dynein 1 intermediate chain 2	DYNC1I2	No	71	8.057E-05		9.327E-05	
Cytoplasmic dynein 1 light intermediate chain 2	DYNC1LI2	No	54	1.489E-04	5.076E-05	1.744E-04	
Cytoskeleton-associated protein 5	CKAP5	No	226	1.288E-04	6.202E-05		9.663E-05
DBIRD complex subunit ZNF326	ZNF326	No	66		4.291E-05	4.301E-04	
Death-inducer obliterator 1	DIDO1	No	244		2.491E-05		1.075E-05
Dehydrogenase/reductase SDR family member 1	DHRS1	No	34	7.024E-04	1.411E-04	5.000E-04	
Dehydrogenase/reductase SDR family member 7	DHRS7	Yes	38	2.586E-04	4.898E-04		
Dehydrogenase/reductase SDR family member on chromosome X	DHRSX	No	36		1.474E-04		1.056E-04
Delta(24)-sterol reductase	DHCR24	Yes	60	7.712E-05	1.431E-04	2.772E-04	
Derlin-1	DERL1	No	29		1.093E-03		7.636E-04
Desmoglein-1	DSG1	No	114	1.169E-04	5.973E-05		
Destrin	DSTN	No	19		2.887E-04	1.444E-03	
Disintegrin and metalloproteinase domain-containing protein 9	ADAM9	No	91	1.165E-04		3.853E-04	
DNA topoisomerase 2-alpha	TOP2A	No	174		7.199E-05	1.633E-04	
DNA topoisomerase 2-beta	TOP2B	No	183	5.957E-05	3.135E-04		7.395E-05
DnaJ homolog subfamily B member 1	DNAJB1	No	38		1.177E-04		7.349E-05
DnaJ homolog subfamily B member 2	DNAJB2	Yes	36	1.653E-04	1.227E-04		1.063E-04
DnaJ homolog subfamily C member 11	DNAJC11	No	63		1.237E-04	8.644E-05	
DnaJ homolog subfamily C member 15	DNAJC15	No	16		2.472E-04		1.666E-04
DnaJ homolog subfamily C member 5	DNAJC5	No	22		1.234E-03		4.500E-04
DnaJ homolog subfamily C member 7	DNAJC7	No	56		8.254E-05		8.884E-05
Dolichol-phosphate mannosyltransferase	DPM1	Yes	30	3.658E-04	2.861E-04		3.512E-04
Dolichol-phosphate mannosyltransferase subunit 3	DPM3	No	10		3.603E-04		3.743E-04
Double-stranded RNA-specific adenosine deaminase	ADAR	No	136	6.492E-05	1.993E-05	4.033E-04	
Dynactin subunit 1	DCTN1	Yes	142	4.793E-05	2.551E-05		
E3 ubiquitin-protein ligase MARCH5	MARCH5	No	31		4.059E-04		2.158E-04
E3 ubiquitin-protein ligase MIB1	MIB1	No	110			7.174E-05	1.953E-05
E3 ubiquitin-protein ligase RNF170	RNF170	No	30			2.784E-04	7.677E-05
E3 ubiquitin-protein ligase TRIP12	TRIP12	No	220	4.578E-05	3.885E-05		4.481E-05
Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial	ETFDH	No	68		4.010E-05	1.345E-04	
ELMO domain-containing protein 2	ELMOD2	No	35		1.265E-04		1.492E-04
Elongation factor 2	EEF2	Yes	95	9.071E-05	2.510E-04		1.951E-04
Elongation of very long chain fatty acids protein 5	ELOVL5	Yes	35		2.400E-04		3.565E-04
Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	MAN1B1	No	80		1.158E-04		3.551E-05
Endoplasmic reticulum resident protein 44	ERP44	No	47		9.662E-05	2.688E-04	
Endoplasmic reticulum-Golgi intermediate compartment protein 3	ERGIC3	No	43	1.296E-04	6.521E-05		
Enhancer of mRNA-decapping protein 4	EDC4	No	152		7.548E-05		2.435E-05
Enhancer of rudimentary homolog	ERH	No	12	8.741E-04	3.416E-04	1.406E-03	
Ephrin type-A receptor 2	EPHA2	Yes	108	1.302E-04			6.918E-04
Epidermal growth factor receptor	EGFR	Yes	134	7.246E-05		7.987E-05	1.021E-04
Epiplakin	EPPK1	No	556		6.356E-05	2.626E-04	
Epithelial cell adhesion molecule	EPCAM	No	35		1.557E-03	1.839E-03	
ER membrane protein complex subunit 2	EMC2	No	35	3.629E-04	4.101E-04		3.971E-04
ER membrane protein complex subunit 3	EMC3	No	30	5.194E-04	4.577E-04		3.527E-04
ER membrane protein complex subunit 6	EMC6	No	12		7.294E-04		4.528E-04
ER membrane protein complex subunit 7	EMC7	No	26	2.720E-04	4.937E-04		5.386E-04
Erlin-1	ERLIN1	No	39	8.998E-04	5.106E-04		5.529E-04
Eukaryotic translation elongation factor 1 epsilon-1	EEF1E1	No	20		3.238E-04		3.339E-04
Eukaryotic translation initiation factor 2 subunit 2	EIF2S2	No	38	1.365E-04		2.532E-04	
Eukaryotic translation initiation factor 2 subunit 3	EIF2S3	No	51	1.052E-04		4.088E-04	
Eukaryotic translation initiation factor 3 subunit A	EIF3A	No	167	5.886E-05	1.149E-04		1.666E-04
Eukaryotic translation initiation factor 3 subunit B	EIF3B	No	92		7.519E-05		1.206E-04
Eukaryotic translation initiation factor 3 subunit E	EIF3E	No	52	1.297E-04	1.081E-04		1.082E-04
Eukaryotic translation initiation factor 3 subunit F	EIF3F	No	38	2.218E-04			1.411E-04
Exocyst complex component 4	EXOC4	No	111	1.078E-04		1.846E-04	7.889E-05
Exocyst complex component 6B	EXOC6B	No	94	6.603E-05			4.844E-05
Exportin-7	XPO7	No	124	6.639E-05	2.759E-05		
Exportin-T	XPOT	No	110		5.024E-05		3.580E-05
Ezrin	EZR	No	69		6.327E-05	1.671E-04	
FACT complex subunit SPT16	SUPT16H	No	120	1.699E-04	4.645E-05		5.518E-05
FACT complex subunit SSRP1	SSRP1	Yes	81	2.837E-04	4.599E-05		3.478E-05
F-actin-capping protein subunit alpha-2	CAPZA2	No	33	3.670E-04		6.293E-04	
Far upstream element-binding protein 1	FUBP1	No	68	2.183E-04	3.489E-04	1.526E-04	
FAS-associated factor 2	FAF2	No	53		3.778E-04		1.976E-04
Fibronectin	FN1	Yes	263	1.557E-04		1.402E-04	3.108E-04
Filaggrin-2	FLG2	No	248		4.699E-05		2.248E-05
Galectin-1	LGALS1	Yes	15	1.594E-03	3.095E-04		9.236E-04
Galectin-3	LGALS3	Yes	26	8.126E-04	1.668E-03		

Galectin-3-binding protein	LGALS3BP	Yes	65			3.751E-04	2.175E-04
Gamma-glutamyltranspeptidase 1	GGT1	No	61	1.114E-04		2.297E-04	1.712E-04
Gem-associated protein 4	GEMIN4	No	120	1.196E-04	6.914E-05		1.168E-04
Gem-associated protein 5	GEMIN5	No	169	5.672E-05	4.278E-05		
GH3 domain-containing protein	GHDC	No	58	1.535E-04		3.625E-04	1.149E-04
Glucose-6-phosphate 1-dehydrogenase	G6PD	No	59		3.798E-04	2.340E-04	
Glucosidase 2 subunit beta	PRKCSH	No	59	2.995E-04	4.229E-04	3.602E-04	
Glutaminyl-peptide cyclotransferase-like protein	QPCTL	No	43		2.209E-04		1.037E-04
Glycerophosphodiester phosphodiesterase domain-containing protein 3	GDPD3	No	37		3.044E-04	1.375E-03	
Glycine--tRNA ligase	GARS	Yes	83		2.670E-05		7.195E-05
Glycogen [starch] synthase, muscle	GYS1	No	84	1.453E-04			7.372E-05
Glycogen phosphorylase, brain form	PYGB	No	97	1.027E-03	1.650E-04		1.123E-04
Glycosyltransferase 8 domain-containing protein 1	GLT8D1	No	42		9.848E-05	1.302E-04	
Golgi pH regulator B	GPR89B	No	53		5.077E-04		2.625E-04
Golgi SNAP receptor complex member 2	GOSR2	No	25		1.327E-04		2.272E-04
Golgin subfamily A member 5	GOLGA5	No	83		8.668E-05		3.396E-05
Golgin subfamily B member 1	GOLGB1	No	376		5.085E-06	4.816E-05	
Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	GBF1	No	206		6.792E-05	9.133E-05	
GPI ethanolamine phosphate transferase 2	PIGG	No	108	6.847E-05	1.793E-04		
G-protein coupled receptor 56	GPR56	No	78		7.630E-05	2.438E-04	
Growth hormone-inducible transmembrane protein	GHITM	No	37	3.675E-04	5.060E-04		3.821E-04
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	Yes	35	5.748E-04	1.658E-04		2.162E-04
HCLS1-associated protein X-1	HAX1	No	32		2.893E-04		3.659E-04
Heat shock protein 75 kDa, mitochondrial	TRAP1	Yes	80		2.386E-04		2.916E-04
Helicase with zinc finger domain 2	HELZ2	No	295		5.409E-05		5.213E-05
Heme oxygenase 2	HMOX2	No	36	2.130E-04	3.530E-04		3.376E-04
Heterochromatin protein 1-binding protein 3	HP1BP3	No	61	1.547E-04	1.965E-04		2.191E-04
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	No	37		1.617E-04	3.725E-04	
Heterogeneous nuclear ribonucleoprotein L	HNRNPL	No	64	1.853E-04		3.224E-04	
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	Yes	70	4.579E-04	4.375E-04	3.483E-04	
Hexokinase-2	HK2	Yes	102			9.319E-05	2.176E-04
High affinity cationic amino acid transporter 1	SLC7A1	Yes	68	2.145E-04	2.013E-04		6.699E-04
Histone H1.5	HIST1H1B	Yes	23	5.347E-04	7.138E-04		
Histone H2A.Z	H2AFZ	Yes	14		4.234E-04		7.993E-04
Histone-lysine N-methyltransferase 2A	KMT2A	No	432		1.305E-05		2.245E-05
Hypoxia up-regulated protein 1	HYOU1	Yes	111		1.415E-04		3.597E-04
Importin-4	IPO4	No	119	1.010E-04	1.333E-04		
Inactive hydroxysteroid dehydrogenase-like protein 1	HSDL1	No	37		3.611E-04		1.178E-04
Inactive tyrosine-protein kinase 7	PTK7	No	118	2.572E-04	3.393E-05		5.108E-05
Inner nuclear membrane protein Man1	LEMD3	No	100		3.638E-05		6.487E-05
Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	Yes	56	1.698E-04	5.428E-05		7.563E-05
Inositol 1,4,5-trisphosphate receptor type 2	ITPR2	No	308		4.109E-05		2.608E-05
Integrator complex subunit 1	INTS1	No	244	3.616E-05	1.668E-05		
Integrin alpha-2	ITGA2	Yes	129	5.282E-04	9.202E-05		1.203E-04
Integrin alpha-3	ITGA3	No	117	1.059E-04			1.916E-04
Integrin alpha-5	ITGA5	No	115	3.032E-04	7.723E-05		1.623E-04
Integrin beta-4	ITGB4	Yes	202	1.143E-03	9.138E-05		7.551E-05
Integrin beta-5	ITGB5	Yes	88	1.692E-04	8.808E-05	5.072E-04	
Integrin beta-6	ITGB6	No	86	5.003E-04		1.358E-04	
Interferon-induced transmembrane protein 3	IFITM3	Yes	15			1.615E-03	1.381E-03
Intron-binding protein aquarius	AQR	Yes	171		1.931E-05	6.508E-05	
Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	Yes	51		6.055E-04	6.920E-04	
Isoform 1 of Plakophilin-2	PKP2	Yes	93	9.709E-05		1.097E-04	
Isoform 1 of Scavenger receptor class B member 1	SCARB1	No	57	3.152E-04	8.993E-05	3.508E-04	
Isoform 11 of CD44 antigen	CD44	Yes	47	1.099E-03	1.297E-04		2.438E-03
Isoform 13 of Sodium bicarbonate cotransporter 3	SLC4A7	Yes	128	8.500E-05	5.676E-05	5.009E-05	
Isoform 2 of CD166 antigen	ALCAM	Yes	64	5.264E-04	1.272E-04		4.214E-04
Isoform 2 of F-actin-capping protein subunit beta	CAPZB	No	31	1.371E-03		1.730E-03	
Isoform 2 of Protein unc-45 homolog A	UNC45A	No	102	8.431E-05	7.096E-05	5.137E-05	
Isoform 2 of Serine/threonine-protein kinase MRCK alpha	CDC42BPA	No	200			3.436E-05	5.656E-05
Isoform 2 of Tropomyosin alpha-3 chain	TPM3	No	29	1.736E-03		2.715E-03	
Isoform 2 of Zinc transporter ZIP11	SLC39A11	No	35		4.305E-04	2.013E-03	
Isoform 3 of Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	No	31		1.235E-04	4.732E-04	
Isoform 3 of Treacle protein	TCOF1	No	152			3.247E-05	4.364E-05
Isoform 4 of 39S ribosomal protein L43, mitochondrial	MRPL43	No	18		9.177E-05		1.334E-04
Isoform 5 of Brain-specific angiogenesis inhibitor 1-associated protein 2	BAIAP2	No	57	1.046E-04	2.904E-05		
Isoform 8 of Dysferlin	DYSF	Yes	241	1.034E-04			4.310E-04
Isoform Alpha of Poliovirus receptor-related protein 2	PVRL2	No	51		2.937E-05		7.226E-05
Isoform SERCA3A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	ATP2A3	No	109		1.352E-03	2.601E-03	

Junctional adhesion molecule A	F11R	No	33	1.291E-03	8.043E-04	1.369E-03	
Keratin, type I cytoskeletal 17	KRT17	Yes	48	2.506E-02			1.384E-03
Keratin, type I cytoskeletal 19	KRT19	Yes	44		3.791E-03	1.313E-02	
Keratinocyte proline-rich protein	KPRP	No	64		2.030E-04		5.919E-05
Kinase D-interacting substrate of 220 kDa	KIDINS220	No	197		4.365E-05		3.379E-05
KN motif and ankyrin repeat domain-containing protein 2	KANK2	No	91		3.306E-05		3.525E-05
Kunitz-type protease inhibitor 2	SPINT2	No	28	5.085E-04	8.234E-04	2.219E-03	
Lactotransferrin	LTF	Yes	78		1.430E-04		1.916E-04
Ladinin-1	LAD1	Yes	57	8.792E-05		5.577E-04	
Laminin subunit beta-3	LAMB3	No	130	1.361E-03			2.055E-05
Large proline-rich protein BAG6	BAG6	No	119		1.056E-04		3.042E-05
Lethal(2) giant larvae protein homolog 2	LLGL2	No	113		1.094E-04	9.475E-05	
Leucine-rich repeat-containing protein 59	LRRC59	No	35	4.152E-04	1.068E-03		1.252E-03
LIM domain only protein 7	LMO7	No	193	5.651E-05		3.536E-05	
Lipase maturation factor 2	LMF2	No	80	1.713E-04	9.180E-05		6.753E-05
Lipolysis-stimulated lipoprotein receptor	LSR	No	71	3.941E-04	2.349E-04	2.429E-04	
L-lactate dehydrogenase A chain	LDHA	Yes	37	7.342E-04	7.005E-04		1.237E-03
Long-chain fatty acid transport protein 1	SLC27A1	No	71	1.499E-04			1.237E-04
Long-chain-fatty-acid--CoA ligase 4	ACSL4	No	79	8.121E-05			4.689E-04
Low-density lipoprotein receptor	LDLR	Yes	95	9.088E-04	8.470E-05		2.448E-04
Ly6/PLAUR domain-containing protein 3	LYPD3	No	36	3.104E-04	2.577E-04	6.669E-04	
Lymphocyte function-associated antigen 3	CD58	No	28		1.726E-04		1.555E-04
Lysine--tRNA ligase	KARS	Yes	68	2.459E-04	4.803E-05		
Lysocardiolipin acyltransferase 1	LCLAT1	No	49		1.264E-04		1.316E-04
Lysophospholipid acyltransferase 2	MBOAT2	No	60		8.642E-05	6.926E-04	
Lysophospholipid acyltransferase 5	LPCAT3	No	56		1.518E-04	4.408E-04	
Lysophospholipid acyltransferase LPCAT4	LPCAT4	No	57	1.398E-04			9.505E-05
Lysosome-associated membrane glycoprotein 2	LAMP2	Yes	45	3.007E-04	8.667E-04		1.099E-04
Magnesium transporter protein 1	MAGT1	No	38		4.678E-04		4.578E-04
Major facilitator superfamily domain-containing protein 10	MFSD10	No	48		4.343E-04		3.277E-04
Major prion protein	PRNP	Yes	28	2.647E-03		5.595E-04	1.104E-03
Major vault protein	MVP	Yes	99	7.982E-04		1.910E-04	1.655E-04
Mannose-P-dolichol utilization defect 1 protein	MPDU1	Yes	27		6.722E-04		8.591E-04
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IB	MAN1A2	Yes	73		4.341E-05	9.487E-05	
MAP7 domain-containing protein 1	MAP7D1	No	93	1.508E-04			4.007E-05
MARCKS-related protein	MARCKSL1	No	20		3.885E-04		2.251E-04
Matrix metalloproteinase-14	MMP14	Yes	66	2.427E-04			2.011E-04
Melanoma inhibitory activity protein 3	MIA3	No	214		8.690E-06		2.525E-05
Melanoma-associated antigen D2	MAGED2	No	65		1.515E-04		1.398E-04
Melanotransferrin	MFI2	No	80	5.160E-04			1.500E-04
Metaxin-2	MTX2	No	30		2.318E-04		1.112E-04
Methylsterol monooxygenase 1	MSMO1	No	35	5.516E-04			6.704E-05
Methyltransferase-like protein 7A	METTL7A	No	28	3.262E-04	1.659E-04	1.160E-03	
Microsomal glutathione S-transferase 2	MGST2	No	17		1.007E-03		1.124E-03
Microtubule-associated protein 4	MAP4	No	121	1.913E-04			5.144E-05
Midasin	MDN1	No	633		2.538E-05		3.566E-05
Mitochondrial carrier homolog 1	MTCH1	No	42		2.970E-04		2.624E-04
Mitochondrial carrier homolog 2	MTCH2	Yes	33		3.187E-04		1.150E-04
Mitochondrial ornithine transporter 1	SLC25A15	No	33		2.830E-04		1.144E-04
Mitochondrial Rho GTPase 2	RHOT2	No	68	1.185E-04	2.777E-04		
Mitotic interactor and substrate of PLK1	MISP	No	75		4.768E-05	8.296E-04	
Monocarboxylate transporter 4	SLC16A3	Yes	49	5.743E-04		1.087E-03	2.075E-03
Motile sperm domain-containing protein 2	MOSPD2	No	60	1.143E-04	4.777E-05		
Multifunctional protein ADE2	PAICS	No	47		2.940E-04		1.961E-04
Myelin protein zero-like protein 2	MPZL2	No	24	6.456E-04	9.177E-05	3.341E-04	
Myeloid-associated differentiation marker	MYADM	No	35		1.087E-04		2.685E-04
NAD(P) transhydrogenase, mitochondrial	NNT	No	114	7.652E-05	6.259E-04		2.651E-04
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 4	NDUFAF4	No	20		2.044E-04		1.122E-04
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	NDUFA12	No	17		8.617E-04		5.838E-04
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	NDUFA6	No	18		4.169E-04	5.390E-04	
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7	NDUFA7	Yes	13		9.501E-04	9.774E-04	
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	NDUFA9	Yes	43	1.467E-04	4.099E-04		2.034E-04
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 1	NDUFB1	Yes	7	1.505E-03	5.077E-04		3.387E-04
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 3	NDUFB3	No	11	7.295E-04	3.862E-04		
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4	NDUFB4	No	15		4.404E-04		4.045E-04
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 8, mitochondrial	NDUFB8	No	22	5.244E-04	7.422E-04	3.801E-04	
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	NDUFB9	No	22		7.510E-04	1.081E-03	
NADH dehydrogenase [ubiquinone] iron-sulfur protein 7, mitochondrial	NDUFS7	No	24		1.853E-04		1.165E-04

NADH-cytochrome b5 reductase 1	CYB5R1	Yes	34	2.875E-04	1.466E-04		2.145E-04
Neuroblastoma-amplified sequence	NBAS	No	269		7.042E-05		1.421E-05
Neuropathy target esterase	PNPLA6	No	150	3.634E-05	6.201E-05		1.987E-04
Neuroplastin	NPTN	No	44		4.133E-04		4.006E-04
NF-kappa-B-repressing factor	NKRF	No	78		4.882E-05	8.789E-05	
NKG2D ligand 2	ULBP2	No	27	1.618E-04			1.574E-04
Non-POU domain-containing octamer-binding protein	NONO	No	54		1.121E-04	2.072E-04	
Nuclear cap-binding protein subunit 1	NCBP1	No	92	9.852E-05		1.369E-04	
Nuclear envelope pore membrane protein POM 121C	POM121C	No	125		3.983E-05		2.428E-05
Nuclear pore complex protein Nup153	NUP153	No	154		7.979E-05		6.269E-05
Nuclear pore complex protein Nup98-Nup96	NUP98	No	198		3.704E-05	8.612E-05	
Nucleolar GTP-binding protein 2	GNL2	No	84		4.429E-05		9.905E-05
Nucleolar protein 9	NOP9	No	69	1.627E-04	1.842E-04	1.114E-04	
Nucleophosmin	NPM1	No	33	5.593E-04			3.354E-04
Nucleoporin NDC1	NDC1	No	76		4.952E-05		5.829E-05
Nucleoporin NUP188 homolog	NUP188	No	196	3.402E-05	1.128E-05		1.123E-05
Nucleoporin NUP53	NUP35	No	35		9.764E-05	2.935E-04	
Nucleoside diphosphate kinase 3	NME3	Yes	19		4.296E-04	7.714E-04	
Occludin	OCLN	No	59		1.859E-04	1.833E-04	
OCIA domain-containing protein 2	OCIAD2	No	17	5.421E-04	7.588E-04		2.081E-03
ORM1-like protein 1	ORMDL1	No	17		6.877E-04	1.250E-03	
Oxysterol-binding protein-related protein 5	OSBPL5	No	99		5.200E-05		8.826E-05
Oxysterol-binding protein-related protein 8	OSBPL8	No	101		2.133E-04		2.878E-04
Peptidyl-prolyl cis-trans isomerase B	PPIB	No	24	5.884E-04	1.039E-03		4.501E-04
Peptidyl-prolyl cis-trans isomerase FKBP11	FKBP11	No	22	3.751E-04	2.788E-04		1.963E-04
Peptidyl-tRNA hydrolase 2, mitochondrial	PTRH2	No	19		1.640E-03		1.336E-03
Perilipin-3	PLIN3	No	47		3.619E-04		2.683E-04
Peroxidasin homolog	PXDN	No	165		3.007E-05		5.057E-05
Peroxisomal membrane protein 2	PXMP2	No	22		1.593E-04		2.251E-04
Peroxisomal membrane protein PEX14	PEX14	No	41		1.061E-04		9.288E-05
Phenylalanine--tRNA ligase alpha subunit	FARSA	No	58	5.969E-04	9.620E-05	1.421E-04	
PH-interacting protein	PHIP	No	207		5.735E-05		1.566E-05
Phosphatidylglycerophosphatase and protein-tyrosine phosphatase 1	PTPMT1	No	23		6.365E-04		2.404E-04
Phosphatidylinositol 4-kinase alpha	PI4KA	No	231	1.118E-04	4.363E-05		1.721E-05
Phosphatidylinositol 4-kinase type 2-alpha	PI4K2A	No	54	6.094E-04	1.289E-04		2.651E-04
Phosphatidylserine synthase 1	PTDSS1	Yes	56		4.517E-04		6.607E-04
Phospholipase D3	PLD3	No	55	1.214E-04	8.919E-05		
Piezo-type mechanosensitive ion channel component 1	PIEZO1	No	287		9.499E-05	6.155E-05	
Plakophilin-1	PKP1	Yes	83	6.803E-04	3.312E-05		
Plakophilin-3	PKP3	No	87	7.283E-04		4.358E-04	2.465E-05
Poliovirus receptor	PVR	No	45	1.520E-04			1.971E-04
Poliovirus receptor-related protein 4	PVRL4	No	55		1.200E-04	5.376E-04	
Poly [ADP-ribose] polymerase 1	PARP1	No	113	4.937E-04		3.221E-04	1.064E-04
Polypeptide N-acetylgalactosaminyltransferase 3	GALNT3	Yes	73	7.843E-05	1.966E-04	1.511E-04	
Polypeptide N-acetylgalactosaminyltransferase 6	GALNT6	No	71		1.788E-04	4.195E-04	
PRA1 family protein 3	ARL6IP5	No	22	2.259E-03	6.851E-04		7.003E-04
Pre-mRNA-processing factor 6	PRPF6	No	107		1.262E-04	3.074E-04	
Pre-mRNA-splicing factor SYF1	XAB2	No	100		5.689E-05	2.305E-04	
Prenylcysteine oxidase-like	PCYOX1L	No	55	2.452E-04	1.785E-04		
pre-rRNA processing protein FTSJ3	FTSJ3	No	97		3.609E-04		3.886E-04
Presenilins-associated rhomboid-like protein, mitochondrial	PARL	No	42		2.955E-04	4.450E-04	
Probable ATP-dependent RNA helicase DDX20	DDX20	No	92		5.426E-05		8.338E-05
Probable ATP-dependent RNA helicase DDX23	DDX23	No	96		3.046E-05	1.634E-04	
Probable ATP-dependent RNA helicase DDX6	DDX6	No	54		1.589E-04		4.101E-05
Probable E3 ubiquitin-protein ligase MYCBP2	MYCBP2	No	510	1.154E-05		6.416E-05	
Probable ergosterol biosynthetic protein 28	C14orf1	No	16	5.406E-04	4.955E-04		
Probable phospholipid-transporting ATPase IIA	ATP9A	No	119		2.430E-04	4.355E-04	
Probable ubiquitin carboxyl-terminal hydrolase FAF-X	USP9X	No	292	3.154E-05	2.672E-05		9.658E-06
Procollagen galactosyltransferase 1	COLGALT1	No	72	1.308E-04			5.603E-05
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	No	61			2.019E-04	2.313E-04
Prostaglandin F2 receptor negative regulator	PTGFRN	No	99	3.281E-04	8.401E-05	3.819E-04	
Prostasin	PRSS8	No	36	3.131E-04	1.395E-04	3.182E-04	
Proteasomal ubiquitin receptor ADRM1	ADRM1	No	42	4.621E-04	7.045E-05	4.641E-04	
Proteasome subunit alpha type-1	PSMA1	No	30		1.210E-04		7.249E-05
Proteasome subunit alpha type-6	PSMA6	No	27		6.517E-05		7.985E-05
Proteasome subunit alpha type-7	PSMA7	No	28		1.293E-04		1.597E-04
Proteasome subunit beta type-2	PSMB2	Yes	23		1.020E-04		1.206E-04
Proteasome subunit beta type-5	PSMB5	Yes	28		1.509E-04		1.675E-04
Proteasome-associated protein ECM29 homolog	KIAA0368	No	204	1.200E-04	1.015E-04		3.764E-05
Protein arginine N-methyltransferase 5	PRMT5	No	73	9.616E-05	5.456E-05		1.369E-04
Protein disulfide-isomerase A4	PDIA4	Yes	73		1.353E-04		3.848E-05
Protein disulfide-isomerase TMX3	TMX3	No	52			1.314E-04	1.555E-04

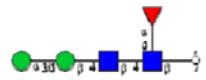
Protein dopey-2	DOPEY2	No	258		1.945E-05	4.707E-05	
Protein EFR3 homolog A	EFR3A	No	93	8.922E-05			3.614E-05
Protein ELYS	AHCTF1	No	253		2.517E-05		4.211E-05
Protein FAM134C	FAM134C	No	51	3.068E-04		1.785E-04	3.032E-04
Protein FAM83H	FAM83H	No	127	4.223E-04	4.550E-05	2.340E-04	
Protein ITFG3	ITFG3	Yes	60		5.907E-05	8.754E-05	
Protein jagunal homolog 1	JAGN1	No	21		6.006E-04		1.537E-03
Protein kinase C and casein kinase substrate in neurons protein 3	PACSIN3	Yes	48	1.594E-04	1.123E-04		
Protein LAP2	ERBB2IP	Yes	158		1.135E-05		3.161E-05
Protein lifeguard 3	TMBIM1	No	35	8.449E-04		4.636E-04	1.578E-04
Protein NipSnap homolog 1	NIPSNAP1	No	33		4.288E-04	8.387E-04	
Protein odr-4 homolog	ODR4	No	51		2.253E-04		6.317E-05
Protein PRRC2A	PRRC2A	No	229		1.476E-05		1.356E-05
Protein RER1	RER1	No	23		2.242E-04		1.803E-04
Protein RRP5 homolog	PDCD11	No	209	3.180E-05	3.518E-05		
Protein S100-A14	S100A14	Yes	12	6.816E-03	1.515E-03	9.199E-03	
Protein S100-A16	S100A16	No	12	5.488E-03	2.535E-03	9.914E-03	
Protein SCO2 homolog, mitochondrial	SCO2	No	30		9.389E-05	4.438E-04	
Protein transport protein Sec24C	SEC24C	No	118			4.313E-05	3.118E-05
Protein transport protein Sec61 subunit gamma	SEC61G	Yes	8	6.786E-03	2.113E-03		
Protein YIF1B	YIF1B	No	34	3.755E-04	3.091E-04		3.863E-04
Protein YIPF6	YIPF6	No	26		4.314E-04	6.191E-04	
Protein-glutamine gamma-glutamyltransferase K	TGM1	Yes	90	9.444E-05		1.872E-04	
Proteolipid protein 2	PLP2	Yes	17	1.732E-03			9.084E-04
Puromycin-sensitive aminopeptidase	NPEPPS	No	103		8.035E-05		4.821E-05
Putative ATP-dependent RNA helicase DHX30	DHX30	No	134			8.203E-05	7.758E-05
Putative deoxyribose-phosphate aldolase	DERA	No	35	1.994E-04	1.159E-04		3.704E-04
Putative helicase MOV-10	MOV10	No	114		9.176E-05		4.892E-05
Putative sodium-coupled neutral amino acid transporter 10	SLC38A10	No	120		1.037E-04	8.667E-05	
Pyridoxal-dependent decarboxylase domain-containing protein 1	PDXDC1	No	87		7.836E-05	1.210E-04	
Ragulator complex protein LAMTOR2	LAMTOR2	No	14		2.041E-04		1.986E-04
Ras GTPase-activating protein-binding protein 1	G3BP1	No	52	1.777E-04	2.597E-04		
Ras GTPase-activating-like protein IQGAP3	IQGAP3	No	185		1.016E-05		6.476E-05
Ras-related protein Rab-12	RAB12	No	27		1.156E-03	1.162E-03	
Ras-related protein Rab-18	RAB18	No	23	1.351E-03	1.018E-03		3.774E-03
Ras-related protein Rab-23	RAB23	No	27		6.993E-05		1.845E-04
Ras-related protein Rab-31	RAB31	Yes	22		6.730E-04		2.788E-04
Ras-related protein Rap-2b	RAP2B	No	21		1.068E-03		6.647E-04
Regulator of nonsense transcripts 1	UPF1	No	124		8.827E-05		6.625E-05
Regulatory-associated protein of mTOR	RPTOR	No	149	2.981E-05		4.372E-05	
Renin receptor	ATP6AP2	No	39	2.043E-04	1.180E-04		
Retinol dehydrogenase 13	RDH13	No	36		2.212E-04	5.129E-04	
Retinol dehydrogenase 14	RDH14	No	37		1.354E-04	2.182E-04	
Rho guanine nucleotide exchange factor 1	ARHGEF1	No	102		4.091E-05		7.002E-05
Rho guanine nucleotide exchange factor 2	ARHGEF2	No	112			1.220E-04	1.144E-04
Ribosomal RNA processing protein 1 homolog A	RRP1	No	53		7.875E-05	1.586E-04	
Ribosome biogenesis protein BMS1 homolog	BMS1	No	146		2.814E-05	6.604E-05	
Ribosome biogenesis protein BOP1	BOP1	Yes	84		2.749E-05		6.577E-05
Ribosome biogenesis protein BRX1 homolog	BRIX1	No	41		1.707E-04		9.965E-05
Ribosome biogenesis regulatory protein homolog	RRS1	No	41		1.989E-04		2.632E-04
RNA-binding protein 28	RBM28	No	86		1.019E-04	3.043E-04	
RNA-binding protein 39	RBM39	No	59		9.977E-05	2.270E-04	
RNA-binding protein 4	RBM4	No	40	2.449E-04	4.553E-05	1.675E-04	
RNA-binding protein Raly	RALY	No	32		1.204E-04		2.299E-04
rRNA 2'-O-methyltransferase fibrillarin	FBL	Yes	34	2.160E-04	4.149E-04		1.147E-03
Saccharopine dehydrogenase-like oxidoreductase	SCCPDH	No	47		2.176E-04		3.011E-04
Selenoprotein T	SELT	No	22		1.660E-04	4.889E-04	
Sentrin-specific protease 3	SENP3	No	65		4.445E-05		1.637E-04
Sequestosome-1	SQSTM1	No	48			4.990E-04	2.559E-04
Serine hydroxymethyltransferase, mitochondrial	SHMT2	No	56	3.187E-04	1.064E-04		9.735E-05
Serine palmitoyltransferase 2	SPTLC2	No	63	1.298E-04	1.985E-04	3.829E-04	
Serine/arginine-rich splicing factor 3	SRSF3	No	19		7.139E-04		5.829E-04
Serine/arginine-rich splicing factor 6	SRSF6	No	40		9.186E-05	7.382E-04	
Serine/arginine-rich splicing factor 7	SRSF7	No	27		4.589E-04	4.662E-04	
Serine/threonine-protein kinase mTOR	MTOR	No	289	1.050E-04	2.250E-05	7.055E-05	
Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	PPP3CA	No	59	2.589E-04	1.772E-04	2.313E-04	
Serine--tRNA ligase, cytoplasmic	SARS	No	59	1.158E-04	1.342E-04		
Serpin H1	SERPINH1	No	46	9.941E-05	7.062E-04		2.444E-04
Serum albumin	ALB	No	69	4.827E-02	1.218E-03		2.072E-03
SH3 domain-binding protein 4	SH3BP4	Yes	107			1.629E-04	8.714E-05
Sideroflexin-2	SFXN2	No	36		1.903E-04		1.077E-04
Sideroflexin-3	SFXN3	No	36	5.267E-04	4.888E-04		1.232E-03

Sideroflexin-4	SFXN4	No	38		1.429E-04		5.829E-05
Signal peptidase complex catalytic subunit SEC11A	SEC11A	No	21	4.446E-04	1.163E-03		6.181E-04
Signal peptidase complex subunit 2	SPCS2	No	25	4.924E-04	7.159E-04		4.113E-04
Signal peptidase complex subunit 3	SPCS3	No	20	5.321E-04	1.493E-03		6.895E-04
Signal recognition particle subunit SRP72	SRP72	No	75		1.212E-04	2.533E-04	
Signal transducer and activator of transcription 1-alpha/beta	STAT1	Yes	87		5.961E-05		6.542E-05
Single-stranded DNA-binding protein, mitochondrial	SSBP1	No	17	7.883E-04		5.763E-04	1.927E-04
Sister chromatid cohesion protein PDS5 homolog A	PDS5A	No	151		6.670E-05		4.863E-05
Small integral membrane protein 1	SMIM1	No	9		3.676E-04	1.738E-03	
Small integral membrane protein 13	SMIM13	No	10		3.609E-04		7.144E-04
Small VCP/p97-interacting protein	SVIP	No	8		2.562E-04		5.027E-04
SNW domain-containing protein 1	SNW1	No	61		4.660E-05	4.317E-04	
Soluble calcium-activated nucleotidase 1	CANT1	No	45		2.105E-04	4.210E-04	
Solute carrier family 12 member 4	SLC12A4	No	121	7.111E-05			4.083E-05
Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1	Yes	54	8.338E-04		4.815E-04	1.120E-03
Solute carrier family 25 member 40	SLC25A40	No	38		1.316E-04		1.863E-04
Sorting and assembly machinery component 50 homolog	SAMM50	No	52	1.870E-04	4.562E-04		2.225E-04
Spectrin beta chain, non-erythrocytic 2	SPTBN2	No	271		1.074E-04	6.295E-04	
Sperm-specific antigen 2	SSFA2	No	138	3.300E-05	3.826E-05		4.941E-05
Sphingosine-1-phosphate phosphatase 1	SGPP1	No	49		7.393E-05		7.641E-05
Splicing factor 3B subunit 2	SF3B2	No	100		4.874E-05	1.882E-04	
Squalene monooxygenase	SQLE	No	64	1.760E-04	3.254E-04		
Squalene synthase	FDFT1	No	48	1.111E-03	2.858E-04		5.952E-05
Starch-binding domain-containing protein 1	STBD1	No	39		1.117E-04		1.107E-04
Stearoyl-CoA desaturase 5	SCD5	No	38		4.041E-04		7.472E-05
Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL	No	42	7.468E-04	5.059E-04		4.953E-04
Structural maintenance of chromosomes flexible hinge domain-containing protein 1	SMCHD1	No	226		2.164E-05		1.746E-05
Structural maintenance of chromosomes protein 3	SMC3	No	142		3.414E-05	3.970E-05	
Structural maintenance of chromosomes protein 4	SMC4	No	147		1.287E-05		5.335E-05
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	Yes	73		6.744E-05	1.644E-04	
Sulfhydryl oxidase 1	QSOX1	No	83			1.098E-04	3.994E-05
Sulfhydryl oxidase 2	QSOX2	No	78		4.749E-05		1.003E-04
Supervillin	SVIL	No	248	2.620E-04			2.175E-05
Suppressor of tumorigenicity 14 protein	ST14	Yes	95	2.959E-04	6.148E-05	1.276E-04	
Sushi domain-containing protein 2	SUSD2	No	90			5.292E-03	1.219E-04
Symplekin	SYMPK	No	141		1.900E-05	3.793E-05	
Synaptogyrin-3	SYNGR3	No	25		1.043E-04	8.284E-04	
Synaptojanin-2-binding protein	SYNJ2BP	No	16	5.098E-04	3.266E-03		6.156E-04
Synaptophysin-like protein 1	SYPL1	No	29	2.420E-03		3.715E-03	
Syndecan-1	SDC1	Yes	32	3.555E-04		9.293E-04	
Syntaxin-10	STX10	No	28	1.923E-04	1.470E-04		1.792E-04
Syntaxin-5	STX5	No	40		1.689E-04		2.378E-04
Syntaxin-6	STX6	No	29		2.992E-04	1.902E-04	
Syntaxin-binding protein 3	STXBP3	No	68		8.438E-05		1.058E-04
T-complex protein 1 subunit zeta	CCT6A	Yes	58	2.355E-04	7.621E-05		7.460E-05
Telomere-associated protein RIF1	RIF1	No	274		1.350E-05		2.796E-05
Tenascin	TNC	No	241	7.201E-05			2.196E-05
Tetratricopeptide repeat protein 37	TTC37	No	175	3.804E-05	1.868E-05		
Thioredoxin domain-containing protein 5	TXNDC5	No	48		4.746E-05		4.585E-05
THO complex subunit 4	ALYREF	No	27		6.313E-04		7.816E-04
Threonine--tRNA ligase, cytoplasmic	TARS	Yes	83		6.802E-05		1.371E-04
Thrombospondin-1	THBS1	Yes	129	1.851E-04	2.338E-04		1.608E-04
Thyroid hormone receptor-associated protein 3	THRAP3	No	109	7.047E-05	2.591E-05	3.284E-04	
Tight junction protein ZO-1	TJP1	No	195		9.172E-06	8.303E-05	
Tight junction protein ZO-2	TJP2	Yes	134	5.532E-05	1.723E-05	1.416E-04	
Transcriptional repressor p66-alpha	GATAD2A	No	68		6.271E-05		6.050E-05
Transformation/transcription domain-associated protein	TRRAP	Yes	438		2.611E-05	1.223E-05	
Transformer-2 protein homolog beta	TRA2B	Yes	34		1.916E-04	7.460E-04	
Transient receptor potential cation channel subfamily M member 4	TRPM4	No	134		1.172E-04	3.980E-05	
Translation initiation factor eIF-2B subunit delta	EIF2B4	No	58	3.521E-04		6.430E-04	1.240E-04
Translation initiation factor eIF-2B subunit epsilon	EIF2B5	No	80	8.793E-05		6.702E-05	
Translocation protein SEC62	SEC62	No	46		7.978E-05		4.964E-05
Translocon-associated protein subunit gamma	SSR3	No	21		1.205E-03		2.284E-03
Transmembrane 9 superfamily member 1	TM9SF1	Yes	69		4.193E-04	1.373E-04	
Transmembrane emp24 domain-containing protein 5	TMED5	No	26	2.769E-04	1.447E-04		1.723E-04
Transmembrane protein 106B	TMEM106B	No	31		2.926E-04	6.604E-04	
Transmembrane protein 126A	TMEM126A	No	22		2.305E-04		4.093E-04
Transmembrane protein 134	TMEM134	No	22		1.269E-04	3.052E-04	
Transmembrane protein 14C	TMEM14C	No	12		7.185E-04		9.723E-04
Transmembrane protein 161A	TMEM161A	No	54		2.296E-04		1.163E-04

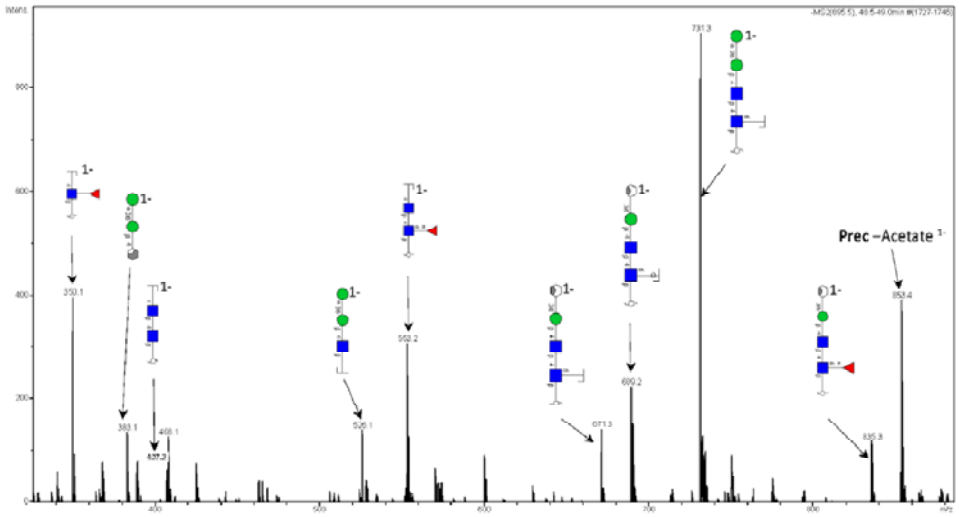
Transmembrane protein 179B	TMEM179B	No	24	2.895E-04		4.369E-04	
Transmembrane protein 192	TMEM192	No	31		7.566E-05	2.202E-04	
Transmembrane protein 2	TMEM2	No	154	2.396E-04			5.663E-05
Transmembrane protein 33	TMEM33	No	28	1.467E-03	4.409E-04		1.122E-03
Transmembrane protein 41A	TMEM41A	No	30		6.278E-05		1.880E-04
Tricarboxylate transport protein, mitochondrial	SLC25A1	No	34		4.679E-04		4.984E-04
Trifunctional enzyme subunit alpha, mitochondrial	HADHA	Yes	83	2.053E-04	8.308E-04	4.696E-04	
Trifunctional enzyme subunit beta, mitochondrial	HADHB	No	51		2.559E-04	1.019E-04	
tRNA (cytosine(34)-C(5))-methyltransferase	NSUN2	No	86		3.668E-05		4.372E-05
tRNA-splicing ligase RtcB homolog	RTCB	No	55	8.228E-05		2.851E-04	
Trophoblast glycoprotein	TPBG	Yes	46		7.050E-04		3.141E-04
Tropomodulin-3	TMOD3	No	40	8.573E-04		2.718E-04	
Two pore calcium channel protein 1	TPCN1	Yes	94		3.542E-05		6.104E-05
Tyrosine-protein kinase FRK	FRK	Yes	58		6.456E-05	9.343E-05	
U3 small nucleolar RNA-associated protein 18 homolog	UTP18	No	62		5.961E-05	2.406E-04	
U4/U6 small nuclear ribonucleoprotein Prp31	PRPF31	No	55		5.005E-05	1.439E-04	
U4/U6.U5 tri-snRNP-associated protein 1	SART1	No	90		2.563E-05	1.627E-04	
U5 small nuclear ribonucleoprotein 40 kDa protein	SNRNP40	No	39		8.982E-05	3.052E-04	
UAP56-interacting factor	FYTTD1	No	36		7.611E-05		6.229E-05
Ubiquitin carboxyl-terminal hydrolase 10	USP10	Yes	87		9.630E-05		3.018E-05
Ubiquitin conjugation factor E4 A	UBE4A	No	123		1.502E-04		4.102E-05
Ubiquitin-associated domain-containing protein 2	UBAC2	No	39		6.811E-05		1.164E-04
Ubiquitin-like protein 3	UBL3	No	13		4.137E-04	6.042E-04	
Ubiquitin-protein ligase E3C	UBE3C	No	124	2.370E-04			2.292E-05
Uncharacterized protein C10orf35	C10orf35	No	13		4.325E-04	8.845E-04	
Uncharacterized protein C19orf52	C19orf52	No	29		1.346E-04		1.318E-04
Uncharacterized protein C2orf47, mitochondrial	C2orf47	No	33		2.739E-04		1.198E-04
Uncharacterized protein KIAA2013	KIAA2013	No	69		1.871E-04	2.463E-04	
Unconventional myosin-Id	MYO1D	No	116	9.009E-04	1.498E-04	2.386E-04	
Unconventional myosin-Va	MYO5A	No	215	3.620E-04	1.443E-04	2.250E-04	
Unconventional myosin-Vb	MYO5B	Yes	214	1.424E-04		1.171E-04	
Unconventional myosin-Vc	MYO5C	No	203		2.065E-04	2.813E-04	
Unconventional myosin-XVIIIa	MYO18A	No	233	1.116E-04		3.778E-04	
Vacuolar ATPase assembly integral membrane protein VMA21	VMA21	No	11	6.692E-04	7.762E-04		9.073E-04
Vacuolar protein sorting-associated protein 45	VPS45	No	65		3.084E-04	2.330E-04	
Very-long-chain (3R)-3-hydroxyacyl-[acyl-carrier protein] dehydratase 3	PTPLAD1	No	43		1.173E-03		1.132E-03
Vesicle transport protein GOT1B	GOLT1B	No	15	3.011E-04	2.760E-03		3.827E-03
Vesicle transport protein USE1	USE1	No	29		1.738E-04		1.517E-04
Vesicle transport through interaction with t-SNAREs homolog 1B	VTI1B	No	27		2.588E-04	6.802E-04	
Vesicle-associated membrane protein 4	VAMP4	No	16			8.521E-04	2.442E-04
Vigilin	HDLBP	No	141		4.584E-05		1.958E-05
Vimentin	VIM	Yes	54	4.801E-03		3.737E-04	5.699E-02
VIP36-like protein	LMAN2L	No	40		1.172E-04		3.642E-04
Vitamin K-dependent gamma-carboxylase	GGCX	Yes	88		4.789E-05		6.277E-05
Voltage-dependent anion-selective channel protein 3	VDAC3	No	31	9.157E-04	5.167E-04		5.849E-04
von Willebrand factor A domain-containing protein 8	VWA8	No	215		1.493E-05	4.298E-05	
V-type proton ATPase 116 kDa subunit a isoform 2	ATP6V0A2	Yes	98		1.585E-04		5.233E-05
V-type proton ATPase subunit d 1	ATP6V0D1	No	40	4.063E-04	3.318E-04		1.953E-04
V-type proton ATPase subunit E 1	ATP6V1E1	No	26		1.630E-04		8.764E-05
V-type proton ATPase subunit G 1	ATP6V1G1	No	14		1.692E-04	1.736E-03	
V-type proton ATPase subunit H	ATP6V1H	No	56		4.963E-05	2.958E-04	
V-type proton ATPase subunit S1	ATP6AP1	Yes	52		1.981E-04		5.158E-05
WD repeat-containing protein 5	WDR5	Yes	37		4.962E-05	2.190E-04	
WD repeat-containing protein 74	WDR74	No	42		4.305E-05		6.447E-05
WW domain-containing oxidoreductase	WWOX	Yes	47		1.659E-04	3.807E-04	
Zinc finger RNA-binding protein	ZFR	No	117		2.275E-05	1.140E-04	
Zinc transporter SLC39A7	SLC39A7	No	50		4.786E-04	1.080E-03	

APPENDIX 3: Supplementary data from Publication II

Glycan No 1
Type = Paucimannose
Precursor = *m/z* 895.4^{1-}
[M-H]$^{1-}$ = 895.4Da
LC retention time = 49 min

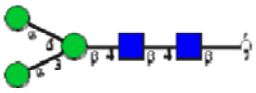


Supplemental Figure S1

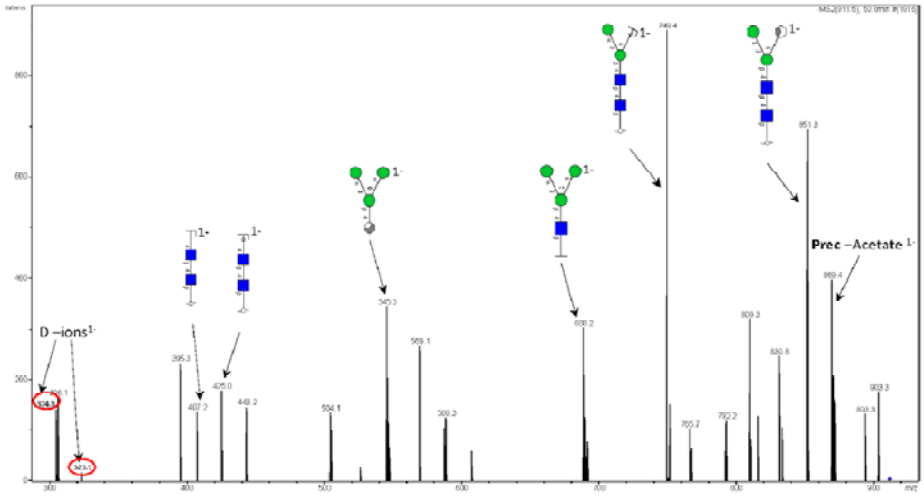


Positive match to MS2 spectrum in UnicarbKB

Glycan No 2
Type = Paucimannose
Precursor = *m/z* 911.3^{1-}
[M-H]$^{1-}$ = 911.3Da
LC retention time = 50 min

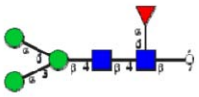


Supplemental Figure S1

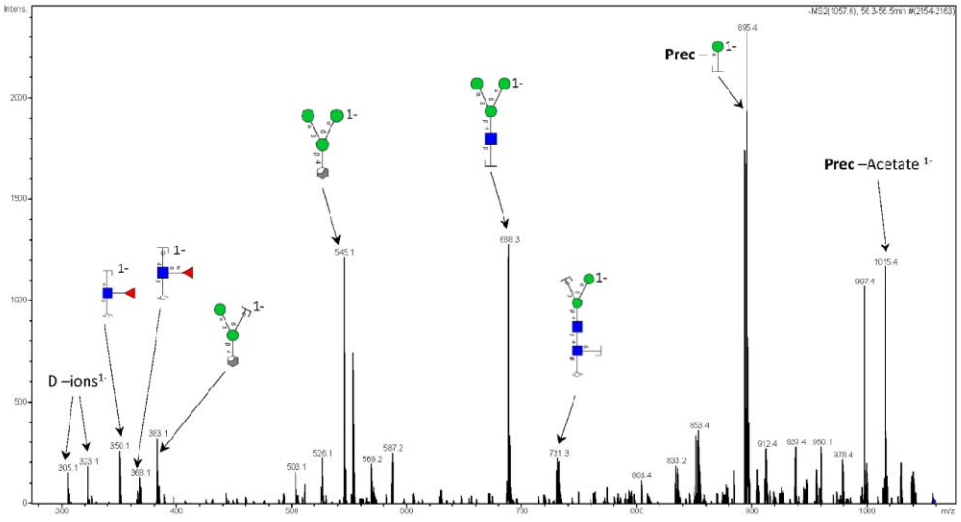


Positive match to MS2 spectrum in UnicarbKB

Glycan No 3
Type = Paucimannose
Precursor = *m/z* 1057.4^{1-}
[M-H]$^{1-}$ = 1057.4Da
LC retention time = 56.5 min

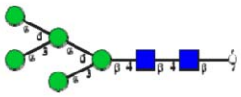


Supplemental Figure S1

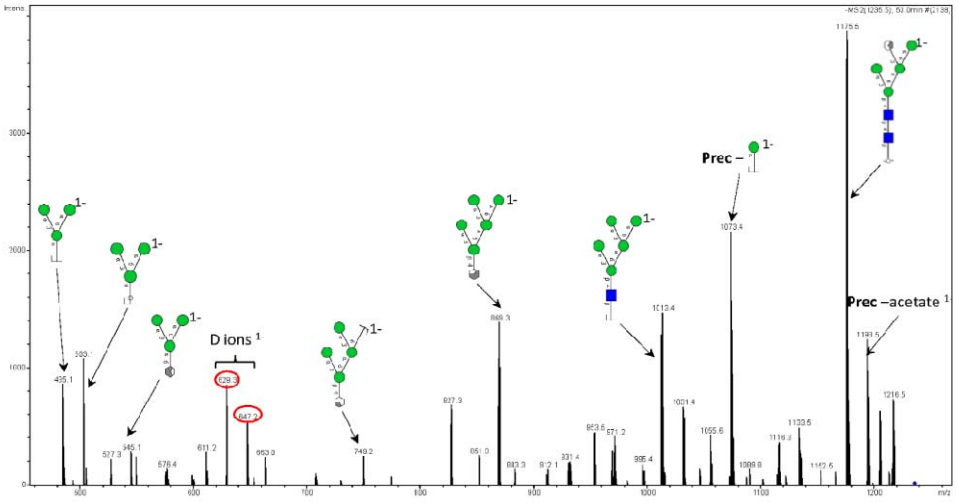


Positive match to MS2 spectrum in UnicarbKB

Glycan No 4
Type = High Mannose
Precursor = *m/z* 1235.5^{1-}
[M-H]$^{1-}$ = 1235.5 Da
LC retention time = 53 min

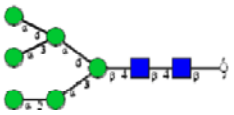


Supplemental Figure S1

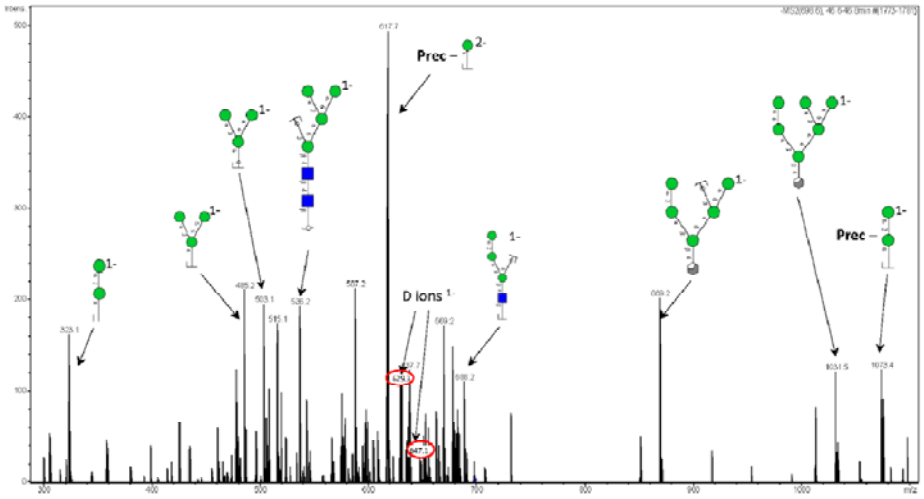


Positive match to MS2 spectrum in UnicarbKB

Glycan No 5
Type = High Mannose
Precursor = *m/z* 698.3^{2-}
$[M-H]^{1-}$ = 1397.4Da
LC retention time = 46.7 min

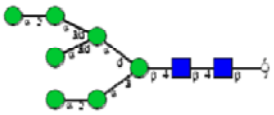


Supplemental Figure S1

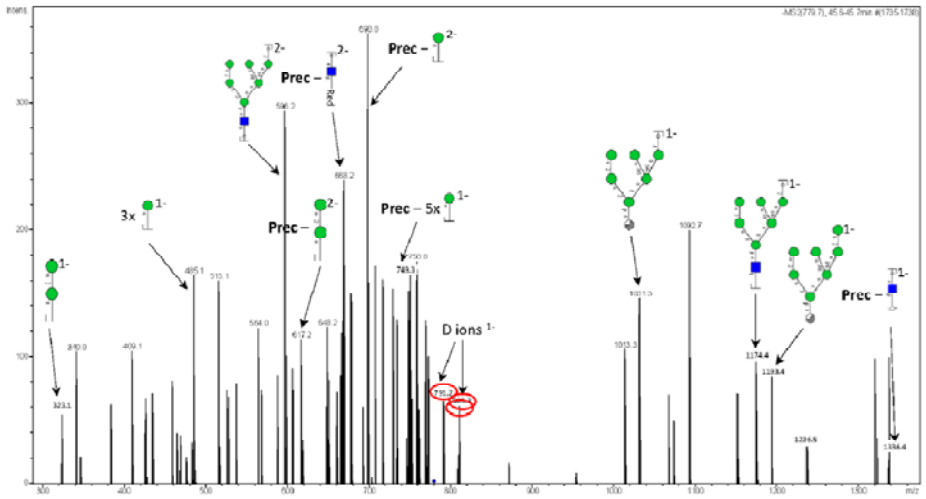


Positive match to MS2 spectrum in UnicarbKB

Glycan No 6a
Type = High Mannose
Precursor = *m/z* 779.3^{2-}
$[M-H]^{1-}$ = 1559.6 Da
LC retention time = 45.7 min

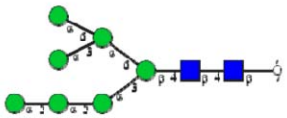


Supplemental Figure S1

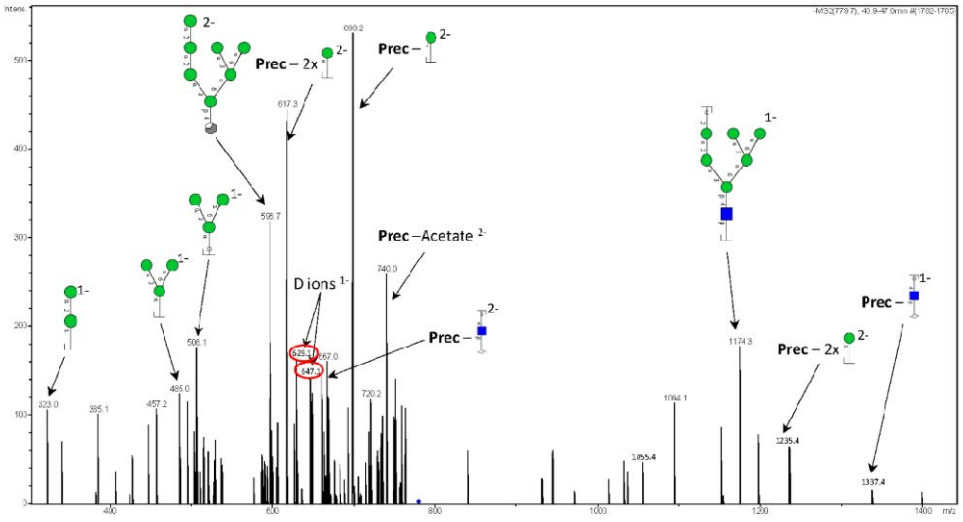


Positive match to MS2 spectrum in UnicarbKB

Glycan No 6b
Type = High Mannose
Precursor = *m/z* 779.3^{2-}
$[M-H]^{1-}$ = 1559.6 Da
LC retention time = 47 min

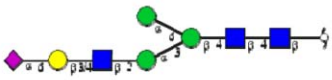


Supplemental Figure S1

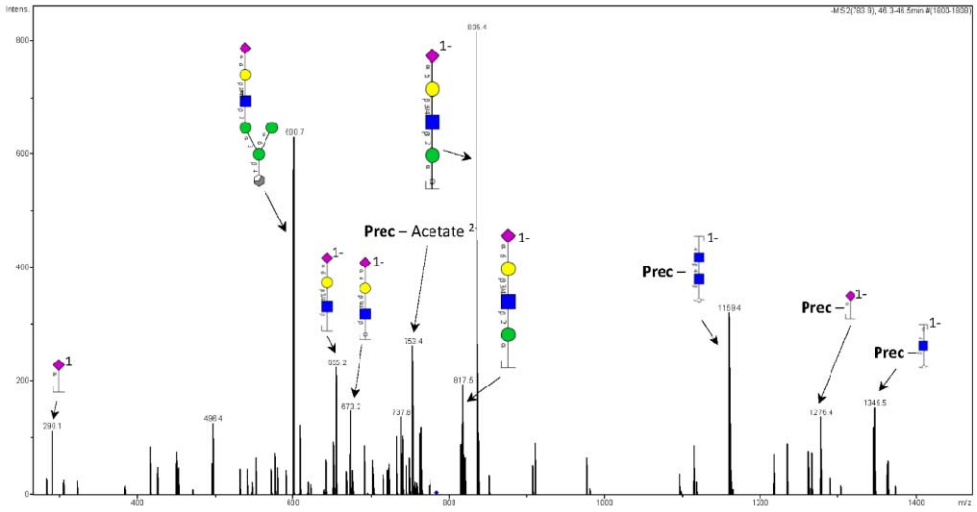


No match to MS2 spectrum in UnicarbKB

Glycan No 7
Type = Complex
Precursor = *m/z* 783.3^{2-}
$[M-H]^{1-}$ = 1567.6 Da
LC retention time = 46.4 min

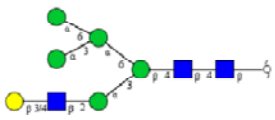


Supplemental Figure S1

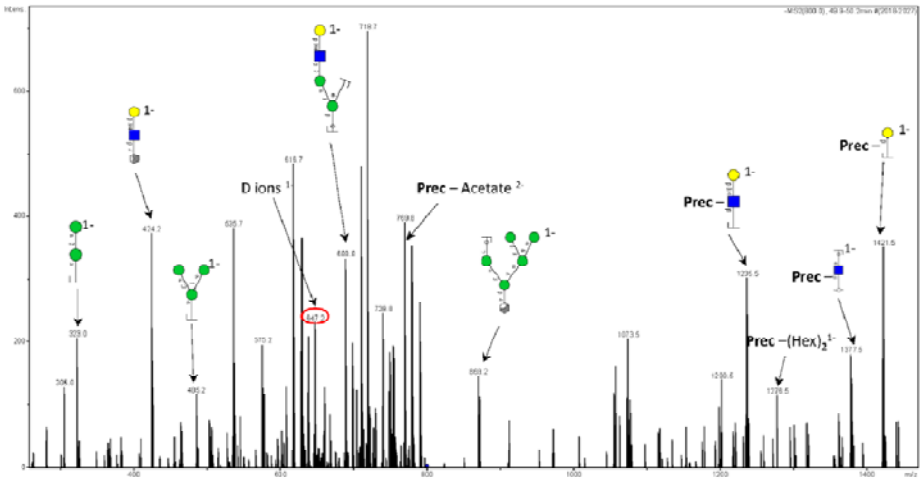


No match to MS2 spectrum in UnicarbKB

Glycan No 8
Type = Hybrid
Precursor = *m/z* 799.8^{2-}
[M-H]$^{1-}$ = 1600.6 Da
LC retention time = 50.1 min

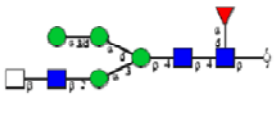


Supplemental Figure S1

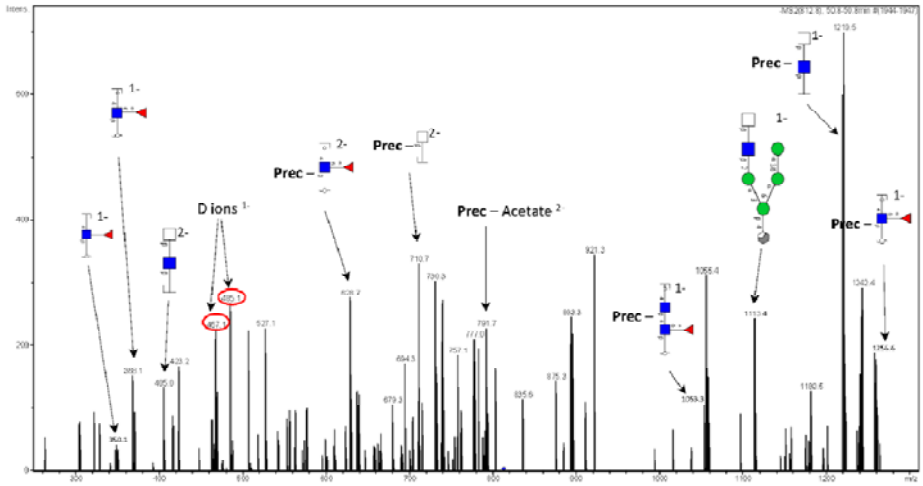


No match to MS2 spectrum in UnicarbKB

Glycan No 9
Type = Hybrid
Precursor = *m/z* 812.3^{2-}
[M-H]$^{1-}$ = 1625.6 Da
LC retention time = 50.8 min

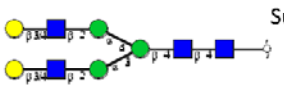


Supplemental Figure S1

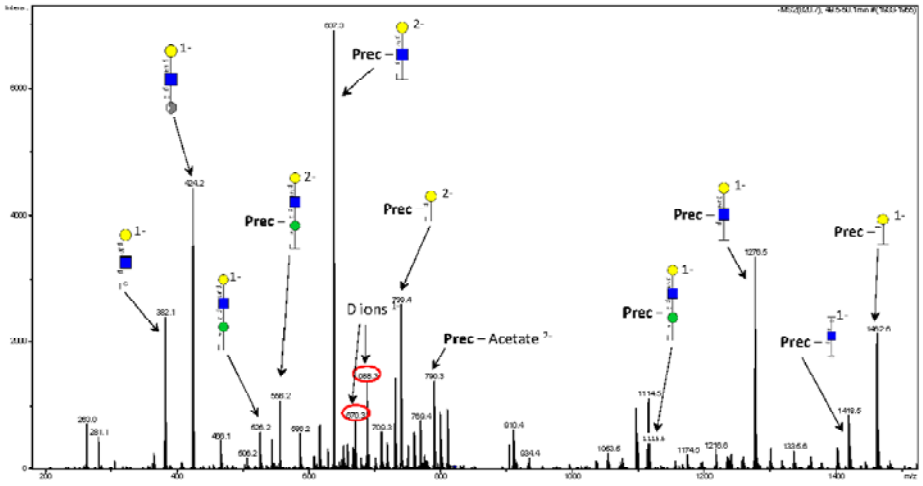


No match to MS2 spectrum in UnicarbKB

Glycan No 10
Type = Complex
Precursor = *m/z* 820.4^{2-}
[M-H]$^{1-}$ = 1641.8 Da
LC retention time = 49.5 min

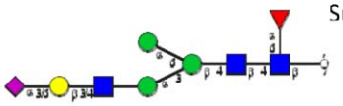


Supplemental Figure S1

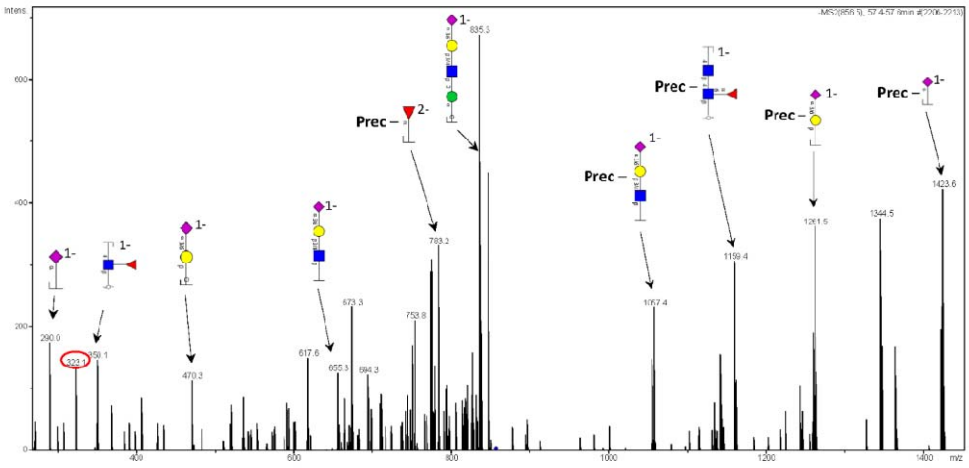


Positive match to MS2 spectrum in UnicarbKB

Glycan No 11a
Type = Complex
Precursor = *m/z* 856.3^{2-}
[M-H]$^{1-}$ = 1713.6 Da
LC retention time = 57.2 min



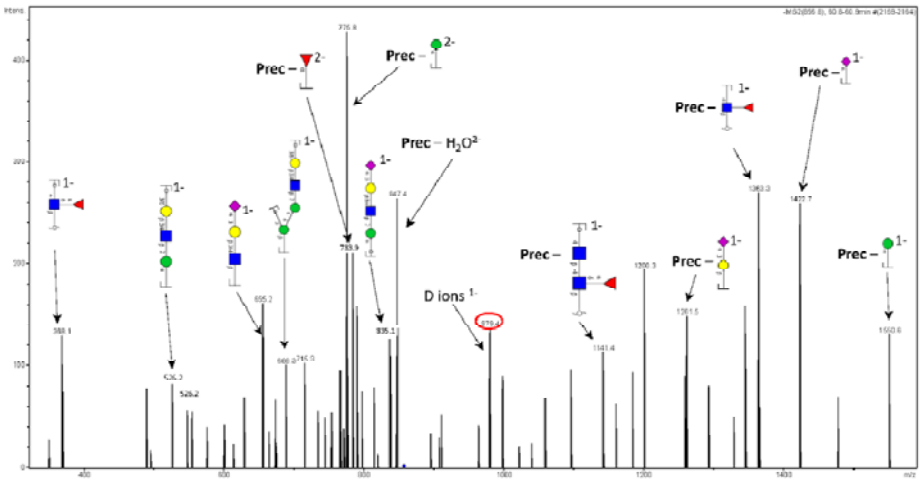
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

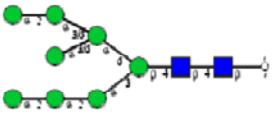
Glycan No 11b
Type = Complex
Precursor = *m/z* 856.3^{2-}
[M-H]$^{1-}$ = 1713.6 Da
LC retention time = 60.9 min

Supplemental Figure S1

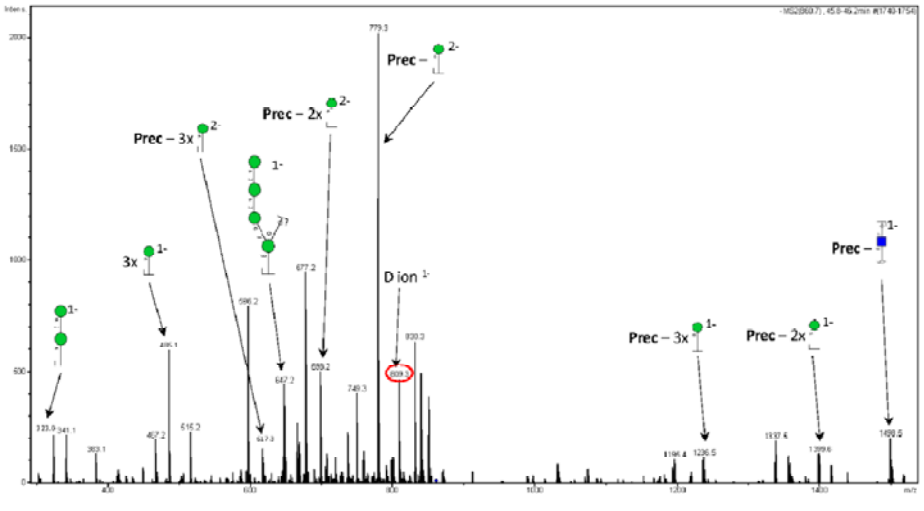


No match to MS2 spectrum in UnicarbKB

Glycan No 12
Type = High Mannose
Precursor = *m/z* 860.3^{2-}
[M-H]$^{1-}$ = 1721.6 Da
LC retention time = 45.9 min



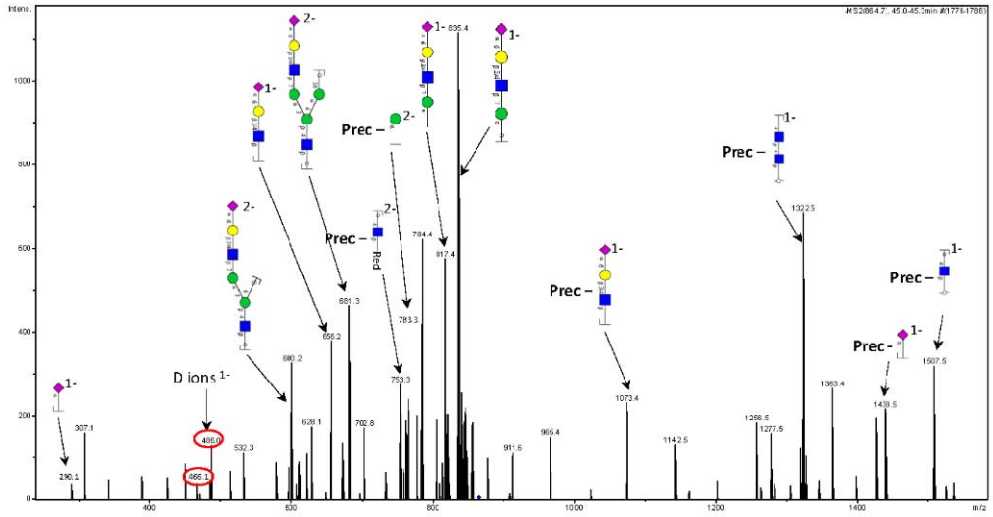
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

Glycan No 13a
Type = Hybrid
Precursor = *m/z* 864.4^{2-}
[M-H]$^{1-}$ = 1729.8 Da
LC retention time = 45.3 min

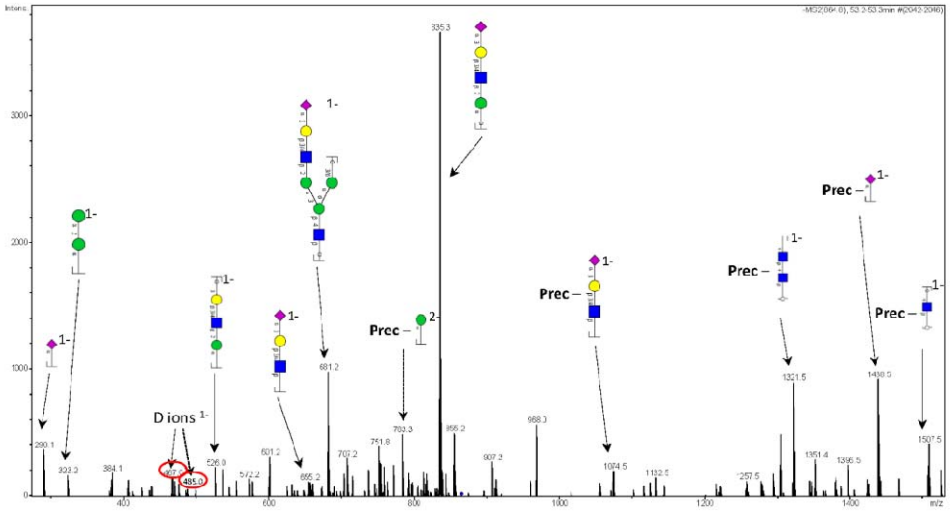
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

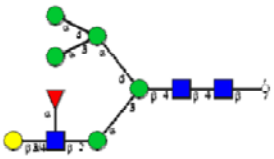
Glycan No 13b
Type = Hybrid
Precursor = *m/z* 864.4^{2-}
[M-H]$^{1-}$ = 1729.8 Da
LC retention time = 53.3 min

Supplemental Figure S1

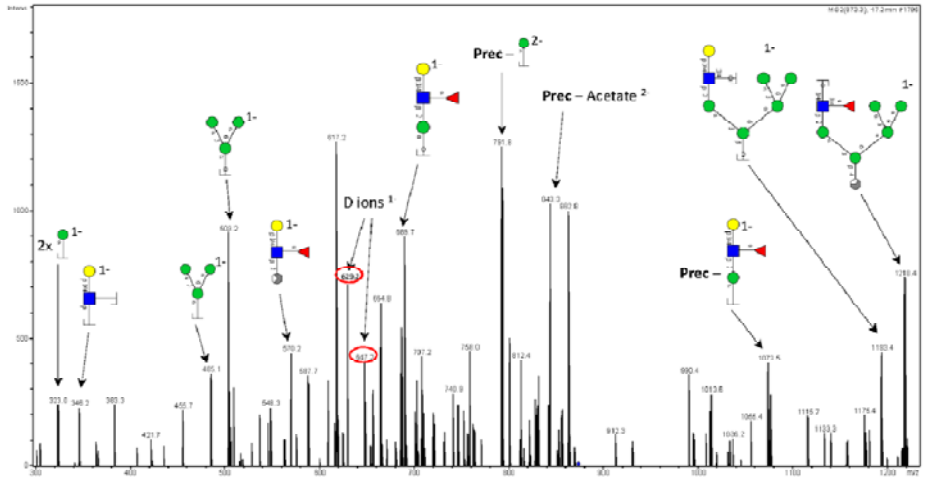


No match to MS2 spectrum in UnicarbKB

Glycan No 14a
Type = Hybrid
Precursor = *m/z* 872.9^{2-}
[M-H]$^{1-}$ = 1746.8 Da
LC retention time = 47.2 min

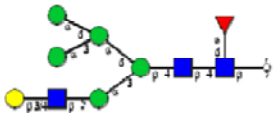


Supplemental Figure S1

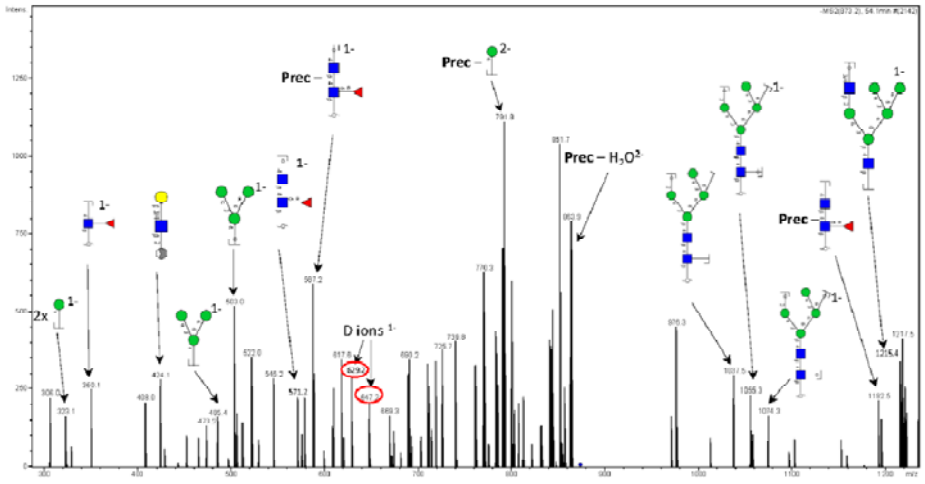


No match to MS2 spectrum in UnicarbKB

Glycan No 14b
Type = Hybrid
Precursor = *m/z* 872.9^{2-}
[M-H]$^{1-}$ = 1746.8 Da
LC retention time = 54.1 min

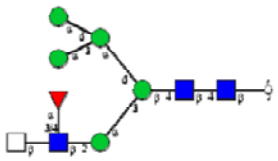


Supplemental Figure S1

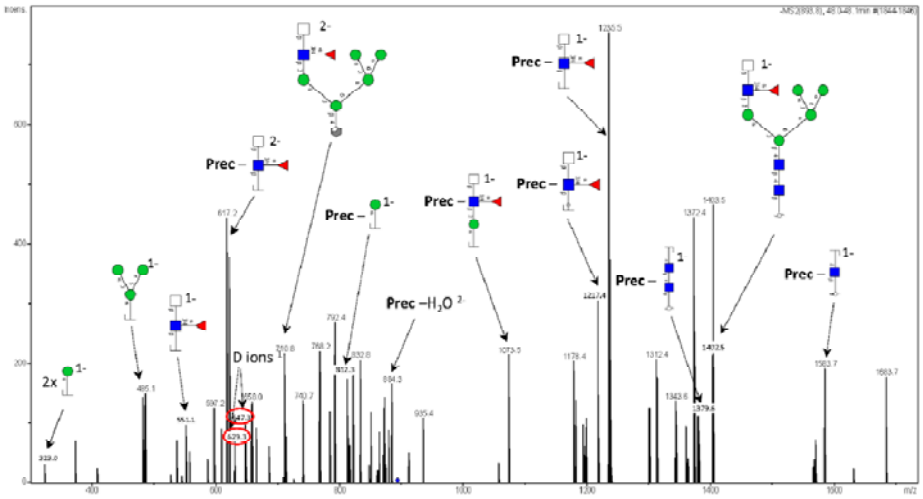


No match to MS2 spectrum in UnicarbKB

Glycan No 15a
Type = Hybrid
Precursor = *m/z* 893.3^{2-}
[M-H]$^{1-}$ = 1787.6 Da
LC retention time = 48 min

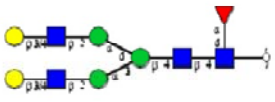


Supplemental Figure S1

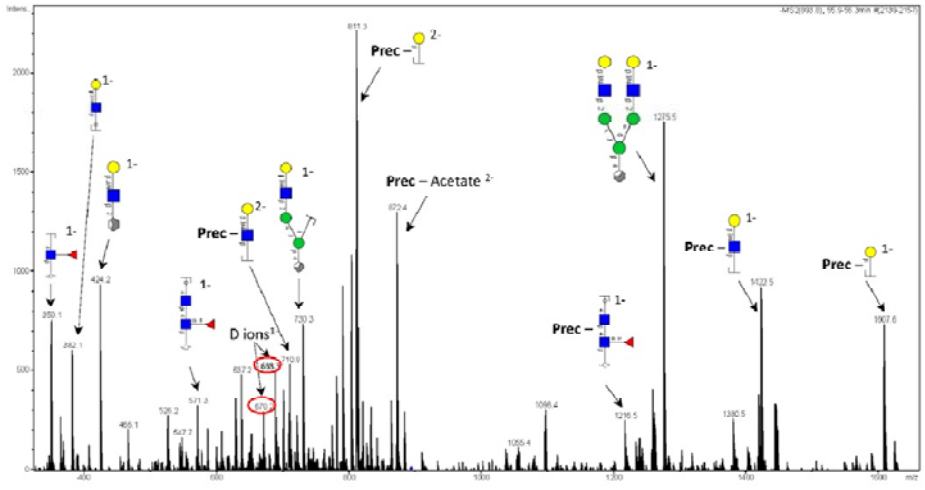


No match to MS2 spectrum in UnicarbKB

Glycan No 15b
Type = Complex
Precursor = *m/z* 893.3^{2-}
[M-H]$^{1-}$ = 1787.6 Da
LC retention time = 56.0 min

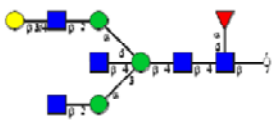


Supplemental Figure S1

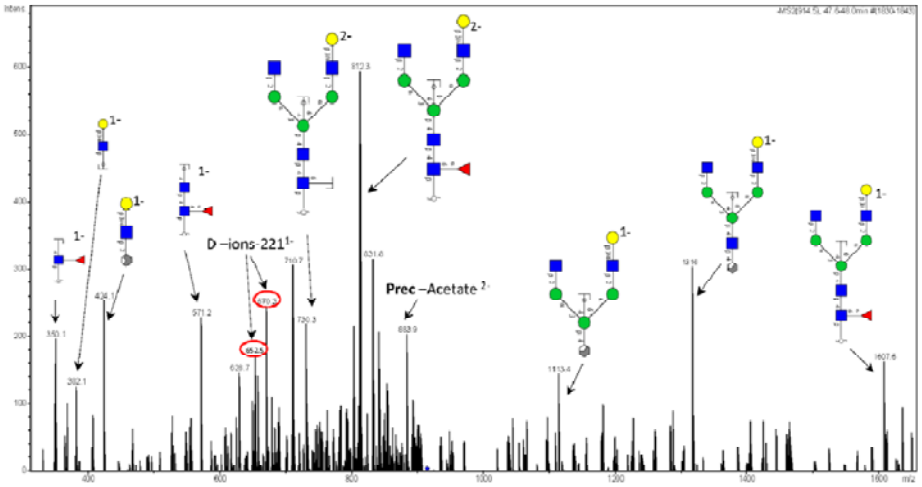


Positive match to MS2 spectrum in UnicarbKB

Glycan No 16a
Type = Complex
Precursor = *m/z* 913.9^{2-}
[M-H]$^{1-}$ = 1828.8Da
LC retention time = 48 min

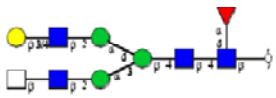


Supplemental Figure S1

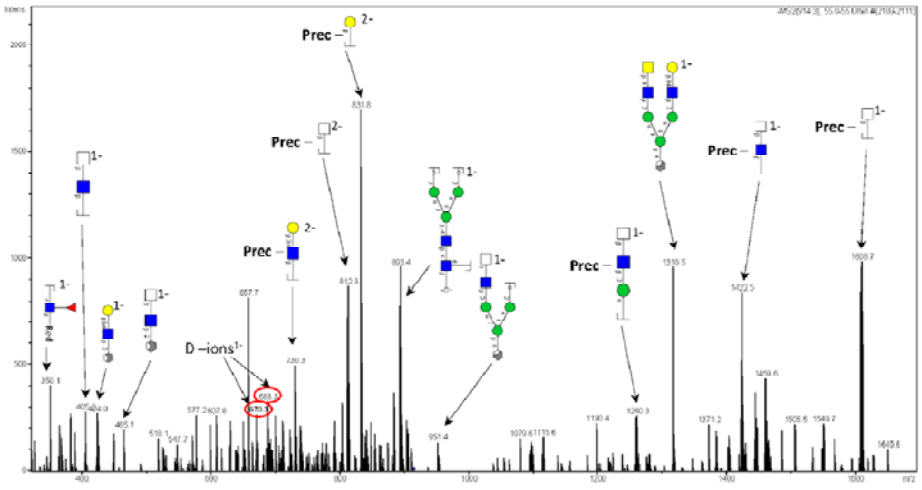


Positive match to MS2 spectrum in UnicarbKB

Glycan No 16b
Type = Complex
Precursor = *m/z* 913.9^{2-}
[M-H]$^{1-}$ = 1828.8Da
LC retention time = 55 min



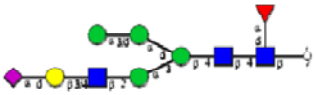
Supplemental Figure S1

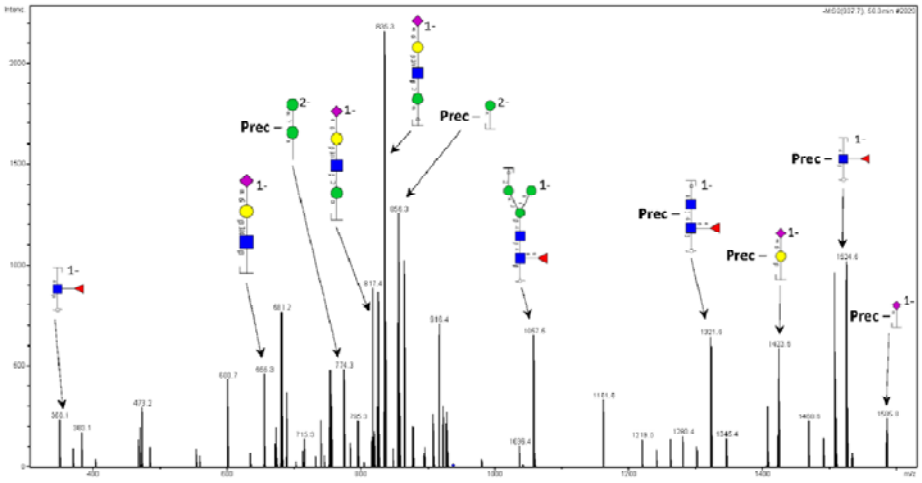


No match to MS2 spectrum in UnicarbKB

Glycan No 17a
Type = Hybrid
Precursor = *m/z* 937.3^{2-}
$[M-H]^{1-}$ = 1875.6 Da
LC retention time = 50.3 min

Supplemental Figure S1

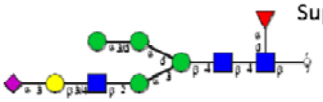


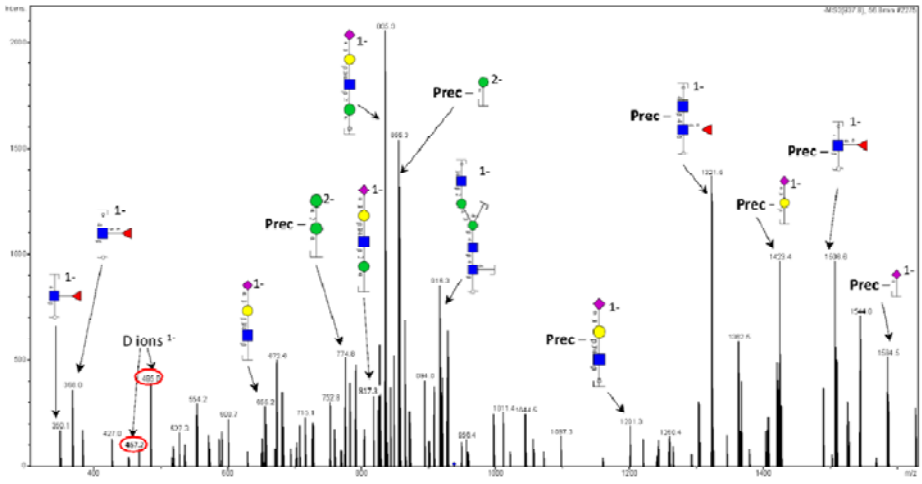


Positive match to MS2 spectrum in UnicarbKB

Glycan No 17b
Type = Hybrid
Precursor = *m/z* 937.3^{2-}
$[M-H]^{1-}$ = 1875.6 Da
LC retention time = 56.8 min

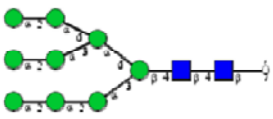
Supplemental Figure S1



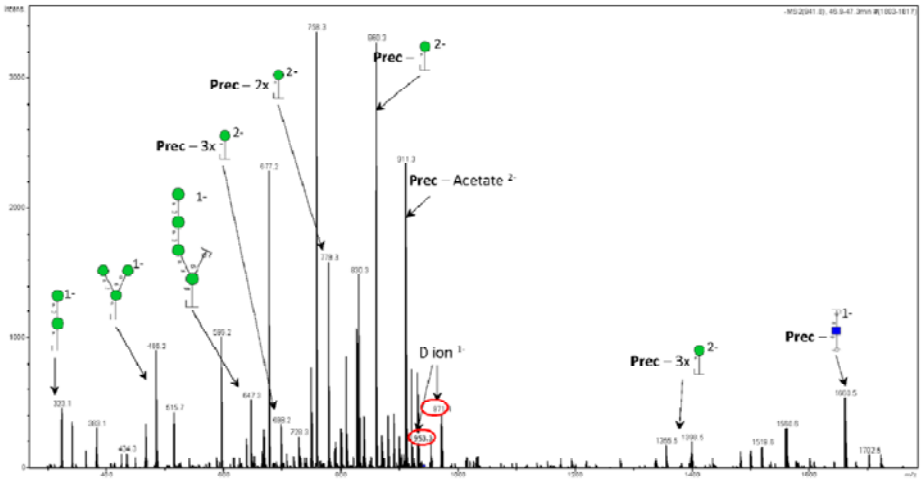


Positive match to MS2 spectrum in UnicarbKB

Glycan No 18
Type = High Mannose
Precursor = *m/z* 941.4^{2-}
[M-H]$^{1-}$ = 1883.8 Da
LC retention time = 47.0 min

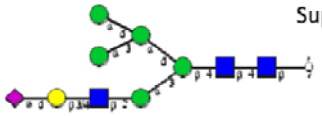


Supplemental Figure S1

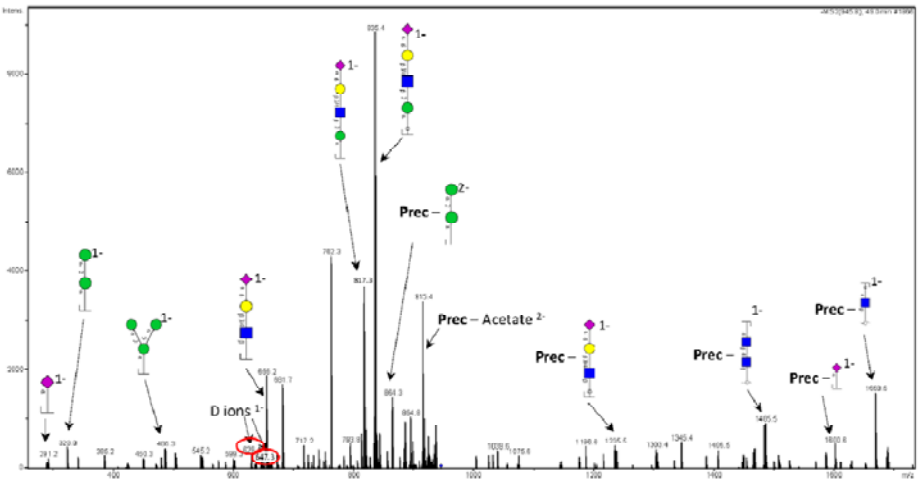


Positive match to MS2 spectrum in UnicarbKB

Glycan No 19a
Type = Hybrid
Precursor = *m/z* 945.4^{2-}
[M-H]$^{1-}$ = 1891.8 Da
LC retention time = 49.0 min

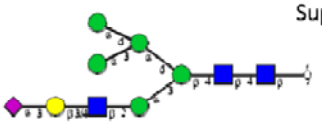


Supplemental Figure S1

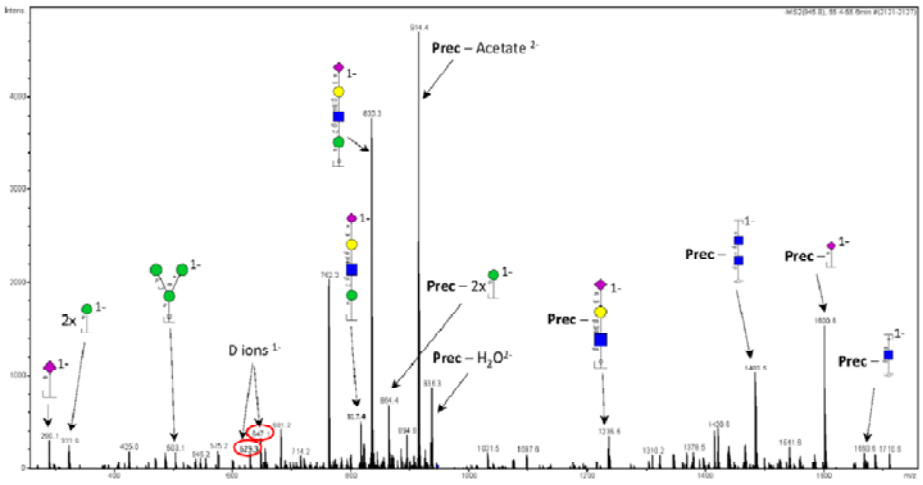


Positive match to MS2 spectrum in UnicarbKB

Glycan No 19b
Type = Hybrid
Precursor = *m/z* 945.4^{2-}
$[M-H]^{1-}$ = 1891.8 Da
LC retention time = 55.4 min

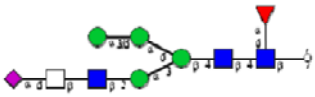


Supplemental Figure S1

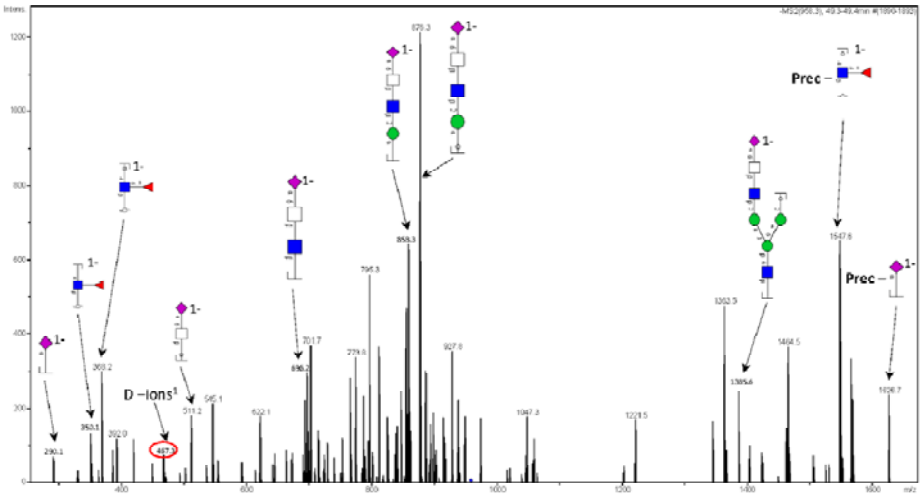


Positive match to MS2 spectrum in UnicarbKB

Glycan No 20
Type = Hybrid
Precursor = *m/z* 957.9^{2-}
$[M-H]^{1-}$ = 1916.8Da
LC retention time = 49.3 min

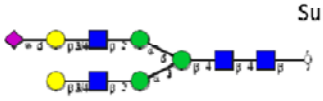


Supplemental Figure S1

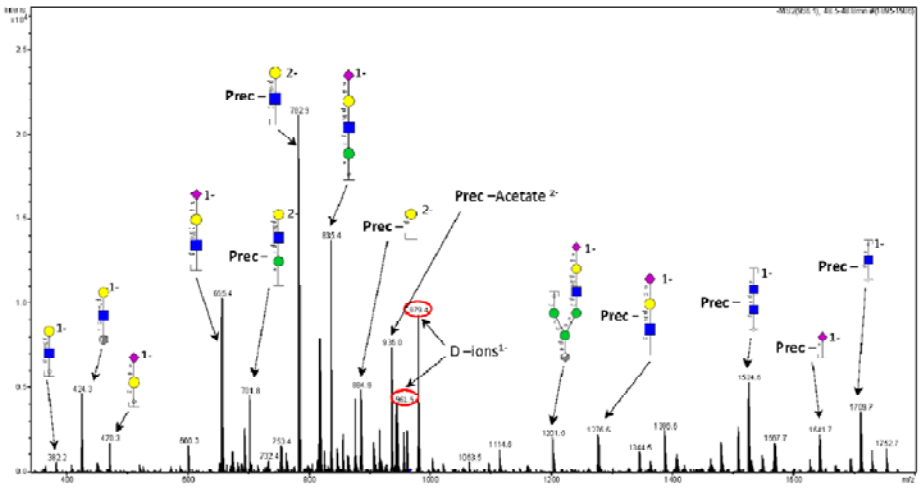


No match to MS2 spectrum in UnicarbKB

Glycan No 21a
Type – Complex
Precursor = *m/z* 965.9^{2-}
[M-H]$^{1-}$ = 1932.8Da
LC retention time = 48.5 min

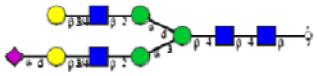


Supplemental Figure S1

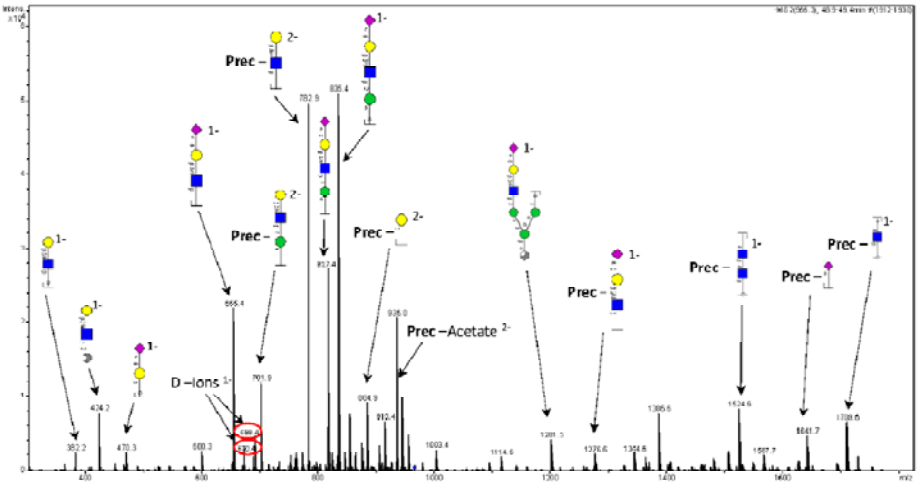


Positive match to MS2 spectrum in UnicarbKB

Glycan No 21b
Type = Complex
Precursor = *m/z* 965.9^{2-}
[M-H]$^{1-}$ = 1932.8Da
LC retention time = 49 min



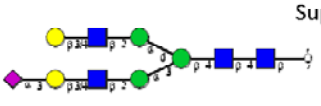
Supplemental Figure S1

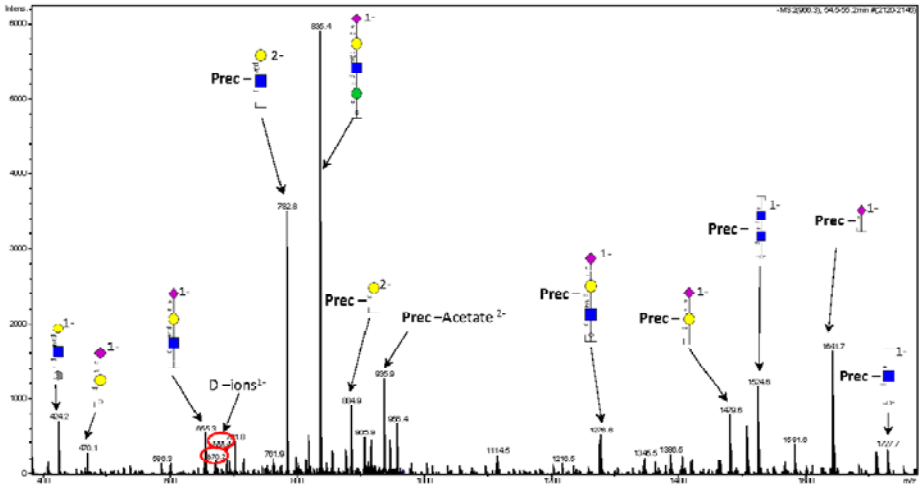


No match to MS2 spectrum in UnicarbKB

Supplemental Figure S1

Glycan No 21c
Type = Complex
Precursor = *m/z* 965.9^{2}
[M-H]$^{1-}$ = 1932.8Da
LC retention time = 54.5 min

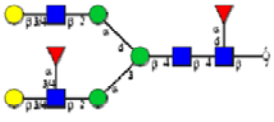


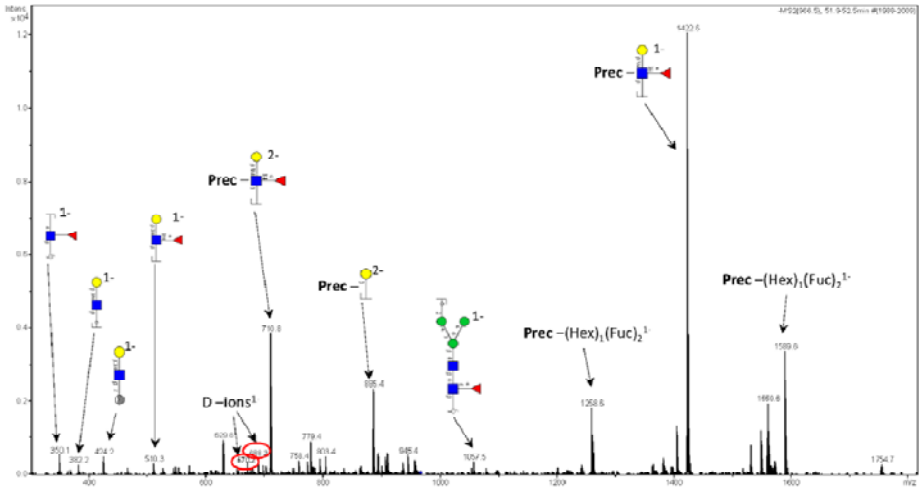


No match to MS2 spectrum in UnicarbKB

Supplemental Figure S1

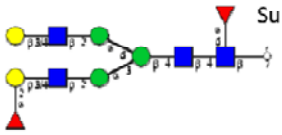
Glycan No 22a
Type = Complex
Precursor = *m/z* 966.4^{2-}
[M-H]$^{1-}$ = 1933.8Da
LC retention time = 52.2 min



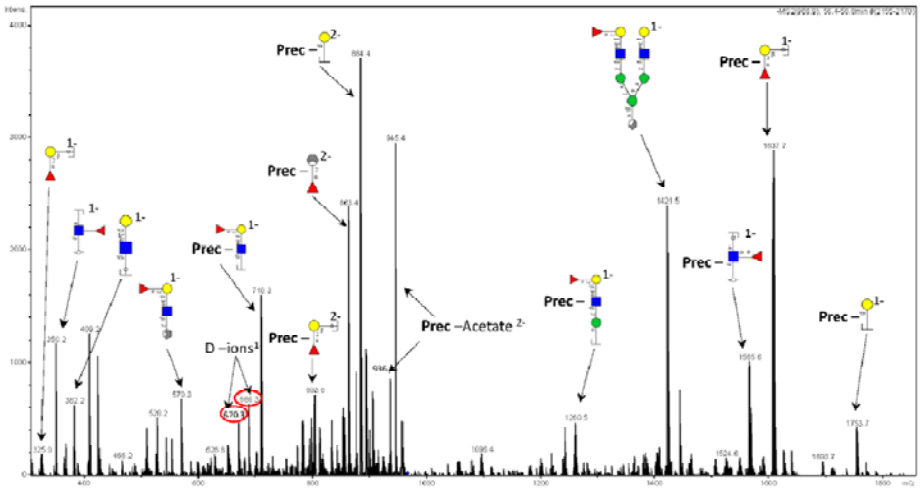


No match to MS2 spectrum in UnicarbKB

Glycan No 22b
Type = Complex
Precursor = *m/z* 966.4^{2-}
[M-H]$^{1-}$ = 1933.8Da
LC retention time = 56.5 min

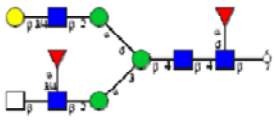


Supplemental Figure S1

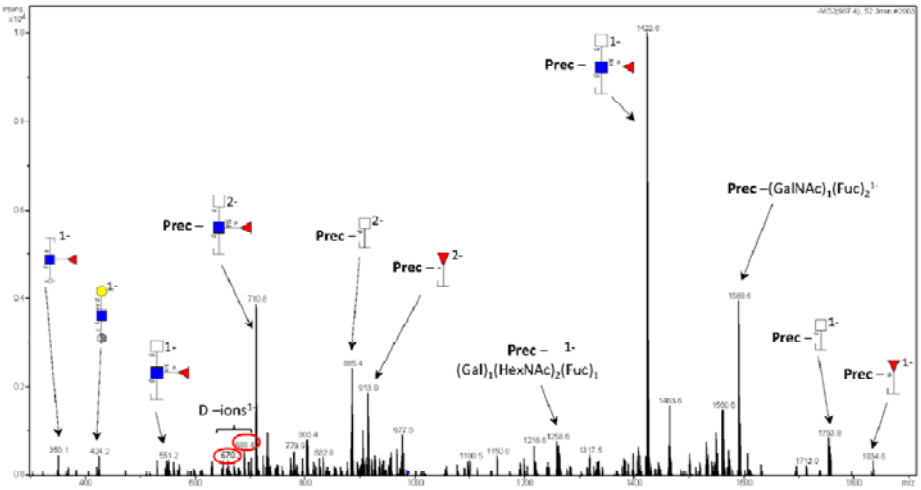


No match to MS2 spectrum in UnicarbKB

Glycan No 23a
Type = Complex
Precursor = *m/z* 986.9^{2-}
[M-H]$^{1-}$ = 1974.8Da
LC retention time = 52.3 min

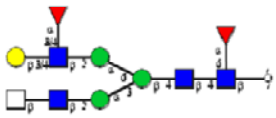


Supplemental Figure S1

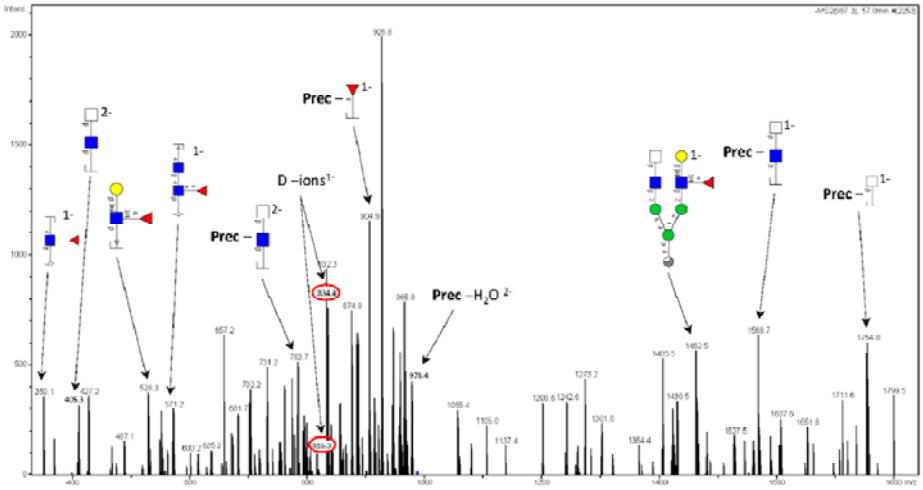


No match to MS2 spectrum in UnicarbKB

Glycan No 23b
Type = Complex
Precursor = *m/z* 986.9^{2-}
[M-H]$^{1-}$ = 1974.8Da
LC retention time = 57 min

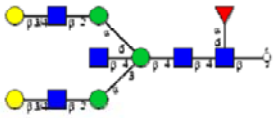


Supplemental Figure S1

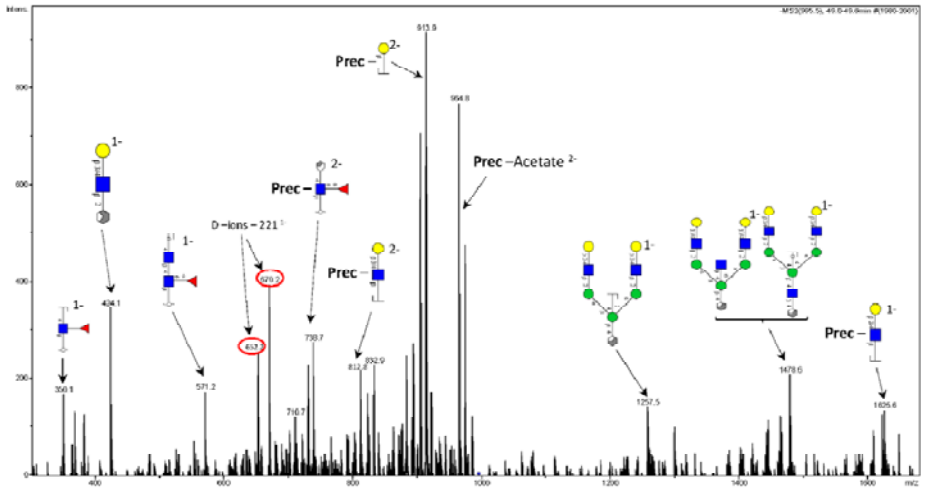


No match to MS2 spectrum in UnicarbKB

Glycan No 24
Type = Complex
Precursor = *m/z* 994.9^{2-}
[M-H]$^{1-}$ = 1990.8 Da
LC retention time = 49.3 min

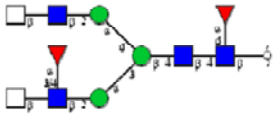


Supplemental Figure S1

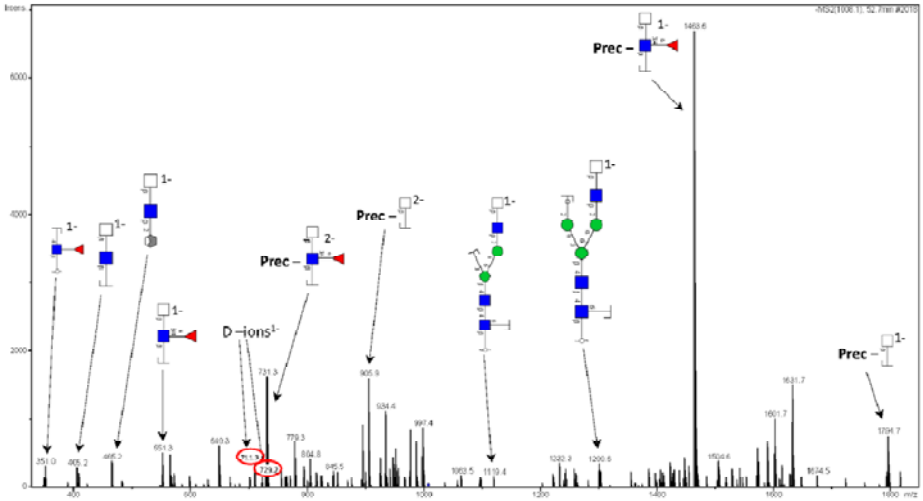


Positive match to MS2 spectrum in UnicarbKB

Glycan No 25
Type = Complex
Precursor = *m/z* 1007.4^{2-}
[M-H]$^{1-}$ = 2015.8 Da
LC retention time = 52.7 min

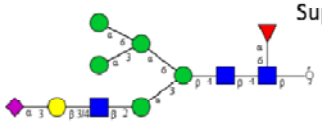


Supplemental Figure S1

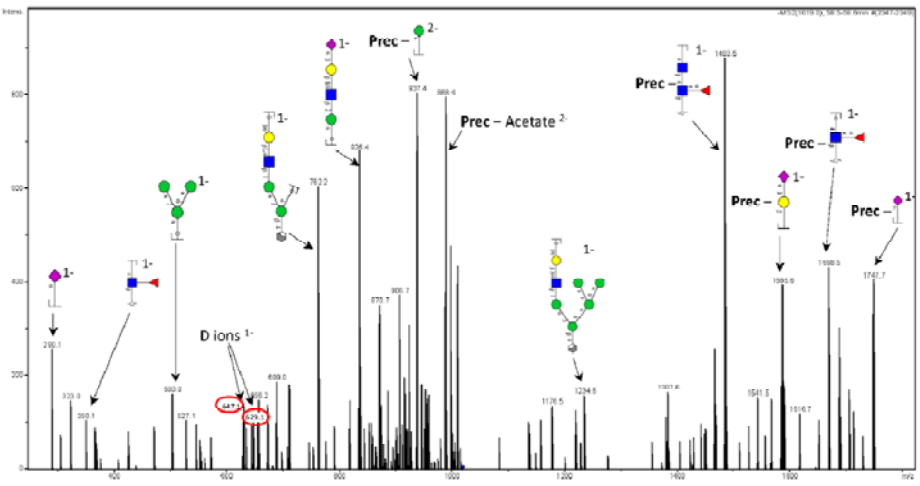


No match to MS2 spectrum in UnicarbKB

Glycan No 26
Type = Hybrid
Precursor = *m/z* 1018.4^{2-}
[M-H]$^{1-}$ = 2037.8 Da
LC retention time = 58.6 min

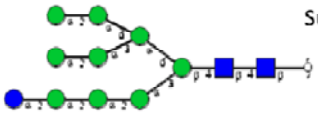


Supplemental Figure S1

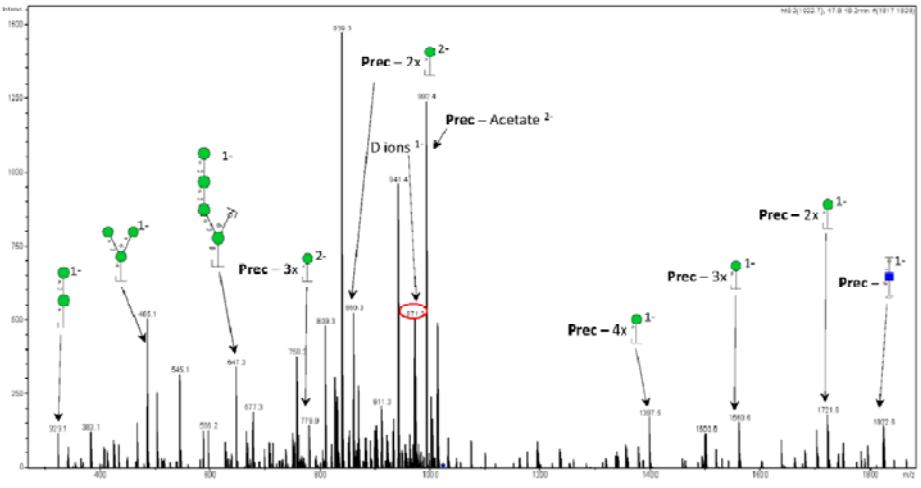


No match to MS2 spectrum in UnicarbKB

Glycan No 27
Type = High Mannose
Precursor = *m/z* 1022.4^{2-}
$[M-H]^{1-}$ = 2045.8 Da
LC retention time = 48 min

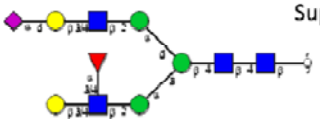


Supplemental Figure S1

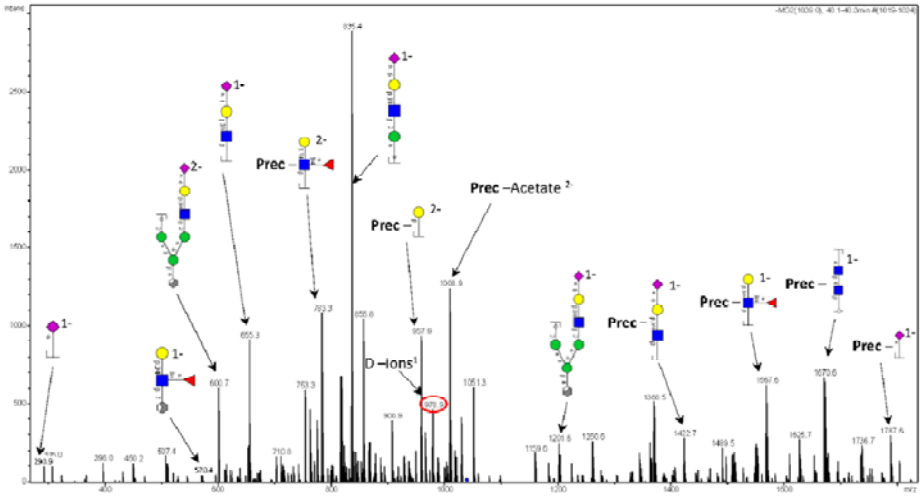


Positive match to MS2 spectrum in UnicarbKB

Glycan No 28a
Type = Complex
Precursor = *m/z* 1038.9^{2-}
$[M-H]^{1-}$ = 2078.8 Da
LC retention time = 48.2 min



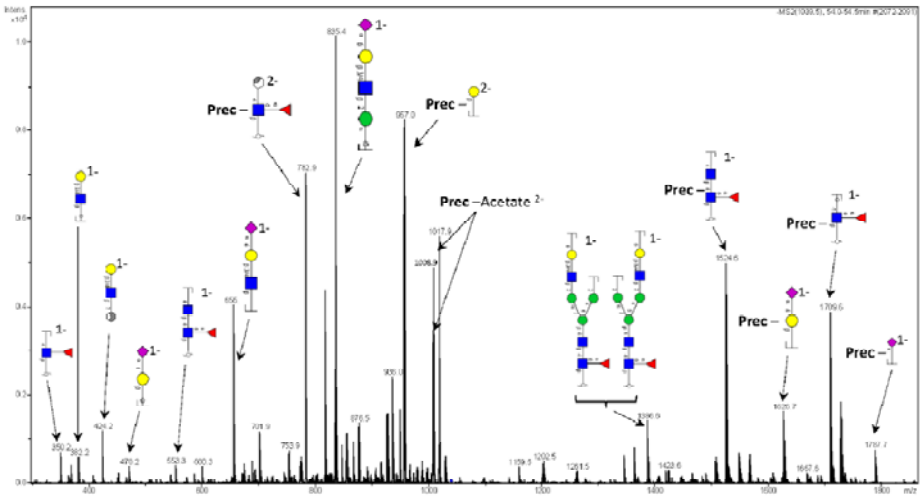
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

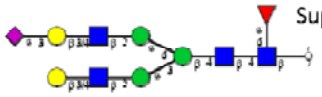
Glycan No 28b
Type = Complex
Precursor = *m/z* 1038.9^{2-}
$[M-H]^{1-}$ = 2078.8 Da
LC retention time = 54.1 min

Supplemental Figure S1

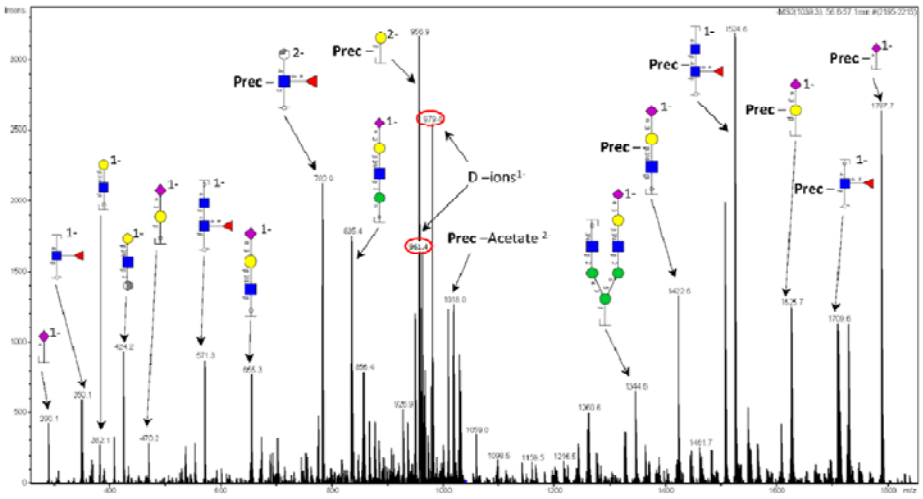


Positive match to MS2 spectrum in UnicarbKB

Glycan No 28c
Type = Complex
Precursor = *m/z* 1038.9^{2-}
$[M-H]^{1-}$ = 2078.8 Da
LC retention time = 56.7 min

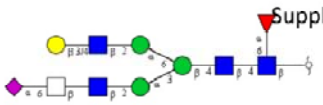


Supplemental Figure S1

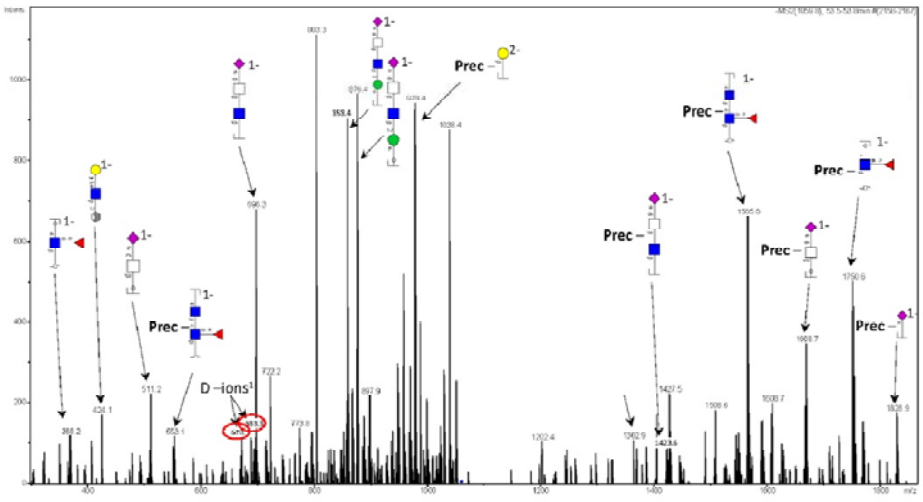


Positive match to MS2 spectrum in UnicarbKB

Glycan No 29
Type = Complex
Precursor = *m/z* 1059.4^{2-}
$[M-H]^{1-}$ = 2119.8Da
LC retention time =53.6 min



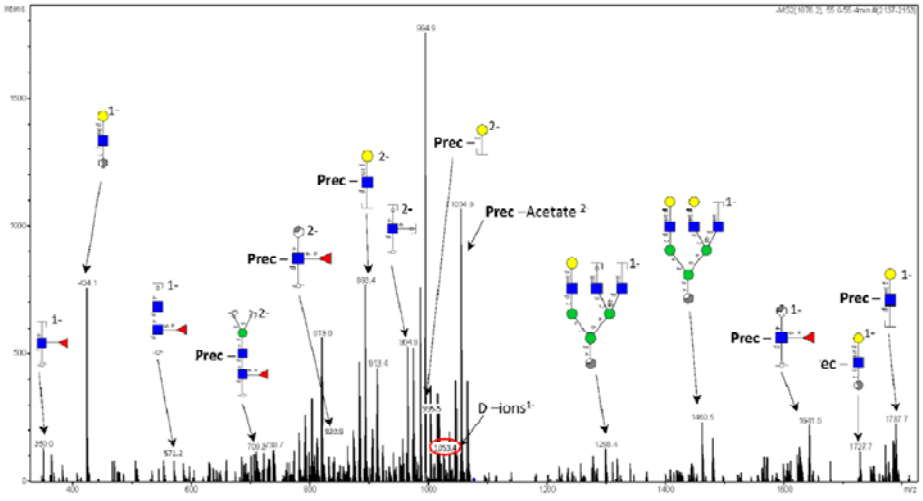
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

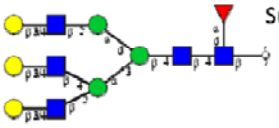
Glycan No 30a
Type = Complex
Precursor = *m/z* 1076.0^{2}
$[M-H]^{1-}$ = 2153 Da
LC retention time = 55.1 min

Supplemental Figure S1

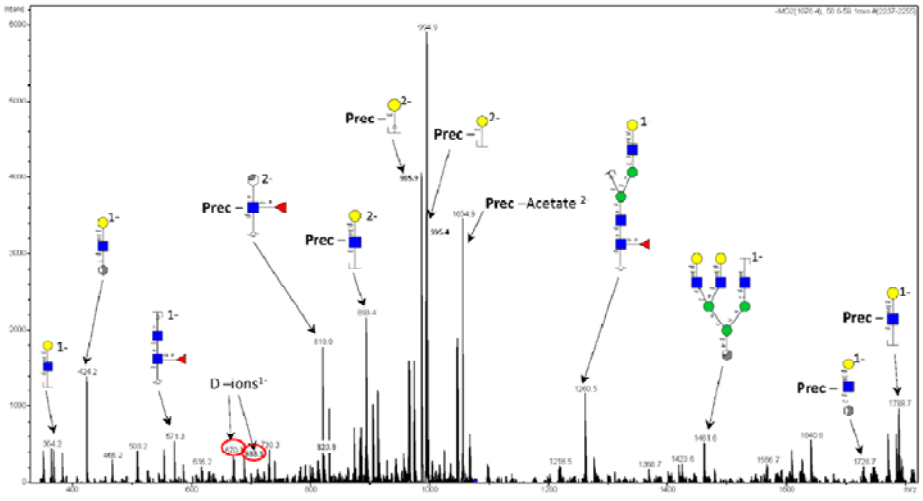


No match to MS2 spectrum in UnicarbKB

Glycan No 30b
Type = Complex
Precursor = *m/z* 1076.0^{2-}
[M-H]$^{1-}$ = 2153 Da
LC retention time = 59.1 min

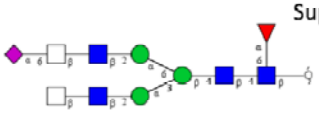


Supplemental Figure S1

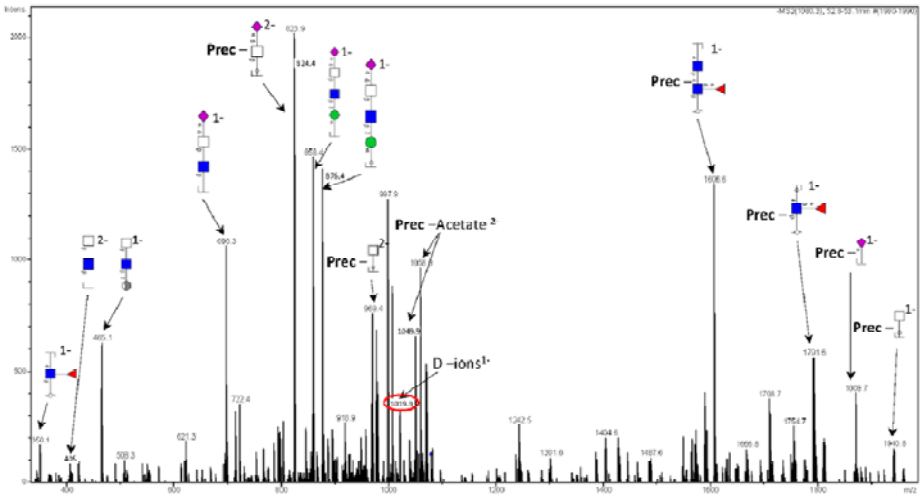


No match to MS2 spectrum in UnicarbKB

Glycan No 31
Type = Complex
Precursor = *m/z* 1079.9^{2-}
[M-H]$^{1-}$ = 2160.8 Da
LC retention time = 52.9 min



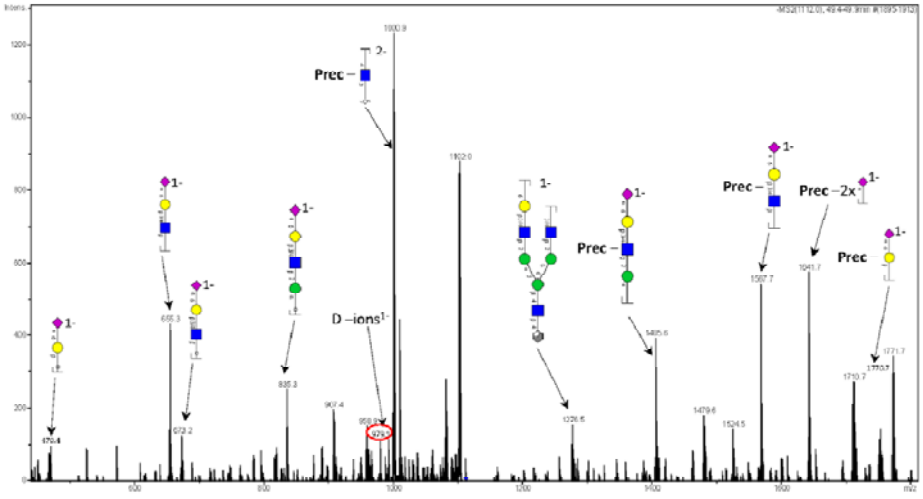
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

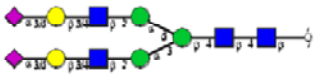
Glycan No 32a
Type = Complex
Precursor = *m/z* 1111.4^{2-}
[M-H]$^{1-}$ = 2223.8 Da
LC retention time = 49.8 min

Supplemental Figure S1

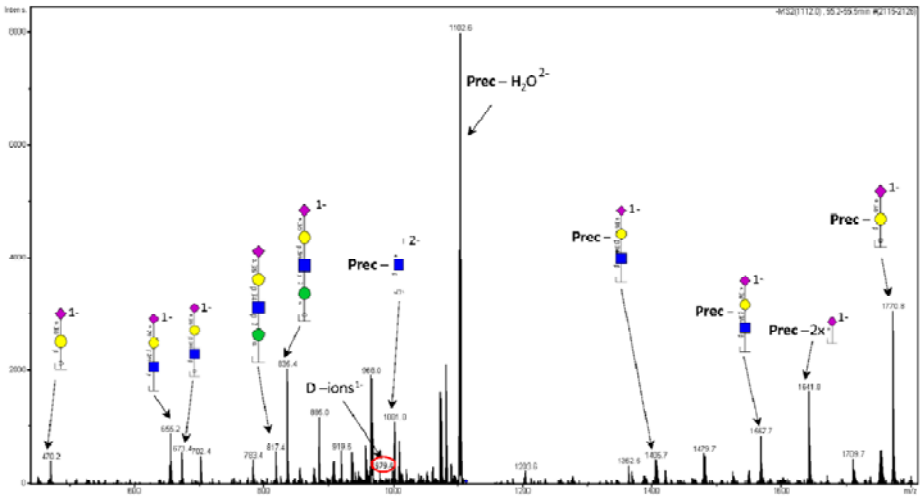


Positive match to MS2 spectrum in UnicarbKB

Glycan No 32b
Type = Complex
Precursor = *m/z* 1111.4^{2-}
[M-H]$^{1-}$ = 2223.8 Da
LC retention time = 55.4 min

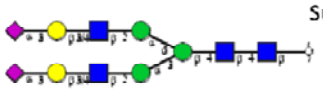


Supplemental Figure S1

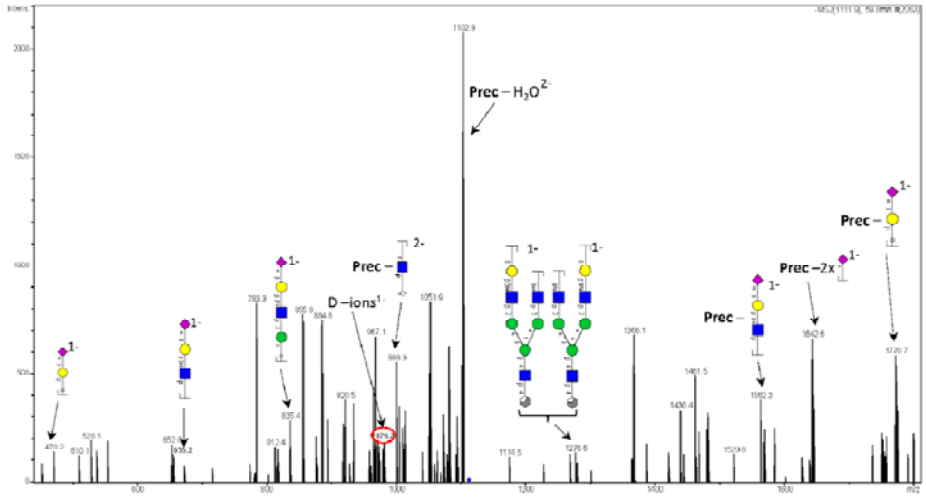


Positive match to MS2 spectrum in UnicarbKB

Glycan No 32c
Type = Complex
Precursor = *m/z* 1111.4^{2-}
[M-H]$^{1-}$ = 2223.8 Da
LC retention time = 59.7 min

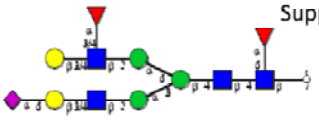


Supplemental Figure S1

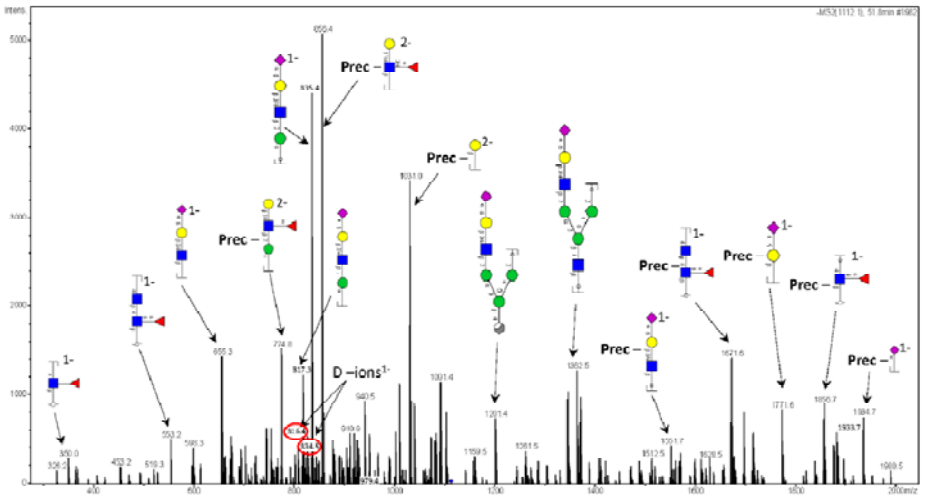


Positive match to MS2 spectrum in UnicarbKB

Glycan No 33a
Type = Complex
Precursor = *m/z* 1111.9^{2-}
[M-H]$^{1-}$ = 2224.8 Da
LC retention time = 51.8 min

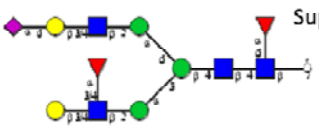


Supplemental Figure S1

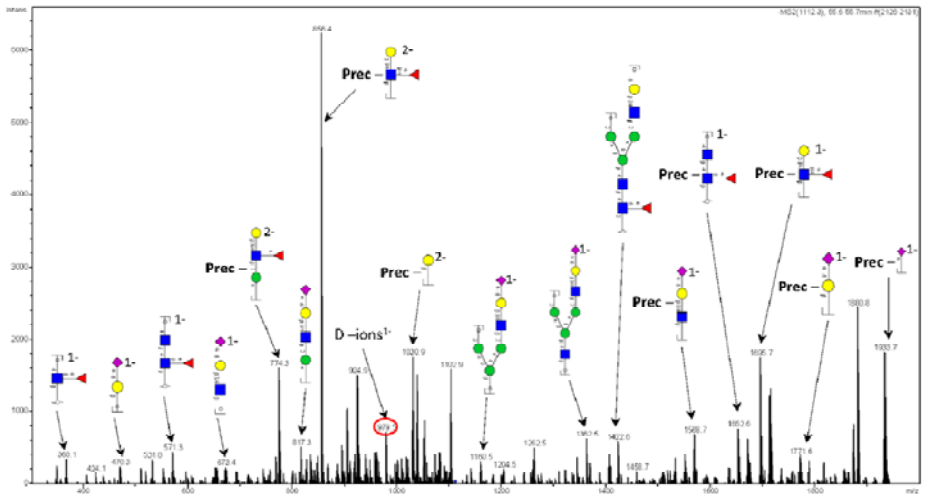


Positive match to MS2 spectrum in UnicarbKB

Glycan No 33b
Type = Complex
Precursor = *m/z* 1111.9^{2-}
$[M-H]^{1-}$ = 2224.8 Da
LC retention time = 55.7 min

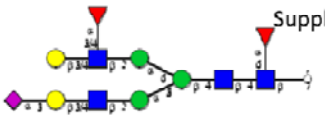


Supplemental Figure S1

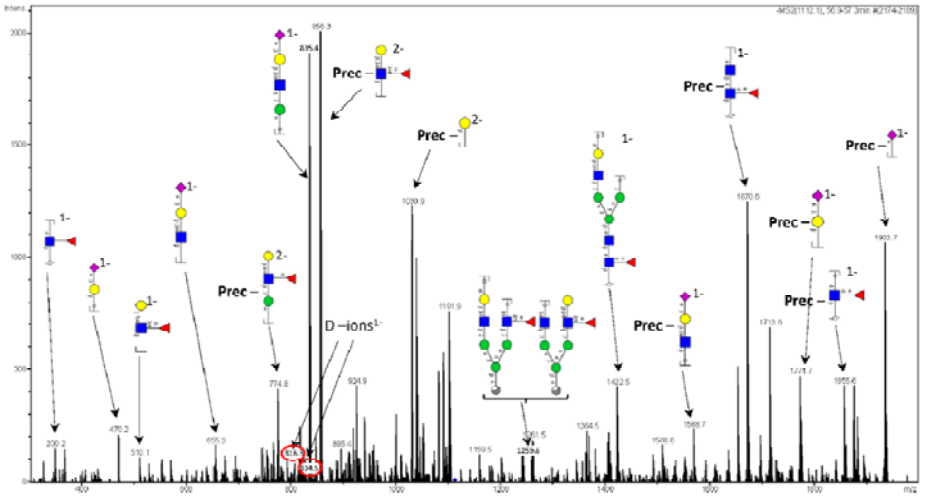


No match to MS2 spectrum in UnicarbKB

Glycan No 33c
Type = Complex
Precursor = *m/z* 1111.9^{2-}
$[M-H]^{1-}$ = 2224.8 Da
LC retention time = 57.0 min

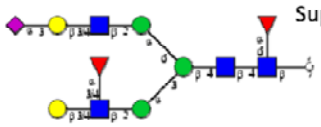


Supplemental Figure S1

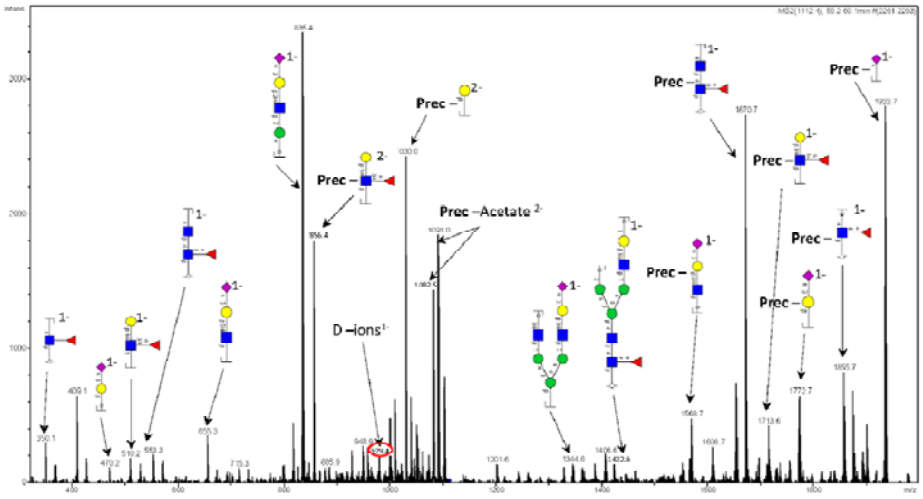


No match to MS2 spectrum in UnicarbKB

Glycan No 33d
Type = Complex
Precursor = *m/z* 1111.9^{2-}
$[M-H]^{1-}$ = 2224.8 Da
LC retention time = 59.6 min

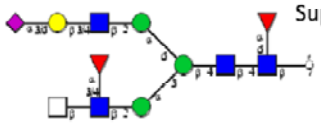


Supplemental Figure S1

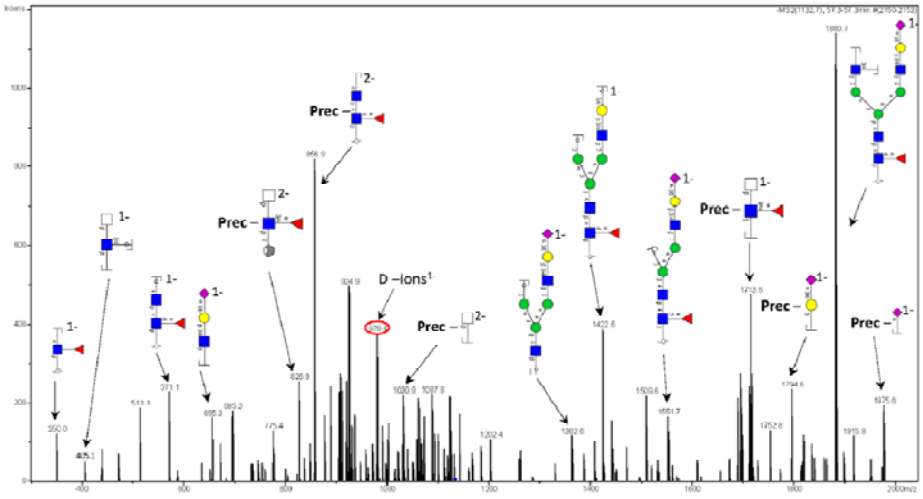


No match to MS2 spectrum in UnicarbKB

Glycan No 34
Type = Complex
Precursor = *m/z* 1132.5^{2-}
$[M-H]^{1-}$ = 2266 Da
LC retention time = 57.3 min

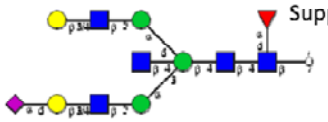


Supplemental Figure S1

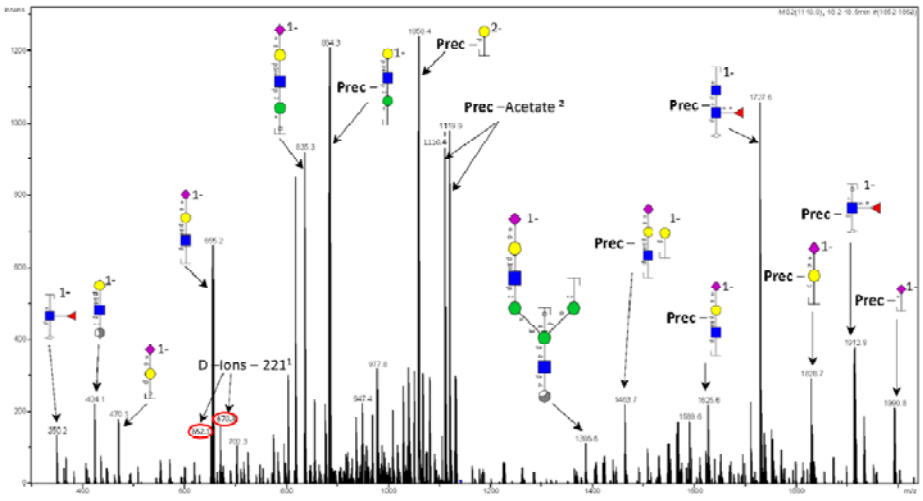


No match to MS2 spectrum in UnicarbKB

Glycan No 35a
Type = Complex
Precursor = *m/z* 1140.4^{2-}
[M-H]$^{1-}$ = 2281.8 Da
LC retention time = 48.3 min

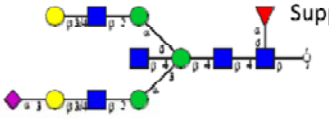


Supplemental Figure S1

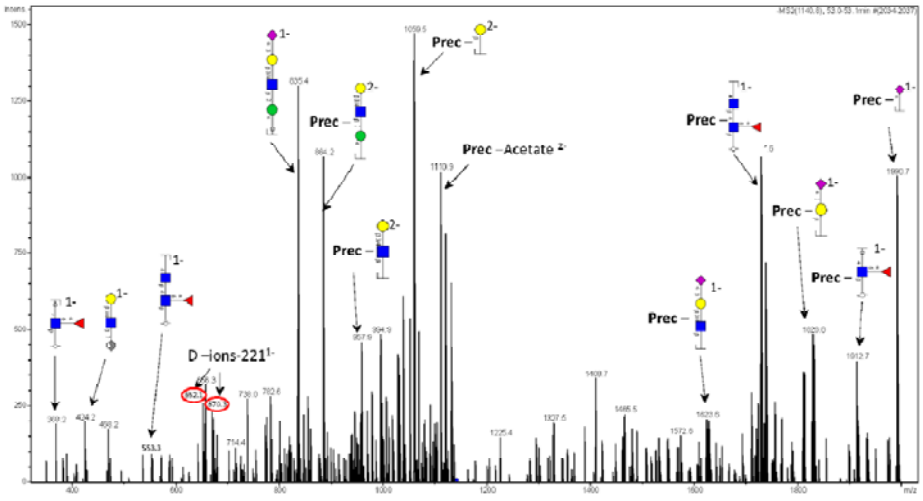


Positive match to MS2 spectrum in UnicarbKB

Glycan No 35b
Type = Complex
Precursor = *m/z* 1140.4^{2-}
[M-H]$^{1-}$ = 2281.8 Da
LC retention time = 53 min

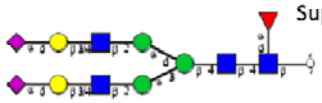


Supplemental Figure S1

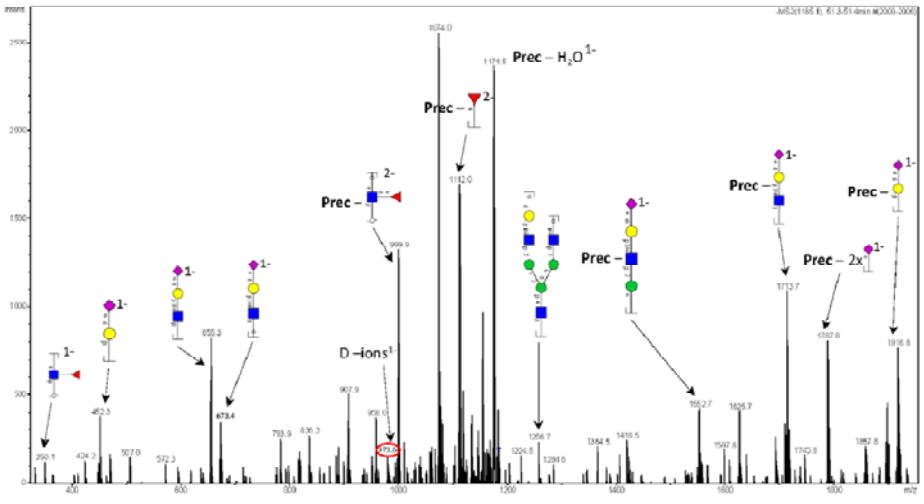


Positive match to MS2 spectrum in UnicarbKB

Glycan No 36a
Type = Complex
Precursor = *m/z* 1184.5^{2-}
[M-H]$^{1-}$ = 2370 Da
LC retention time = 51.4 min



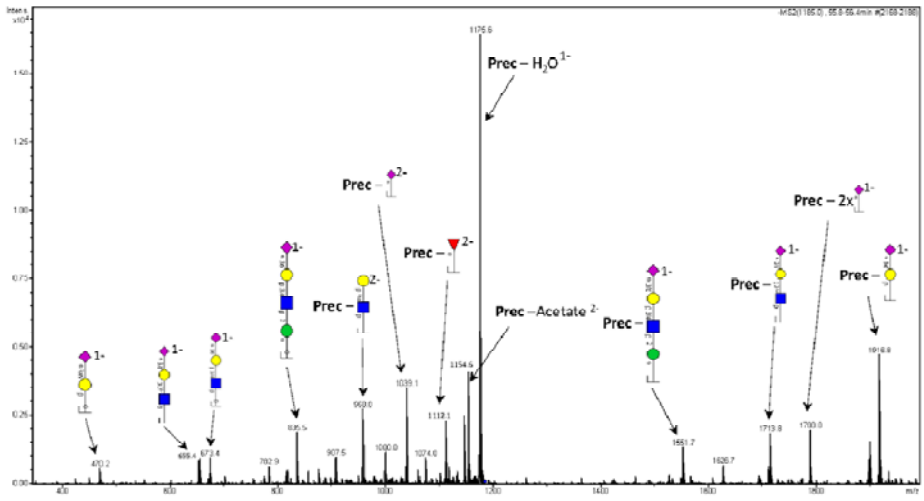
Supplemental Figure S1



No match to MS2 spectrum in UnicarbKB

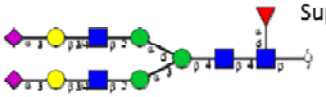
Glycan No 36b
Type = Complex
Precursor = *m/z* 1184.5^{2-}
[M-H]$^{1-}$ = 2370 Da
LC retention time = 56 min

Supplemental Figure S1

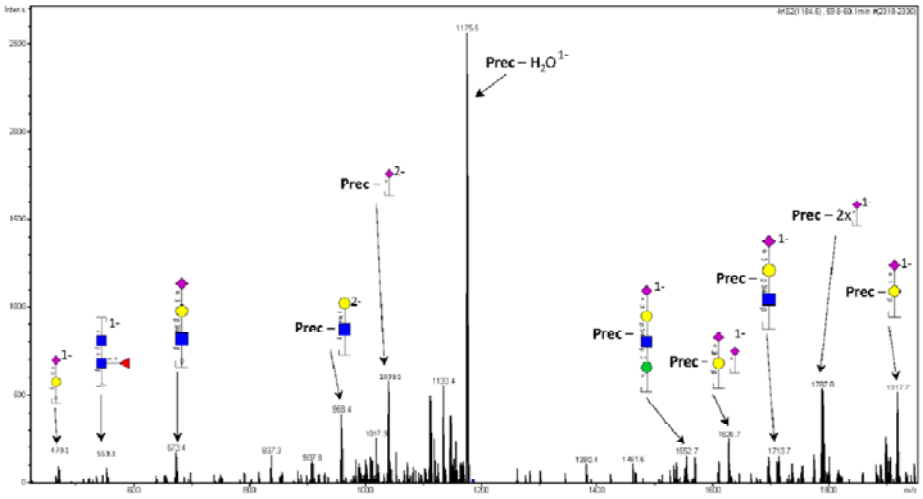


No match to MS2 spectrum in UnicarbKB

Glycan No 36c
Type = Complex
Precursor = *m/z* 1184.5^{2-}
$[M-H]^{1-}$ = 2370 Da
LC retention time = 60 min

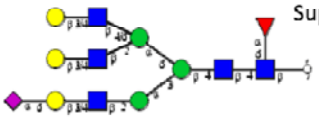


Supplemental Figure S1

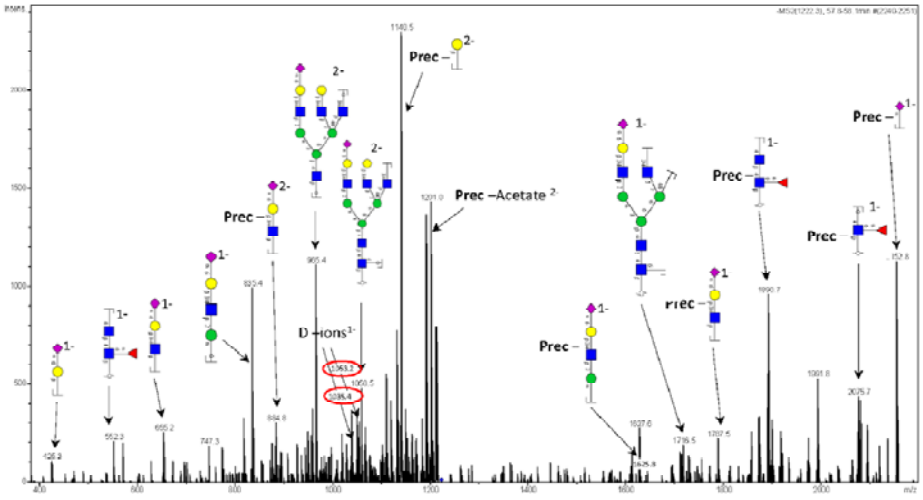


No match to MS2 spectrum in UnicarbKB

Glycan No 37a
Type = Complex
Precursor = *m/z* 1221.5^{2-}
$[M-H]^{1-}$ = 2444 Da
LC retention time = 57.9 min

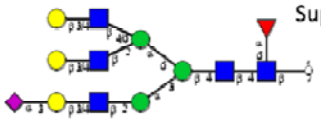


Supplemental Figure S1

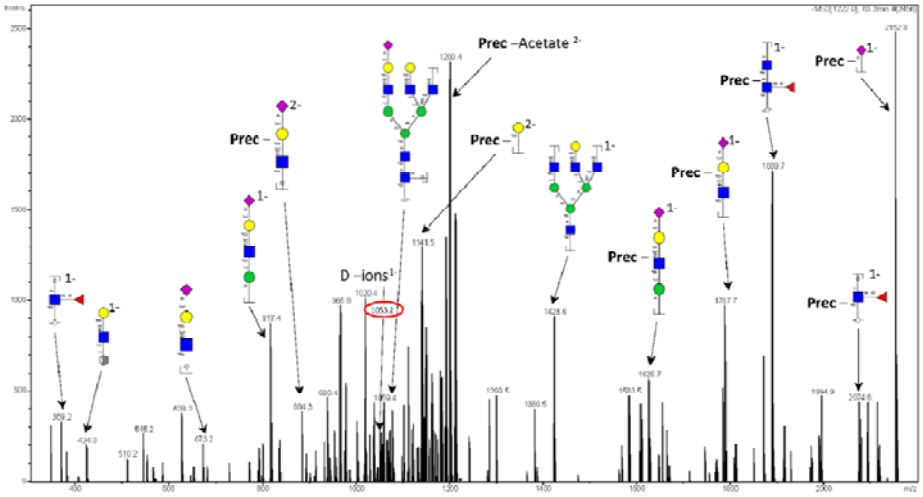


No match to MS2 spectrum in UnicarbKB

Glycan No 37b
Type = Complex
Precursor = *m/z* 1221.5^{2-}
[M-H]$^{1-}$ = 2444 Da
LC retention time = 63.2 min

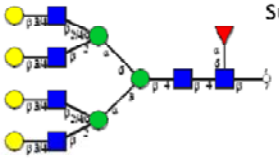


Supplemental Figure S1

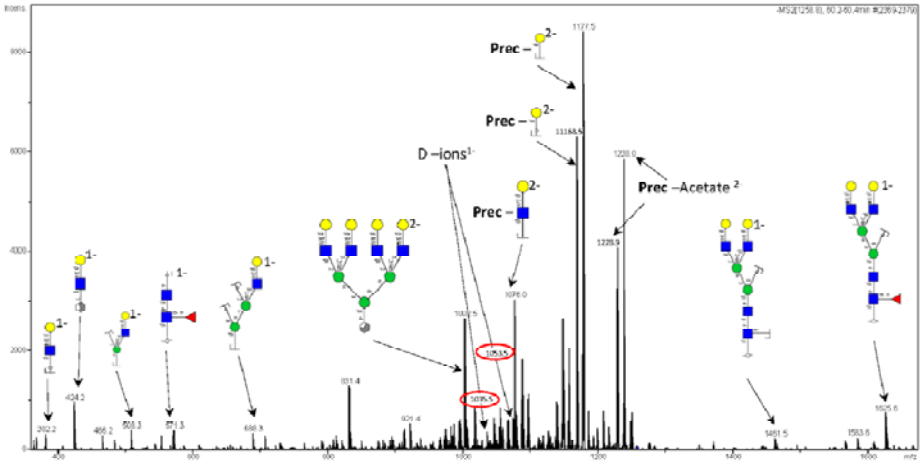


No match to MS2 spectrum in UnicarbKB

Glycan No 38
Type = Complex
Precursor = *m/z* 1258.5^{2-}
[M-H]$^{1-}$ = 2517 Da
LC retention time = 60.3 min

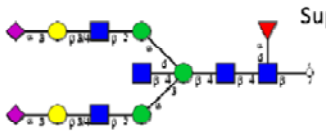


Supplemental Figure S1

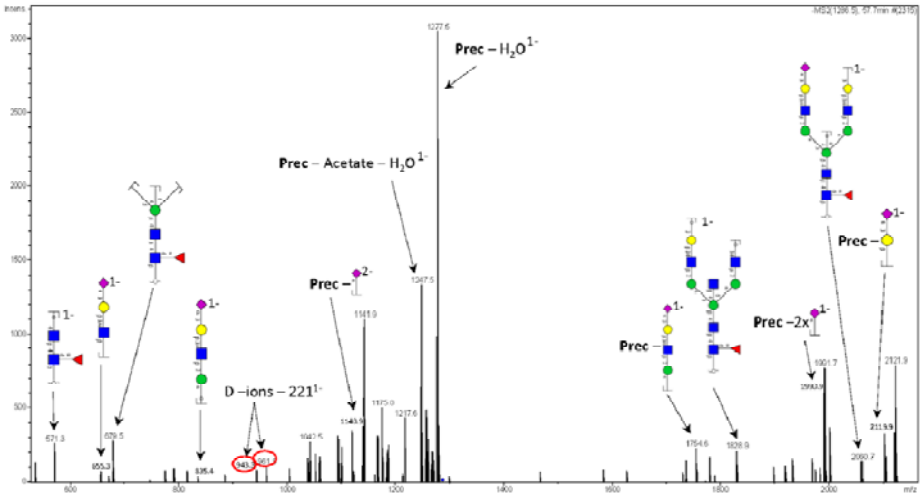


No match to MS2 spectrum in UnicarbKB

Glycan No 39
Type = Complex
Precursor = *m/z* 1286^{2-}
$[M-H]^{1-}$ = 2573 Da
LC retention time = 57.7 min

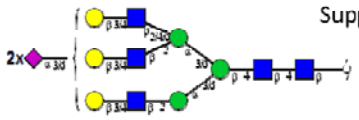


Supplemental Figure S1

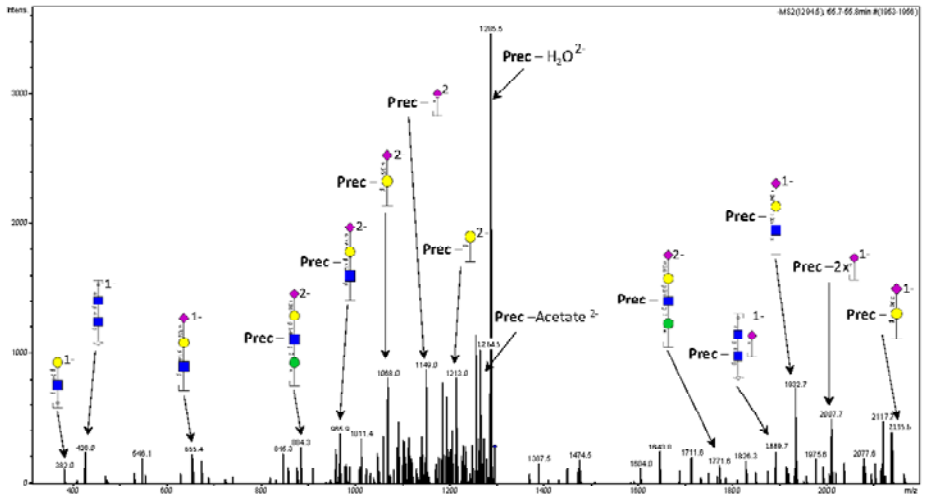


No match to MS2 spectrum in UnicarbKB

Glycan No 40a
Type = Complex
Precursor = *m/z* 1294^{2-}
$[M-H]^{1-}$ = 2589 Da
LC retention time = 55.8 min

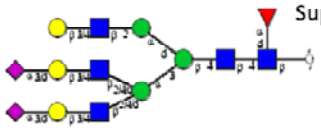


Supplemental Figure S1

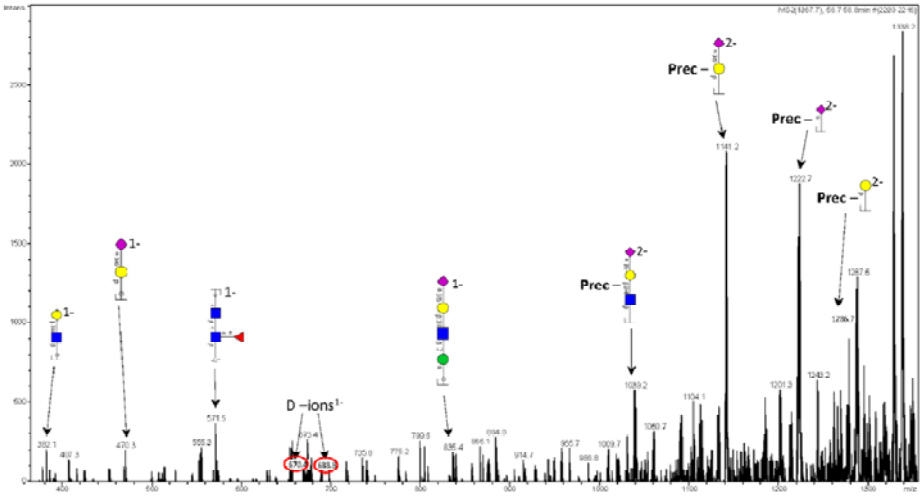


No match to MS2 spectrum in UnicarbKB

Glycan No 41
Type = Complex
Precursor = *m/z* 1367^{2}
$[M-H]^{1-}$ = 2735Da
LC retention time = 58.8 min

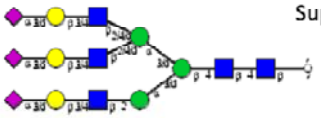


Supplemental Figure S1

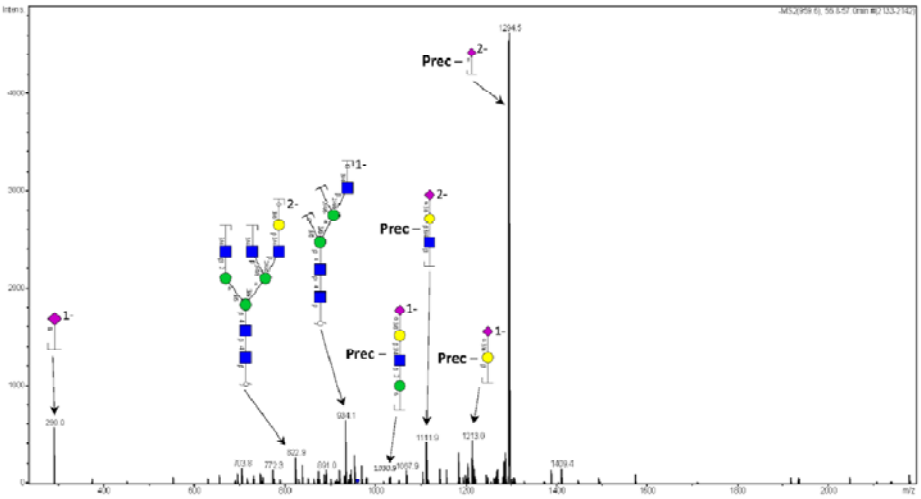


No match to MS2 spectrum in UnicarbKB

Glycan No 42a
Type = Complex
Precursor = *m/z* 959.4^{3}
$[M-H]^{1-}$ = 2880.2Da
LC retention time = 56.9 min

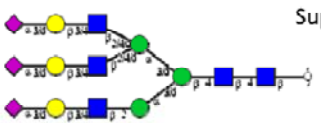


Supplemental Figure S1

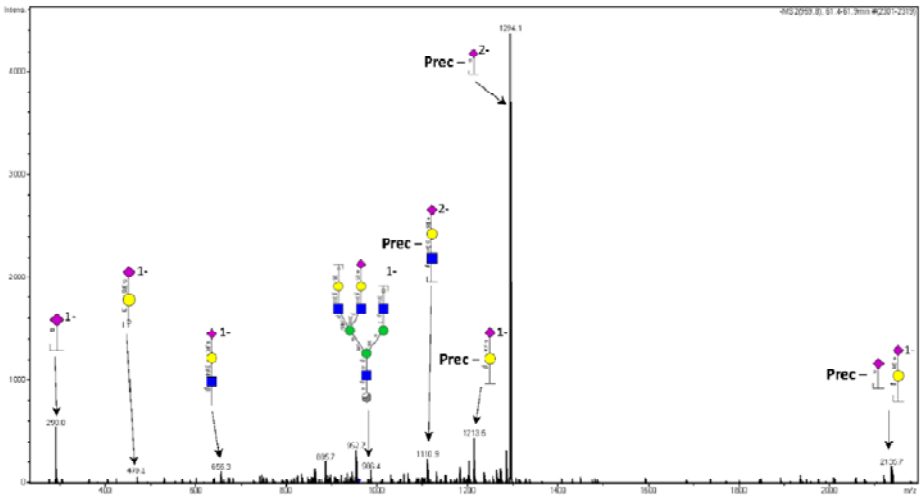


No match to MS2 spectrum in UnicarbKB

Glycan No 42b
Type = Complex
Precursor = *m/z* 959.4^{3-}
$[M-H]^{1-}$ = 2880.2Da
LC retention time = 61.6 min

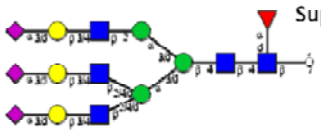


Supplemental Figure S1

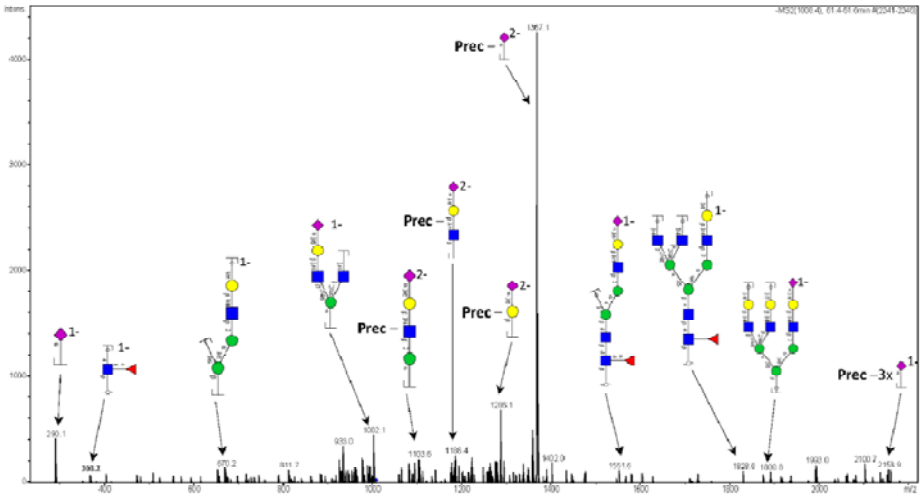


No match to MS2 spectrum in UnicarbKB

Glycan No 43a
Type = Complex
Precursor = *m/z* 1008^{3}
$[M-H]^{1-}$ = 3026.2Da
LC retention time = 61.4 min

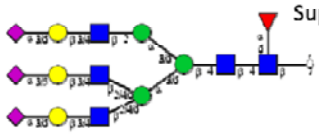


Supplemental Figure S1

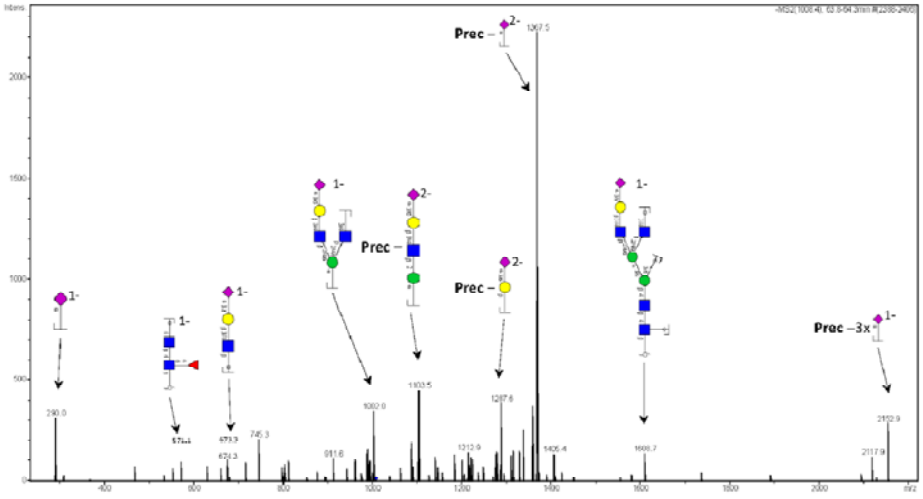


No match to MS2 spectrum in UnicarbKB

Glycan No 43b
Type = Complex
Precursor = *m/z* 1008^{3}
$[M-H]^{1-}$ = 3026.2Da
LC retention time = 64.2 min

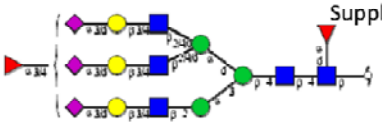


Supplemental Figure S1

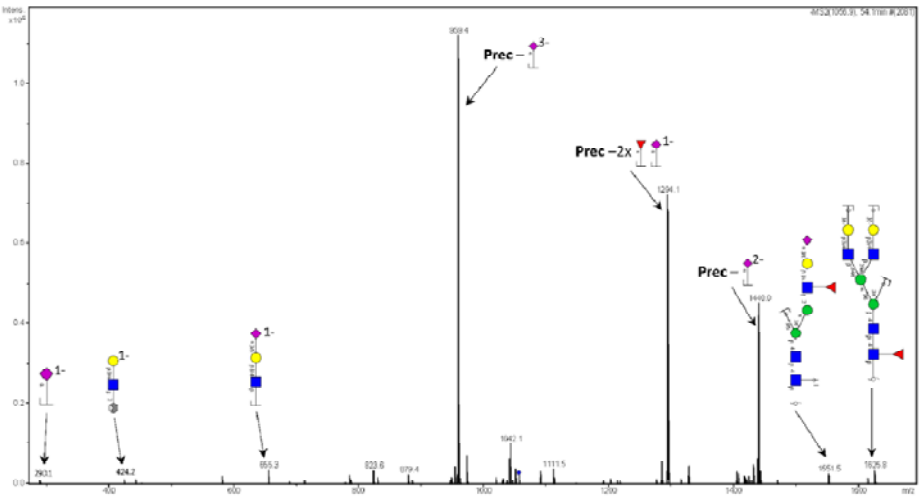


No match to MS2 spectrum in UnicarbKB

Glycan No 44a
Type = Complex
Precursor = *m/z* 1056.4^{3}
$[M-H]^{1-}$ = 3172.2Da
LC retention time = 54.1 min

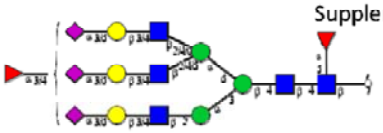


Supplemental Figure S1

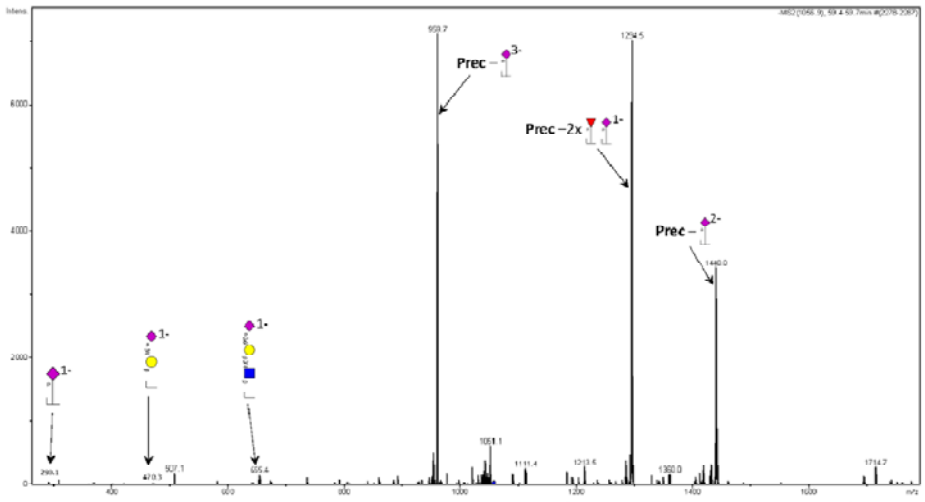


No match to MS2 spectrum in UnicarbKB

Glycan No 44b
Type = Complex
Precursor = *m/z* 1056.4^{3-}
$[M-H]^{1-}$ = 3172.2Da
LC retention time = 59.4 min

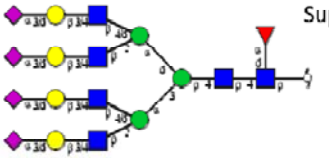


Supplemental Figure S1

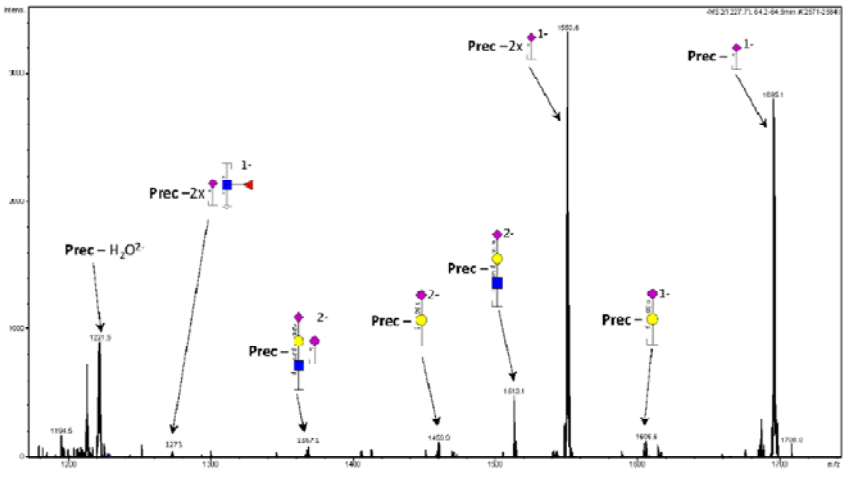


No match to MS2 spectrum in UnicarbKB

Glycan No 45
Type = Complex
Precursor = *m/z* 1226.8^{3}
$[M-H]^{1-}$ = 3682.3 Da
LC retention time = 64.4 min

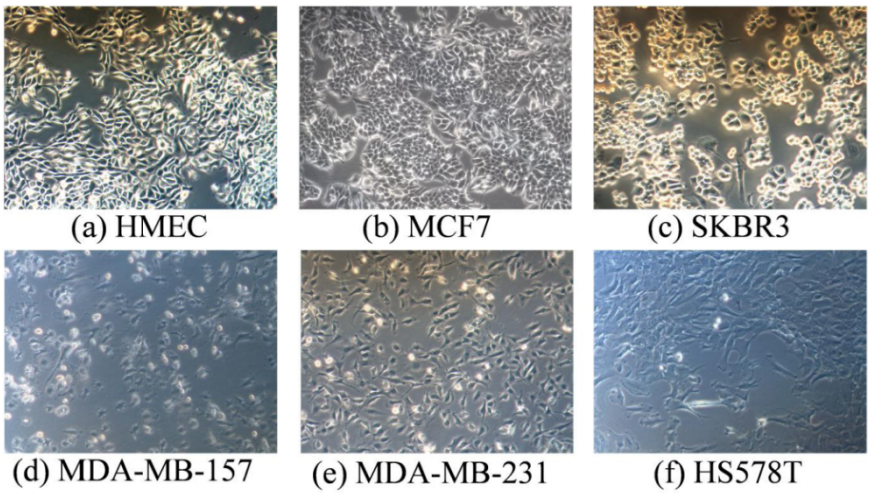


Supplemental Figure S1



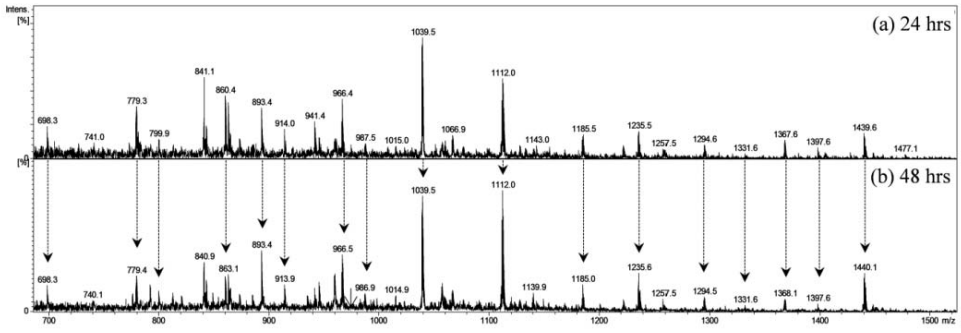
No match to MS2 spectrum in UnicarbKB

Supplemental Figure S2



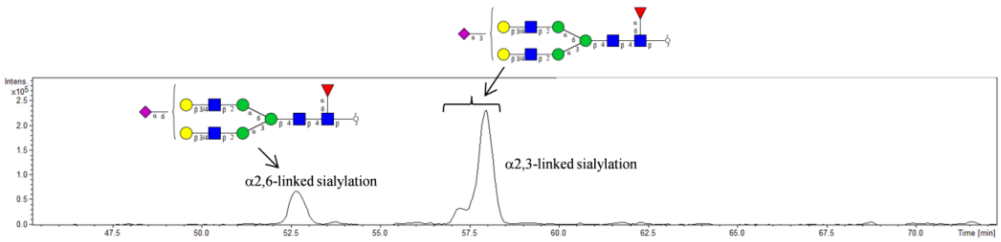
Supplemental Figure S2 - Morphology of the panel of breast epithelial cell lines. Light microscopy based morphology assessment of the cell lines (a) HMEC, (b) MCF7, (c) SKBR3, (d) MDA-MB-157, (e) MDA-MB-231 and (e) HS578T.

Supplemental Figure S3



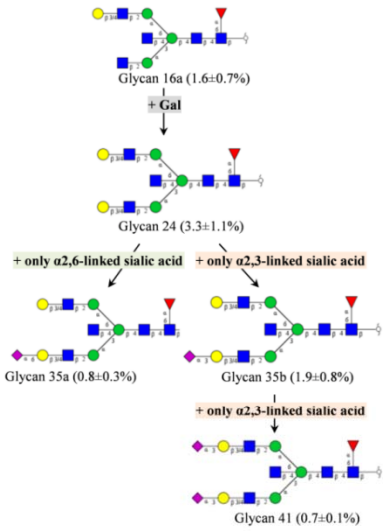
Supplemental Figure S3 - PGC-LC-ESI-IT-MS/MS mass profile (MS1) of secretome *N*-glycans at 24 hours (a) and 48 hours (b) after change to serum-free media. Arrows indicate the *m/z*'s (predominantly of Z=2-3) corresponding to *N*-glycans identified in both profiles. Although no significant changes in the qualitative expression of *N*-glycans were observed, minor quantitative differences, in part reflecting technical LC-MS/MS variation, were observed from 24 h to 48 h.

Supplemental Figure S4



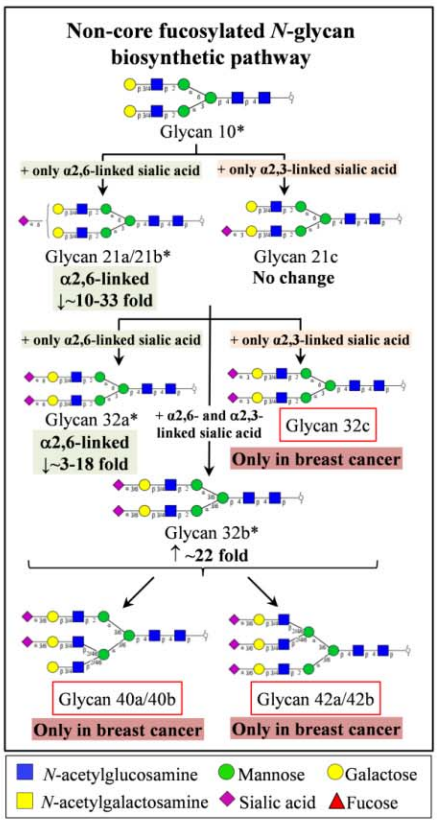
Supplemental Figure S4 - EIC of complex type *N*-glycans of *m/z* 1038.9^{2-} from PGC-LC-ESI-IT-MS/MS. Mono-sialylated *N*-glycan isomers have different sialic acid linkages which were well separated by PGC-LC. *N*-glycans with α2,6-linked sialylation eluted earlier (~52.5 min) than those with α2,3-linked sialylation (57-58 min).

Supplemental Figure S5



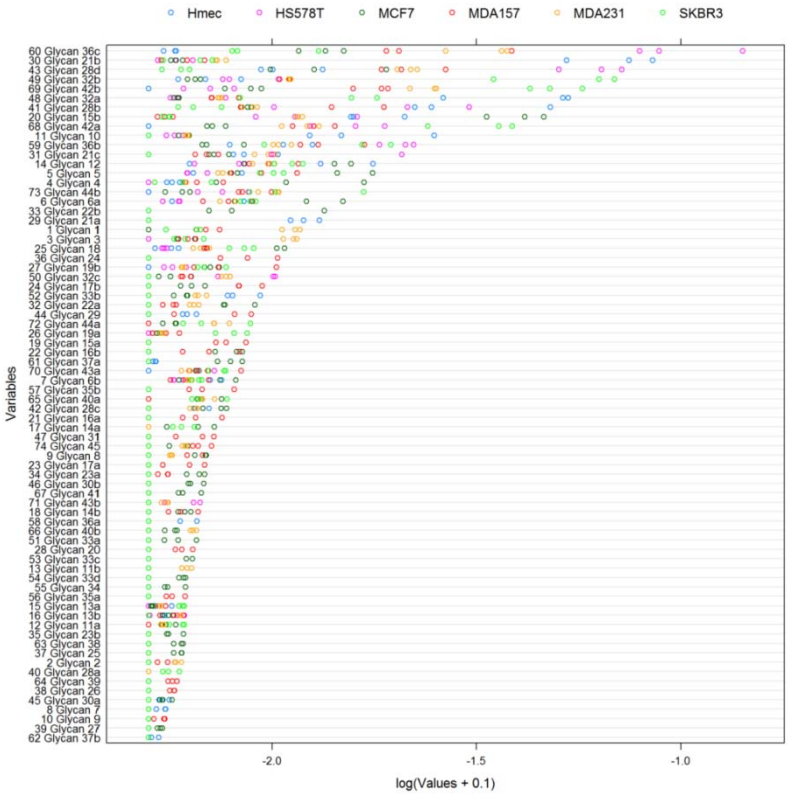
Supplemental Figure S5 - Biosynthetic relationship between the five bisecting GlcNAc containing *N*-glycans in MDA157. Five core-fucosylated bisecting GlcNAc containing *N*-glycans were detected in MDA157, The isomeric glycans 35a and 35b differed by their sialic acid linkages; the α2,3-linked sialylated structure was expressed at twice the level of the α2,6-linked isomer.

Supplemental Figure S6



Supplemental Figure S6 - Non-core fucosylated secretome *N*-glycans mapped to *N*-glycosylation biosynthetic pathways. *N*-glycans were found to be significantly up- or down-regulated (*) in the breast cancer cell lines relative to HMEC. Unmarked *N*-glycans were either not significantly regulated or found exclusively in normal or cancer samples and depicted here to complete the pathways. Expression of α2,6-sialylated *N*-glycans (glycan 21b and 32a) were reduced 3-33 fold whereas the α2,3-sialylated *N*-glycan (glycan 32c) was only detected in the breast cancer cell lines.

Supplemental Figure S7



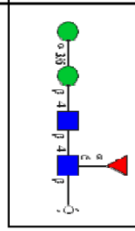
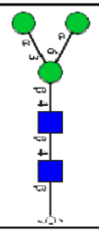
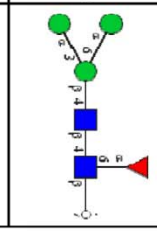
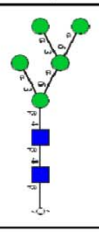
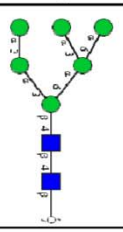
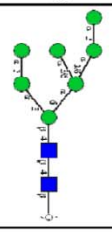
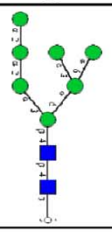
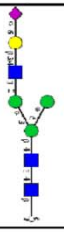
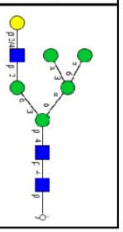
Supplemental Figure S7 - Log transformed glycoprofiling data of the secretomes of the six investigated breast cell lines. The top 12 *N*-glycans (Glycans 36c, 21b, 28d, 32b, 42b, 32a, 28b, 15b, 42a, 10, 26b and 21c) contributed the most to the variance observed in PCA1 of the principal component analysis.

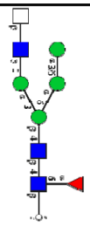
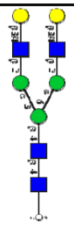
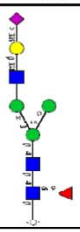
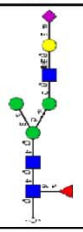
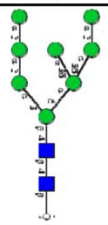
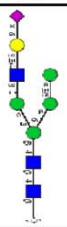
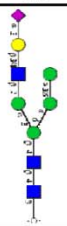
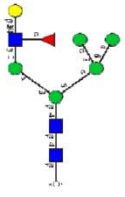
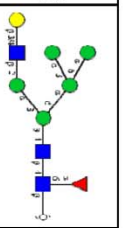
Supplementary Table S1 - Released *N*-glycans identified from conditioned media of 8 breast epithelial cell lines by PGC-LC-MS/MS (CID)

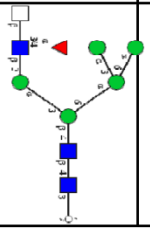
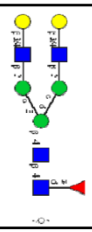
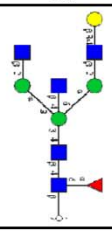
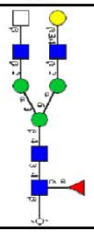
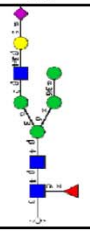
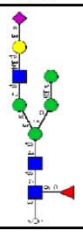
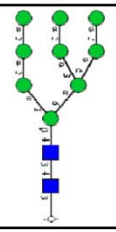
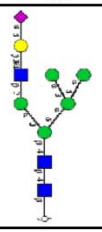
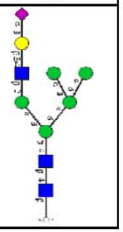
											RT=Retention time															
Glycan #	Theo m/z	charge	Hex	HexNAc	dHex	NeuAc	Isoform	MS/MS	Theo mass (M)	RT_Hmec	Obs. Mass (M)_HMEC	RT_MCF7	Obs. Mass (M)_MCF7	RT_SKBR3	Obs. Mass (M)_SKBR3	RT_MDA157	Obs. Mass (M)_MDA157	RT_MDA231	Obs. Mass (M)_MDA231	RT_HS578T	Obs. Mass (M)_HS578T	Delta mass HMEC	Delta mass MCF7	Delta mass SKBR3	Delta mass MDA157	Delta mass MDA231
1	895.4	1	2	2	1		a	Y	896.3					51	896.2	51.5	896.4	49	896.4					-0.1	0.1	0.1
2	911.3	1	3	2			a	Y	912.3							50	912.3	48	912.4						0.0	0.1
3	1057.5	1	3	2	1		a	Y	1058.4			53.7	1058.5	56.3	1058.5	56.5	1058.6	51	1058.5				-0.1	-0.1	0.0	0.1
4	1255.5	1	5	2			a	Y	1256.4	50.1	1256.5073	52	1256.5	52	53.7	52.3	1256.5	50	1256.4			0.1	0.1		0.1	0.0
5	698.3	2	6	2			a	Y	1396.5	45.6	1396.6146	47	1396.6	47	1396.6	47	1396.6	45	1396.6	45.9	1396.6	0.1	-0.1	-0.1	0.1	0.1
6a	779.3	2	7	2			a	Y	1560.5	44.3	1560.6146	46	1560.6	46	1560.6	46	1560.6	44	1560.6	43.6	1560.6	0.1	0.1	0.1	0.1	0.1
6b	779.3	2	7	2			b	Y	1560.5	45.5	1560.6146	47	1560.6	47	1560.6	48	1560.6	45	1560.6	45.8	1560.6	0.1	0.1	0.1	0.1	0.1
7	781.3	2	4	3		1	a	Y	1562.5	46.2	1562.6146											0.1				
8	799.8	2	6	3			a	Y	1601.6	48.2	1601.6146	49.8	1601.6			50.1	1601.6	47	1601.6			0.1	0.1		0.1	0.1
9	812.3	2	4	4	1		b	Y	1626.6							50.8	1626.6								0.0	
10	820.3	2	5	4			a	Y	1642.6	49.7	1642.6146	51.2	1642.6					49	1642.6	50.4	1642.6	0.0	0.0			0.0
11a	858.3	2	4	3	1	1	a	Y	1714.6			55.8	1714.6	57	1714.6			55	1738.6				0.0	0.0		0.0
11b	858.3	2	4	3	1	1	b	Y	1714.6									54	1734.6							0.0
12	860.3	2	8	2			a	Y	1722.6	47	1722.6146	47	1722.6	47	1722.6	47	1722.6	44	1722.6	46.2	1722.6	0.1	0.0	0.0	0.0	0.0
13a	864.3	2	5	3		1	a	Y	1730.6	45.4	1730.6146	46.1	1730.6	46.2	1730.6	46.4	1730.6	45	1730.6			0.0	0.0	0.0	0.0	0.0
13b	864.3	2	5	3		1	b	Y	1730.6	51.4	1730.6146	51.7	1730.6			51.2	1730.6	51	1730.6	51.4	1730.6	0.0	0.0		0.0	0.0
14a	872.8	2	6	3	1		a	Y	1747.6			47.8	1747.6	48	1747.6								0.0	0.0		
14b	872.8	2	6	3	1		b	Y	1747.6			56	1747.6			54.8	1747.6						0.0		0.0	
15a	891.3	2	5	4	1		a	Y	1782.6							48.3	1782.6									
15b	891.3	2	5	4	1		b	Y	1782.6	53.4	1782.6146	55	1782.6	56	1782.6	55	1782.6	55	1782.6	55	1782.6	0.0	0.0	0.0	0.0	0.0
16a	913.8	2	4	5	1		a	Y	1829.7							48.1	1829.6								0.1	
16b	913.8	2	4	5	1		b	Y	1829.7			54.3	1829.6			55	1829.6						0.1		0.1	
17a	957.3	2	5	3	1	1	a	Y	1876.7							50.2	1876.6								0.0	
17b	957.3	2	5	3	1	1	b	Y	1876.7			56	1876.6			56.7	1876.6						0.2		0.0	
18	941.3	2	9	2			a	Y	1884.6	44.8	1884.6146	47	1884.6	47	1884.6	47	1884.6	44	1884.6	45.5	1884.6	0.0	0.0	0.0	0.0	0.0
19a	945.3	2	6	3		1	a	Y	1892.7			48.8	1892.6	49	1892.6	49	1892.6	48	1892.6				0.0	0.2	0.0	0.0
19b	945.3	2	6	3		1	b	Y	1892.7			54.4	1892.6	55	1892.6	55.2	1892.6	54	1892.6	54.6	1892.6		0.0	0.2	0.0	0.0
20	957.9	2	4	4	1	1	a	Y	1917.7							49.5	1917.6								0.1	
21a	966.9	2	5	4		1	a	Y	1933.7	48.5	1933.6146											0.1				
21b	966.9	2	5	4		1	b	Y	1933.7	49.1	1933.6146	50	1933.6	50	1933.6	51	1933.6	49	1933.6	50	1933.6	0.1	-0.1	-0.1	0.1	0.1
21c	966.9	2	5	4		1	c	Y	1933.7	54.5	1933.6146	53.7	1933.6			53.6	1933.6	53	1933.6	53.7	1933.6	0.1	-0.1		0.1	0.1
22a	966.4	2	5	4	2		a	Y	1934.7			53.2	1934.6			53.8	1934.6	51	1934.6				-0.1		0.1	0.1
22b	966.4	2	5	4	2		b	Y	1934.7			57.1	1934.6										-0.1			
23a	988.9	2	4	5	2		a	Y	1975.7			53.3	1975.6			53.8	1975.6						0.1		0.1	
23b	988.9	2	4	5	2		b	Y	1975.7			57	1975.6										0.1			
24	994.9	2	5	5	1		a	Y	1991.7							49.3	1991.6								-0.1	
25	1007.4	2	5	6	2		a	Y	2016.8			53.3	2016.8										0.1			
26	1018.4	2	6	5	1	1	a	Y	2038.7							58.6	2038.6								0.1	
27	1022.4	2	10	2			a	Y	2046.7			48	2046.8										0.1			
28a	1058.9	2	5	4	1	1	a	Y	2079.7					49	2079.8									0.1		
28b	1058.9	2	5	4	1	1	c	Y	2079.7	52.3	2079.8146	54	2079.8	54	2079.8	54	2079.8	53	2079.8	53.4	2079.8	0.1	0.1	0.1	0.1	0.1
28c	1058.9	2	5	4	1	1	d	Y	2079.7	56.8	2079.8146	58.6	2079.8					57	2079.8			0.1	0.1			0.1
28d	1058.9	2	5	4	1	1	e	Y	2079.7	57.4	2079.8146	58.9	2079.8	59	2079.8	59.7	2079.8	58	2079.8	58.9	2079.8	0.1	0.1	0.1	0.1	0.1
29	1059.4	2	4	5	1	1	b	Y	2120.8	51.7	2120.8146					51.1	2120.8					0.1			0.1	
30a	1075.9	2	6	5	1		a	Y	2153.8	53.3	2153.8146	53.6	2153.8									0.0	0.0			
30b	1075.9	2	6	5	1		b	Y	2153.8			58.7	2153.8										0.0			
31	1080	2	5	6	1	1	a	Y	2161.8										2162.0						0.2	
32a	1111.4	2	5	4		2	a	Y	2224.8	48.2	2224.8146	49	2224.8	50	2224.8	50	2224.8	49	2224.8	49	2224.8	0.0	0.0	0.0	0.0	0.0
32b	1111.4	2	5	4		2	b	Y	2224.8	53.5	2224.8146	55	2224.8	56	2224.8	55	2224.8	55	2224.8	54.6	2224.8	0.0	0.0	0.0	0.0	0.0
32c	1111.4	2	5	4		2	c	Y	2224.8			58.9	2224.8			60.1	2224.8	59	2224.8	59	2224.8		0.0		0.0	0.0
33a	1111.9	2	5	4	2	1	a	Y	2225.8			53.6	2225.8										0.0			
33b	1111.9	2	5	4	2	1	b	Y	2225.8	54.4	2225.8146	55.7	2225.8					55	2225.8			0.0	0.0			0.0
33c	1111.9	2	5	4	2	1	c	Y	2225.8			57.7	2225.8										0.0			
33d	1111.9	2	5	4	2	1	d	Y	2225.8			60.2	2225.8										0.0			
34	1112.5	2	4	5	2	1	b	Y	2226.8			55.8	2226.8										0.0			
35a	1140.4	2	5	5	1	1	a	Y	2282.8							48.4	2282.8								0.0	
35b	1140.4	2	5	5	1	1	b	Y	2282.8							52.8	2282.8								0.0	
36a	1184.5	2	5	4	1	2	a	Y	2370.8	51.4	2371.0146											0.2				
36b	1184.5	2	5	4	1	2	b	Y	2370.8	56	2371.0146	57	2370.8	58	2370.8	57	2371.0	58	2370.8		2371.0	0.2	0.0	0.0	0.2	0.0
36c	1184.5	2	5	4	1	2	c	Y	2370.8	60.1	2371.0146	61	2370.8	61	2370.8	61	2371.0	62	2370.8		2371.0	0.2	0.0	0.0	0.2	0.0
37a	1221.5	2	6	5	1	1	a	Y	2444.9	57.9	2445.0146	59.1	2445.0									0.1	0.1			
37b	1221.5	2	6	5	1	1	b	Y	2444.9	63.2	2445.0											0.1				
38	1258.5	2	7	6	1		a	Y	2518.9			60.3	2519.0										0.1			
39	1288	2	5	5	1	2	a	Y	2575.9							57.7	2574.0								0.1	
40a	1294	2	6	5		2	a	Y	2589.9			56.1	2590.0	56.4	2590.0			56	2590.0				0.1	0.1		0.1
40b	1294	2	6	5		2	b	Y	2589.9			60.3	2590.0					60	2590.0				0.1			0.1
41	1367	2	6	5	1	2	a	Y	2736.0				2736.0										0.1			
42a	1459.6	2	6	5		3	a	Y	2881.0			55	2881.2	55.1	2881.2	55.6	2881.2	55	2881.0	55.1	2881.2		0.2	0.2	0.2	0.0
42b	1459.6	2	6	5		3	b	Y	2881.0			59.6	2881.2	59.7	2881.2	60	2881.2	59	2881.0	59.8	2881.2		0.2	0.2	0.2	0.0
43a	1512.4	2	6	5	1	3	a	Y	3027.1			62.1	3028.8	62	3027.2	62	3027.0	62	3027.0	62	3027.2		-0.2	0.2	0.0	0.0
43b	1512.4	2	6	5	1	3	b	Y	3027.1									64	3027.0	63.1	3027.2					0.0
44a	1585.6	2	6	5	2	3	a	Y	3173.1			53.5	3173.2	54.3	3173.2		3173.2	54	3173.0				0.1	0.1	0.1	-0.1
44b	1585.6	2	6	5	2	3	b	Y	3173.1			58.6	3173.2	58.7	3173.2	59.2	3173.1	59	3173.0	58.5	3173.2		0.1	0.1	0.0	-0.1
45	1226.8	3	7	6	1	4	a	Y	3683.3			63.5	3685.4			64.5	3685.6	64	3685.4				0.1		0.1	0.1

	avg AUC= avergae area under curve of three techical replicates						RA=Relative abundance					
Delta mass	avgAUC_HMEC	avgAUC_MCF7	avgAUC_SKBR3	avgAUC_MDA157	avgAUC_MDA231	avgAUC_HS578T	RA_HMEC	RA_MCF7	RA_SKBR3	RA_MDA157	RA_MDA231	RA_HS578T
HS578T												
			1bb5185.33	4001435.00	321b1412.33				1.01	1.53	4.22	
				1.127384.b7	3b0947.33					0.40	0.73	
		5525118.b7	1808590.00	2448534.00	3235054.33			0.89	1.01	1.00	4.18	
	2.517930.00	30775851.b7	1755214.b7	5.1881129.b7	748409.00		0.61	4.23	0.90	1.40	0.97	
0.1	b.597975.b7	34438250.b7	6108517.00	7521896.00	1536149.b7	1289872.33	1.98	5.20	3.54	3.27	2.02	1.18
0.1	b.3992.74.33	30407561.b7	4b56136.33	5.994050.33	1760055.b7	849778.33	2.04	4.18	2.78	2.16	2.27	0.64
0.1	4931531.00	1.121.4724.33	2123072.b7	2.270230.b7	8096bb.00	812828.b7	1.49	1.69	1.27	1.03	1.04	0.73
	1.123253.b7						0.34					
	2097775.33	9191102.33		2.630600.b7	435855.b7		0.58	1.40		0.98	0.57	
				822845.33						0.51		
0.0	2503.6663.33	6.144493.b7			7b7239.00	817173.b7	b.99	0.94			0.99	0.67
		2945909.00	1181887.00		275806.33			0.49	0.77		0.33	
					7b1198.00						0.99	
0.0	1.424.1.659.00	3.2923.106.33	6874430.33	9.663844.00	25.54074.33	2.80872.b.47	4.54	4.94	4.13	3.55	3.30	2.11
	1.42b451.00	589803.00	1.47b823.00	896113.33	2545.16.00		0.45	0.09	0.85	0.31	0.30	
0.0	1.4011.21.33	3.295.698.b.7		1.435747.00	6.28494.00	651760.00	0.39	0.49		0.60	0.81	0.56
		7b73455.33	1.596.649.00					1.16	0.88			
		5.710253.b.7		2.29.9224.00				0.92		0.87		
				5.945711.33						2.19		
0.0	2.887.b.92b.33	9.2725.832.00	1.809.697.33	900357.b7	3.442955.b.7	4.84491.3.33	7.21	14.76	0.46	0.41	4.46	4.66
				5.1105.49.b.7						1.33		
		1.391.7450.00		4.433349.b7				2.23		1.63		
				3.047.28.b.7						0.96		
		893.850.00		6.486.691.33				1.14		2.72		
0.0	1.596.725.b.7	2.054.6.191.00	4295.549.33	5.749.534.33	10.66.158.33	477.612.b.7	0.30	5.06	2.59	1.49	1.37	0.39
		1.77.118.b.7	3.981.191.b.7	986.296.33	2.600.72.00			0.27	2.28	0.44	0.34	
0.0		1.048.6.100.00	2.477.866.b.7	5.790.361.b.7	6.758.97.00	6.66.157.b.7		1.66	1.61	2.24	0.88	0.65
				1.987.160.00						0.89		
	1.845.5504.b.7						4.65					
0.1	7.72.15.588.b.7	4.414.457.00	1.809.697.33	2.990403.00	1.474490.00	6.912.28.33	21.51	0.71	1.14	0.93	1.90	0.69
0.1	1.095.4706.33	1.151.b.489.b.7		3.743.085.00	25.91.251.b.7	5.11.6.217.b.7	2.89	1.83		1.53	5.34	5.21
		1.589.5.384.00		1.41.64.79.33	9.141.88.33			2.33		0.56	1.18	
		1.789.8.854.33						3.07				
		8.17.7.325.b.7		1.140.431.00				1.25		0.40		
		3.76.1.462.00						0.62				
				6.046.b.7b.b.7						2.79		
		5.08.3.564.33						0.78				
				1.608.983.00						0.61		
		1.91.4.497.b.7						0.29				
			1.015.920.00						0.55			
0.1	4.247.0.896.b.7	1.755.0.906.b.7	2.427.239.b.7	1.656.7.020.b.7	2.203.989.b.7	86.81.717.00	11.97	2.72	1.29	5.32	2.84	8.11
	4.94.8.678.33	1.124.1.004.b.7			917.999.33		1.33	1.82			1.19	
0.1	1.511.143.b.90	3.444.9.203.00	1.405.741.33	2.187.4.828.33	6.87.6.602.b.7	2.453.2.169.00	4.01	5.45	0.74	8.98	8.89	19.79
	3.737.1.06.33			4.88.85.21.b.7			1.04			1.95		
	1.41.6.1.90.33	2.49.1.854.00					0.37	0.41				
		7.45.9.6.20.00						1.18				
				3.1.55.b.45.33						1.28		
0.0	5.587.1.368.00	4.69.1.504.b.7	8.911.223.00	4.802.b.12.33	1.473.354.33	1.027.311.33	15.35	0.72	5.48	1.94	1.91	1.12
0.0	b.270.534.00	7.41.246.6.b.7	2.87.777.6.3.b.7	1.053.0.663.00	3.1.91.264.33	2.780.623.33	1.59	1.19	18.20	4.02	4.13	2.64
0.0		3.98.6.979.33		2.203.3.62.b.7	1.60.4.879.00	3.98.7.314.00		0.64		0.93	2.08	3.02
		5.69.6.522.33						0.80				
	9.48.95.84.33	5.33.5.796.00			1.05.8.518.33		2.52	0.87			1.34	
		7.25.4.111.00						0.70				
		5.68.9.592.33						0.87				
		3.65.9.179.b.7						0.61				
				1.395.437.b.7						0.66		
				3.587.528.00						1.59		
	4.028.752.00						1.11					
0.2	1.589.4213.b.7	1.154.4.486.33	1.142.444.1.00	1.518.781.5.00	291.3.555.b.7	1.041.8.919.33	3.68	1.83	6.08	5.53	3.83	8.49
0.2	2.135.8.1.81.33	3.319.9.121.00	6.405.876.00	2.247.6.212.b.7	98.54.07.6.b.7	3.307.1.810.33	0.58	5.32	3.29	10.22	12.82	26.97
	572.049.33	1.425.0.603.33					0.16	2.21				
	5.46.757.b.7						0.13					
		4.87.1.540.b.7						0.78				
				1.615.855.b.7						0.60		
		9.63.6.598.33	2.651.470.b.7		1.16.7.175.b.7			1.53	1.52		1.52	
		4.01.2.105.00			8.946.45.33			0.60			1.16	
		6.16.2.444.33						1.00				
0.2		1.1.1.11.704.00	2.01.77.544.33	1.3.500.742.b.7	35.70.285.b.7	8.23.7.586.33		1.77	12.58	5.00	4.59	6.42
0.2		1.752.8.893.33	2.990.622.4.00	1.922.9.596.b.7	75.68.146.00	2.145.778.b.7		2.70	17.01	7.59	9.77	1.62
0.2		7.78.2.623.33	3.195.743.33	3.81.4.018.b.7	7.596.78.b.7	1.697.734.00		1.20	1.90	1.79	0.98	1.36
0.2					3.503.50.00	1.53.8.249.b.7					0.43	0.97
		3.45.9.488.00	3.32.4.269.b.7		1.46.6.047.00			0.58	1.70		1.89	
0.1		5.11.6.828.33	7.33.1.376.33	7.162.1.34.33	2.706.392.b.7	2.47.6.661.33		0.78	4.46	2.86	3.52	2.01
		5.42.7.560.b.7		3.2.6.9.454.33	7.146.71.b.7			0.80		1.36	0.93	

(COND'T) Supplementary Table S1 - Released N-glycans identified from conditioned media of 8 breast epithelial cell lines by PGC-LC-MS/MS (CID)

Confidence	High	High	High	High	High	High	High	HIgh	High
Type	Paucimannose	Paucimannose	Paucimannose	High mannose	High mannose	High mannose	High mannose	Complex	Hybrid
Glycan no	1	2	3	4	5	6a	6b	7	8
									

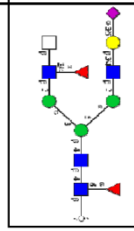
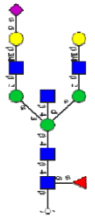
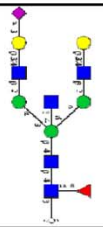
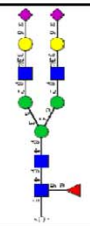
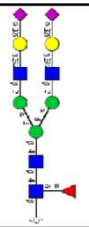
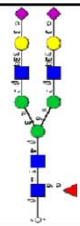
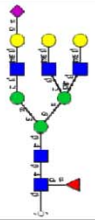
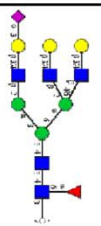
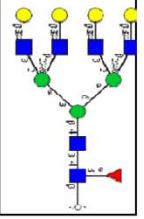
Confidence	HIgh	High	HIgh	HIgh	HIgh	High	High	High	High
Type	Hybrid	Complex	Complex	Complex	High mannose	Hybrid	Hybrid	Hybrid	Hybrid
Glycan no	9	10	11a	11b	12	13a	13b	14a	14b
									

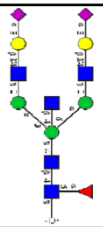
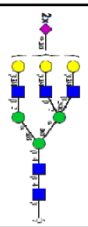
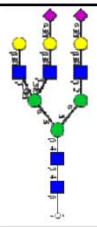
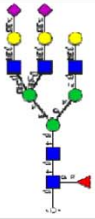
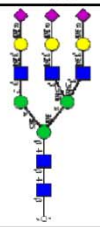
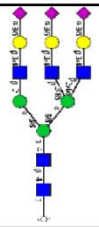
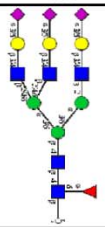
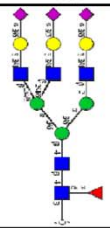
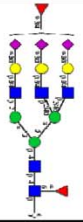
Confidence	High	High	High	High	Medium/High	High	High	High	High
Type	Hybrid	Complex	Complex	Complex	Hybrid	Hybrid	High mannose	Hybrid	Hybrid
Glycan no	15a	15b	16a	16b	17a	17b	18	19a	19b
									

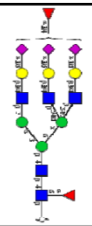
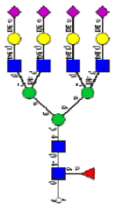
Confidence	High	High	High	High	High	High	High	High	High
Type	Hybrid	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	20	21a	21b	21c	22a	22b	23a	23b	24

Confidence	High	High	High	High	Medium/High	High	High	High	High
Type	Complex	Hybrid	High mannose	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	25	26	27	28a	28b	28c	28d	29	30a

Confidence	High	High	High	High	High	High	High	High	High
Type	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	30b	31	32a	32b	32c	33a	33b	33c	33d

Confidence	High	High	High	High	High	High	High	High	Medium/High
Type	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	34	35a	35b	36a	36b	36c	37a	37b	38
									

Confidence	High	High	High	Medium/High	Medium/High	High	High	High	Medium/High
Type	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	39	40a	40b	41	42a	42b	43a	43b	44a
									

Confidence	Medium/High	Medium/High
Type	Complex	Complex
Glycan no	44b	45
		

Supplemental Table S2

Supplemental Table S2. Number of characterized *N*-glycans grouped in the four *N*-glycan types from the six investigated breast cells.

Cell line	Number of structures in each *N*-glycan type				
	High mannose	Hybrid	Complex	Paucimannose	Total
HMEC	6	4	18	0	28
MCF7	7	9	36	1	53
SKBR3	6	5	15	2	28
MDA157	6	11	26	3	46
MDA231	6	7	23	3	39
HS578T	5	2	16	0	23

Supplemental Table S3

Supplemental S3a. List of significantly regulated secreted *N*-glycans that were commonly expressed between MCF7, SKBR3 and MDA231. Comparison between cell lines are expressed as fold change, where NA = not available; NC = no change. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Sialic acid linkage is indicated as α2,3-linked or α2,6-linked for mono-sialylated structures; α2,6/α2,6, α2,6/α2,3, α2,3/α2,3 for di-sialylated structures and ND = not determined.

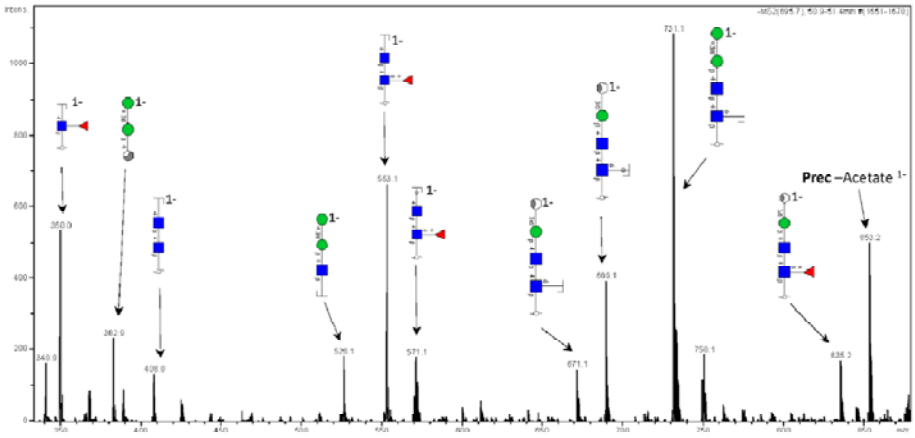
Glycan no	Composition [Core = $(HexNAc)_2(Hex)_3$]	Type	Sialic acid linkage	Core fucosylation	Terminal fucosylation	SKBR3 vs MCF7	MDA231 vs MCF7	MDA231 vs SKBR3
1	$(HexNAc)_2(Hex)_2(Fuc)_1$	Paucimannose	NA	Yes	No	NA	NA	↑4.2 fold***
3	Core + (Fuc)	Paucimannose	NA	Yes	No	NA	↑4.7 fold****	↑4.1 fold****
13a	Core + $(HexNAc)_1(Hex)_2(NeuAc)_1$	Hybrid	α2,6	No	No	↑9.9 fold****	NC	↓3.0 fold****
15b	Core + $(HexNAc)_2(Hex)_2(Fuc)_1$	Complex	NA	Yes	No	↓32.2	↓3.3 fold****	↑9.7 fold**
19a	Core + $(HexNAc)_1(Hex)_3(NeuAc)_1$	Hybrid	α2,6	No	No	↑9.6 fold***	NC	↓6.8 fold***
28d	Core + $(HexNAc)_2(Hex)_2(Fuc)_1(NeuAc)_1$	Complex	α2,3	Yes	No	↓7.4 fold*	NC	↑12.0 fold***
32a	Core + $(HexNAc)_2(Hex)_2(NeuAc)_2$	Complex	α2,6/α2,6	No	No	↑7.6 fold*	NC	NC
32b	Core + $(HexNAc)_2(Hex)_2(NeuAc)_2$	Complex	α2,6/α2,3	No	No	↑15.3 fold***	NC	↓4.4 fold**
32c	Core + $(HexNAc)_2(Hex)_2(NeuAc)_2$	Complex	α2,3/α2,3	No	No	NA	↑3.3 fold**	NA
36b	Core + $(HexNAc)_2(Hex)_2(Fuc)_1(NeuAc)_2$	Complex	α2,6/α2,3	Yes	No	↑3.3 fold*	NC	NC
36c	Core + $(HexNAc)_2(Hex)_2(Fuc)_1(NeuAc)_2$	Complex	α2,3/α2,3	Yes	No	NC	NC	↑3.9 fold**
42a	Core + $(HexNAc)_3(Hex)_3(NeuAc)_3$	Complex	ND	No	No	↑7.1 fold***	NC	↓3.0 fold**
42b	Core + $(HexNAc)_3(Hex)_3(NeuAc)_3$	Complex	ND	No	No	↑6.3 fold****	↑3.6 fold***	NC
44b	Core + $(HexNAc)_3(Hex)_3(Fuc)_2(NeuAc)_3$	Complex	ND	Yes	Yes	↑5.7 fold*	NC	NC

Supplementary Table S3b. List of secreted N-glycans that were unique expressed MCF7 and not found in SKBR3 and MDA231. Sialic acid linkage is indicated as α2,3-linked or α2,6-linked for mono-sialylated structures, NA = not available; ND = not determined.

Glycan no	Composition [Core = $(HexNAc)_2(Hex)_3$]	Type	Sialic acid linkage	Core Fucosylation	Terminal Fucosylation
14b	Core + $(HexNAc)_1(Hex)_3(Fuc)_1$	Hybrid	NA	Yes	No
16b	Core + $(HexNAc)_3(Hex)_1(Fuc)_1$	Complex	NA	Yes	No
17b	Core + $(HexNAc)_1(Hex)_2(Fuc)_1(NeuAc)_1$	Complex	α2,3	Yes	No
22b	Core + $(HexNAc)_2(Hex)_2(Fuc)_2$	Complex	NA	Yes	Yes
23a	Core + $(HexNAc)_3(Hex)_1(Fuc)_2$	Complex	NA	Yes	Yes
23b	Core + $(HexNAc)_3(Hex)_1(Fuc)_2$	Complex	NA	Yes	Yes
25	Core + $(HexNAc)_4(Fuc)_2$	Complex	NA	Yes	Yes
27	Core + $(Hex)_7$	High Mannose	NA	NA	NA
30a	Core + $(HexNAc)_3(Hex)_3(Fuc)_1$	Complex	NA	Yes	No
30b	Core + $(HexNAc)_3(Hex)_3(Fuc)_1$	Complex	NA	Yes	No
33a	Core + $(HexNAc)_2(Hex)_2(Fuc)_2(NeuAc)_1$	Complex	α2,6	Yes	Yes
33c	Core + $(HexNAc)_2(Hex)_2(Fuc)_2(NeuAc)_1$	Complex	α2,3	Yes	Yes
33d	Core + $(HexNAc)_3(Hex)_1(Fuc)_2(NeuAc)_1$	Complex	α2,3	Yes	Yes
34	Core + $(HexNAc)_3(Hex)_1(Fuc)_2(NeuAc)_1$	Complex	ND	Yes	Yes
37a	Core + $(HexNAc)_3(Hex)_3(Fuc)_1(NeuAc)_1$	Complex	α2,6	Yes	No
38	Core + $(HexNAc)_4(Hex)_4(Fuc)_1$	Complex	NA	Yes	No
41	Core + $(HexNAc)_3(Hex)_3(Fuc)_1(NeuAc)_2$	Complex	ND	Yes	No

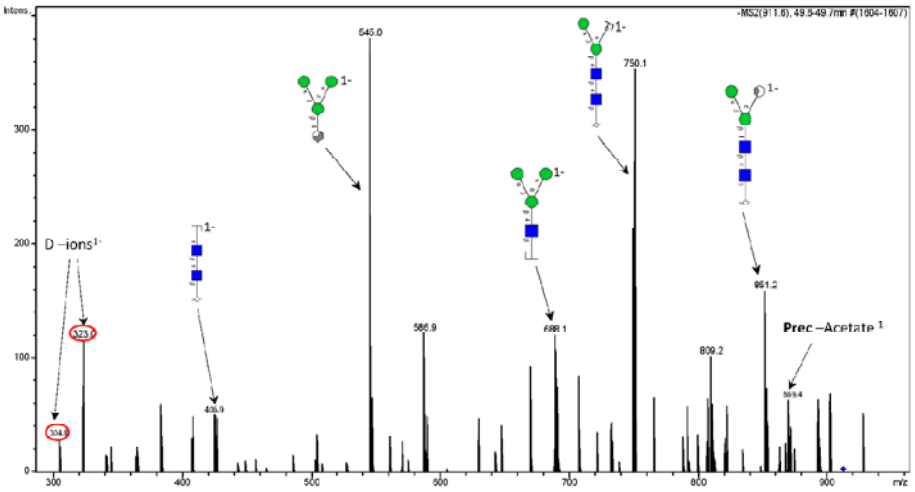
APPENDIX 4

Glycan No 1
Type = Pauciamannose
Precursor = *m/z* 895.3^{1-}
$[M-H]^{1-}$ = 895.3Da
LC retention time = 51.1 min



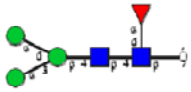
Positive match to MS2 spectrum in UnicarbKB

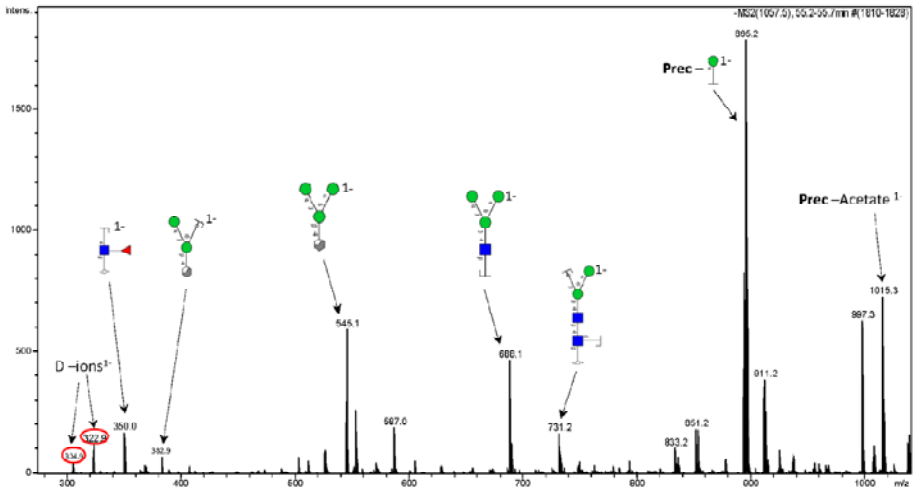
Glycan No 2
Type = Pauciamannose
Precursor = *m/z* 911.3^{1-}
$[M-H]^{1-}$ = 911.3Da
LC retention time = 49.6 min



Positive match to MS2 spectrum in UnicarbKB

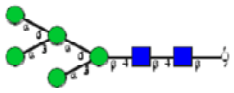
Glycan No 3
Type = Pauciamannose
Precursor = *m/z* 1057.3^{1-}
[M-H]$^{1-}$ = 1057.3Da
LC retention time = 55.4 min

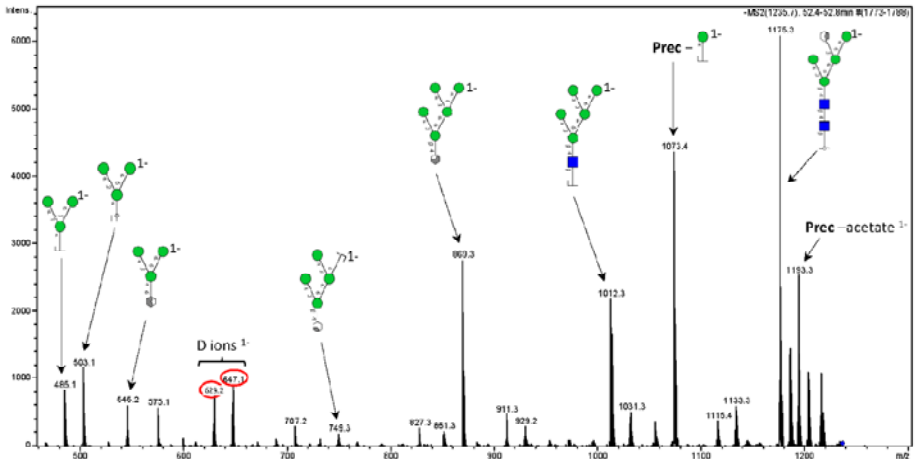




Positive match to MS2 spectrum in UnicarbKB

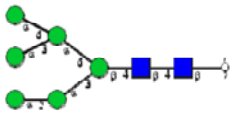
Glycan No 4
Type = High Mannose
Precursor = *m/z* 1235.3^{1-}
[M-H]$^{1-}$ = 1235.3 Da
LC retention time = 52 min

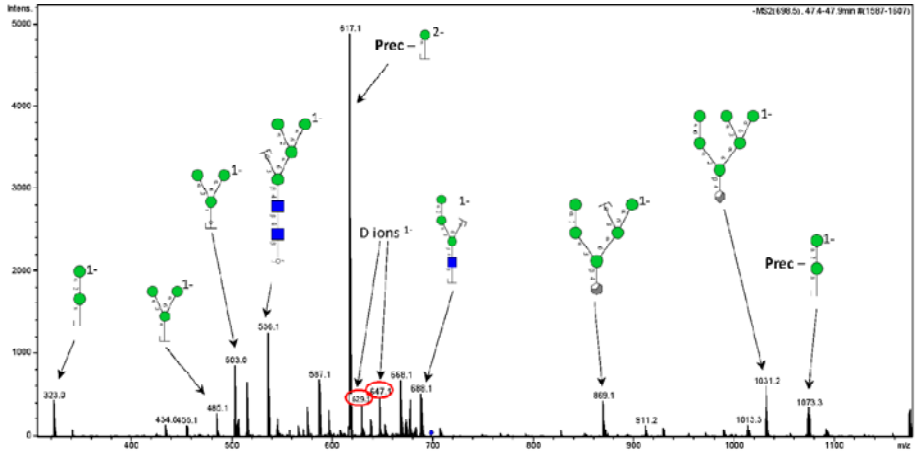




Positive match to MS2 spectrum in UnicarbKB

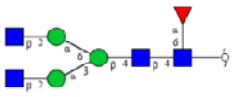
Glycan No 5
Type = High Mannose
Precursor = *m/z* 698.2^{2-}
[M-H]$^{1-}$ = 1397.4 Da
LC retention time = 47.6 min

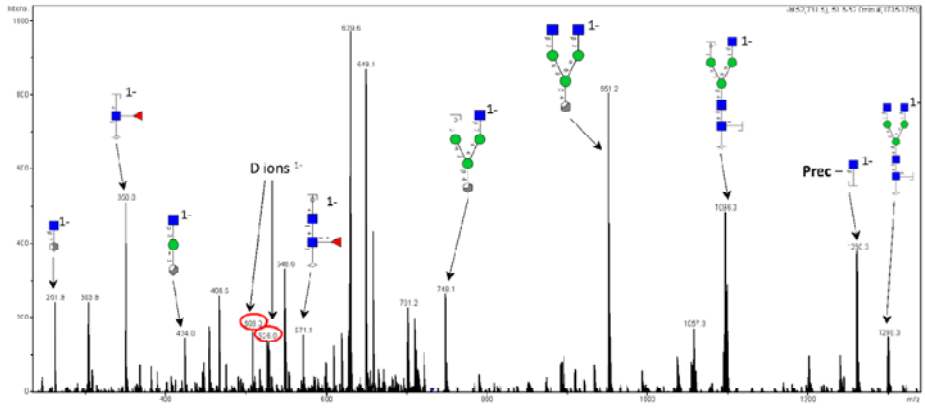




Positive match to MS2 spectrum in UnicarbKB

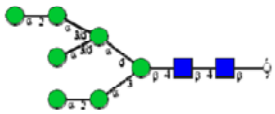
Glycan No 6
Type = Complex
Precursor = *m/z* 731.2^{2-}
[M-H]$^{1-}$ = 1463.4 Da
LC retention time = 51.8 min

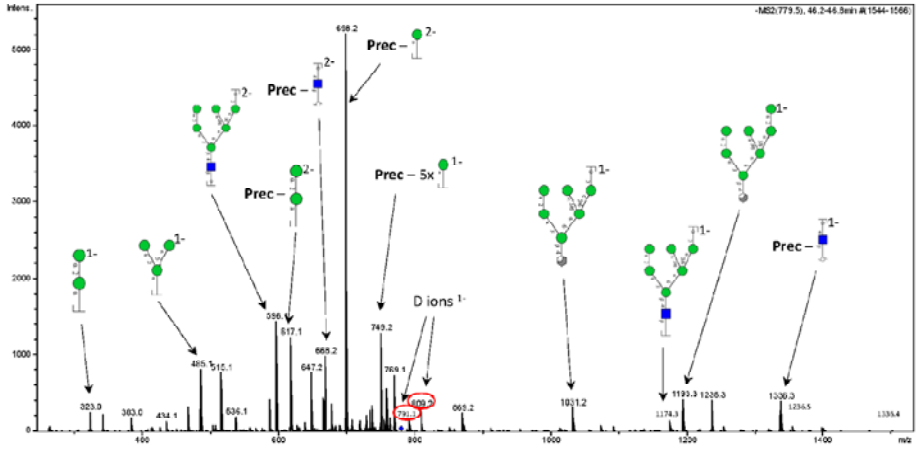




No match to MS2 spectrum in UnicarbKB

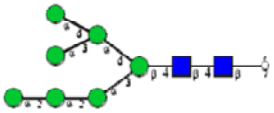
Glycan No 7a
Type = High Mannose
Precursor = *m/z* 779.3^{2-}
[M-H]$^{1-}$ = 1559.6 Da
LC retention time = 46.5 min

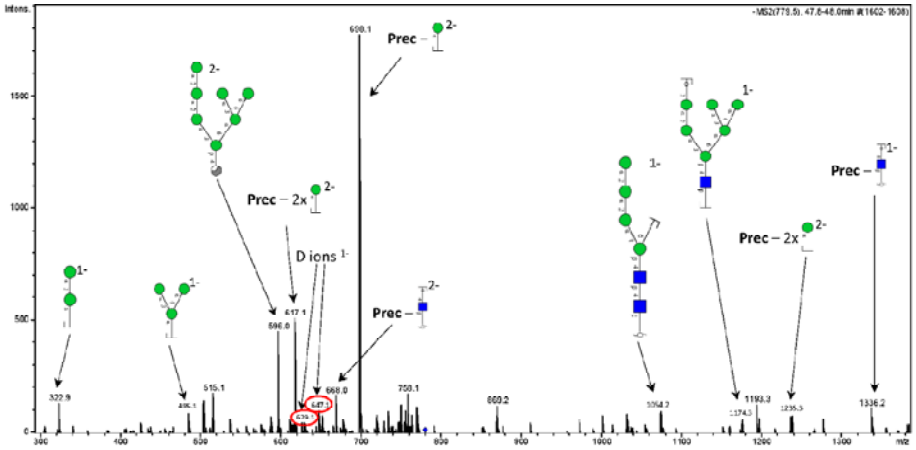




Positive match to MS2 spectrum in UnicarbKB

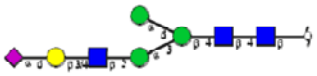
Glycan No 7b
Type = High Mannose
Precursor = *m/z* 779.3^{2-}
[M-H]$^{1-}$ = 1559.6 Da
LC retention time = 47.8 min

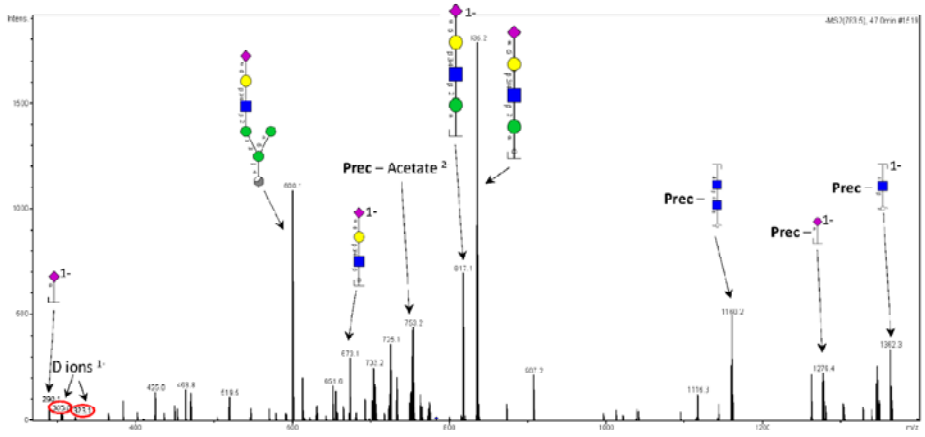




Positive match to MS2 spectrum in UnicarbKB

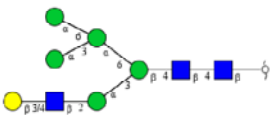
Glycan No 8
Type = Hybrid
Precursor = *m/z* 783.3^{2-}
[M-H]$^{1-}$ = 1567.6 Da
LC retention time = 47.0 min

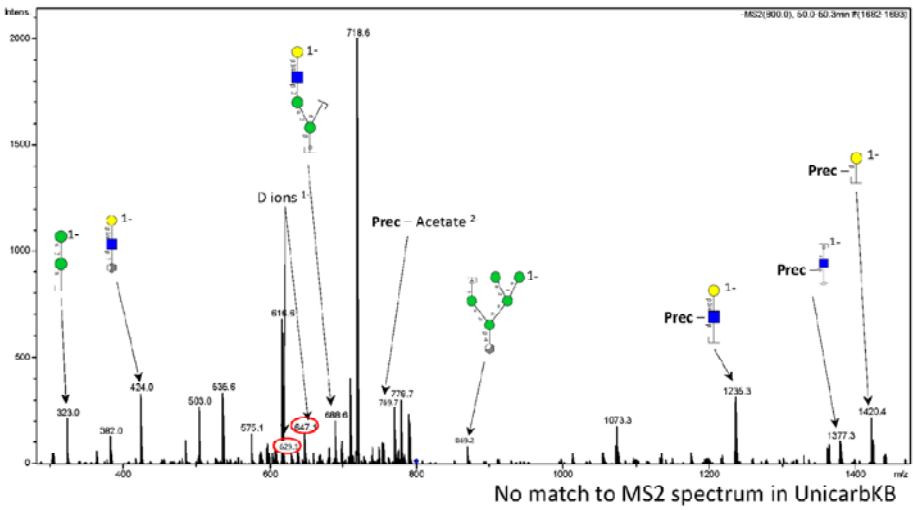




No match to MS2 spectrum in UnicarbKB

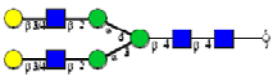
Glycan No 9
Type = Hybrid
Precursor = *m/z* 799.8^{2-}
[M-H]$^{1-}$ = 1600.6 Da
LC retention time = 50.1 min

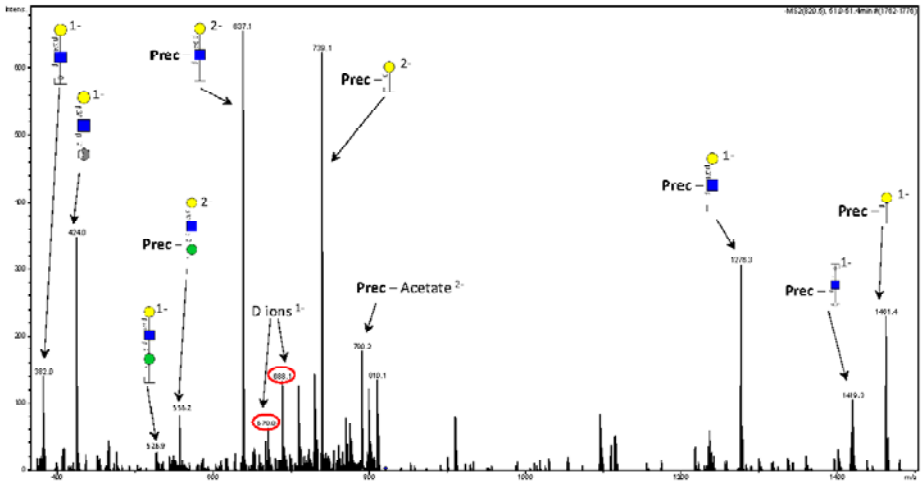




No match to MS2 spectrum in UnicarbKB

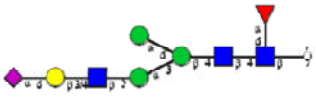
Glycan No 10
Type = Complex
Precursor = *m/z* 820.3^{2-}
[M-H]$^{1-}$ = 1641.6 Da
LC retention time = 51.2 min

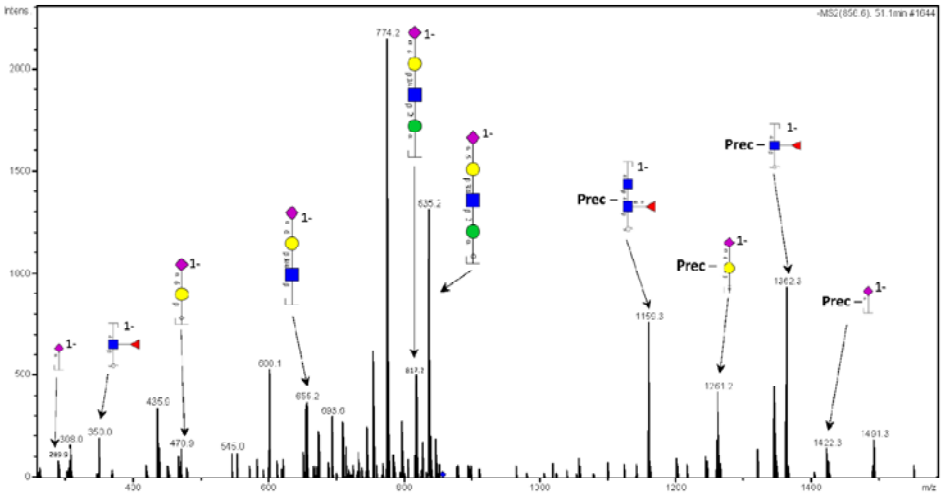




Positive match to MS2 spectrum in UnicarbKB

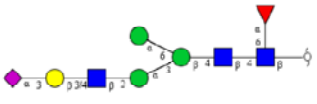
Glycan No 11a
Type = Complex
Precursor = *m/z* 856.3^{2-}
[M-H]$^{1-}$ = 1713.6 Da
LC retention time = 51.1 min

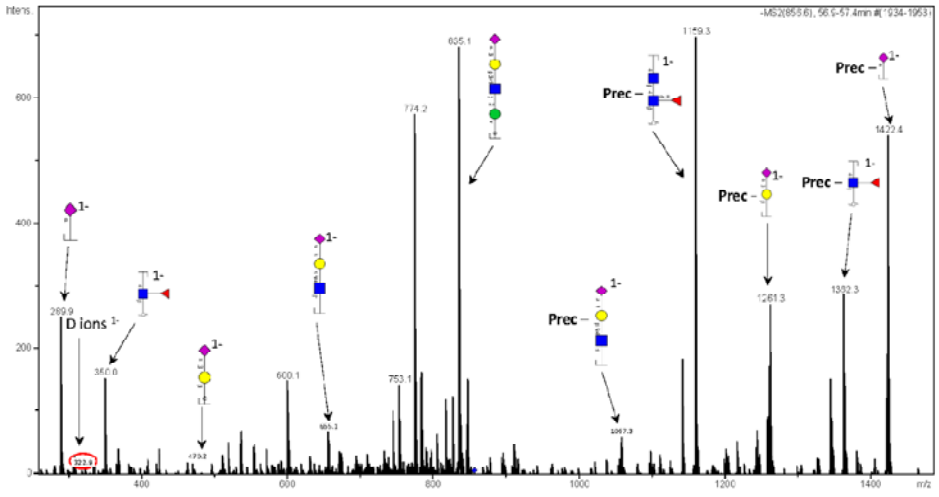




No match to MS2 spectrum in UnicarbKB

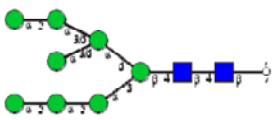
Glycan No 11b
Type = Hybrid
Precursor = *m/z* 856.3^{2-}
[M-H]$^{1-}$ = 1713.6 Da
LC retention time = 57.1 min

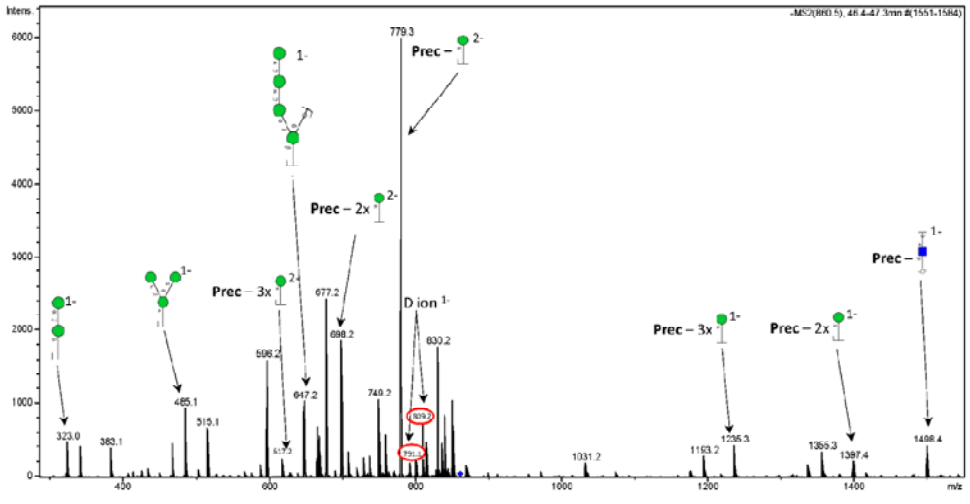




No match to MS2 spectrum in UnicarbKB

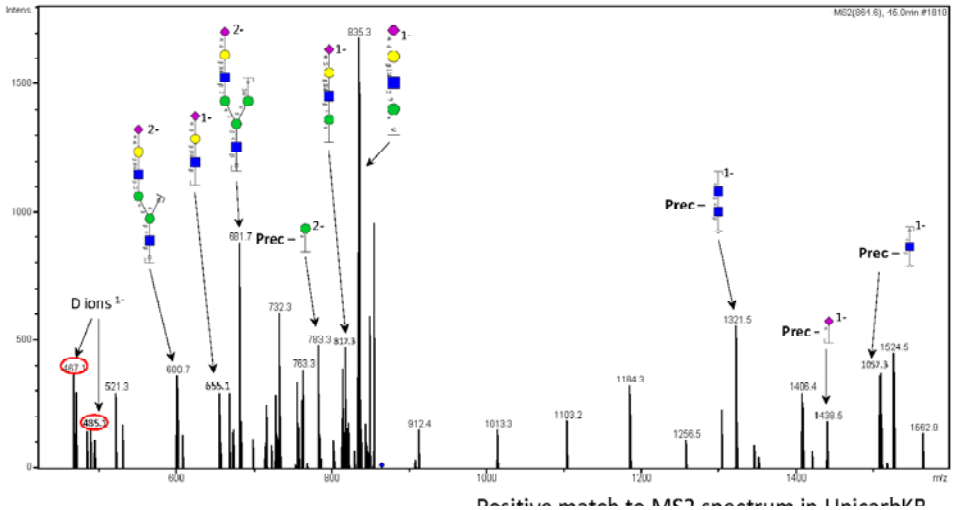
Glycan No 12
Type = High Mannose
Precursor = *m/z* 860.3^{2-}
[M-H]$^{1-}$ = 1721.6 Da
LC retention time = 46.7 min



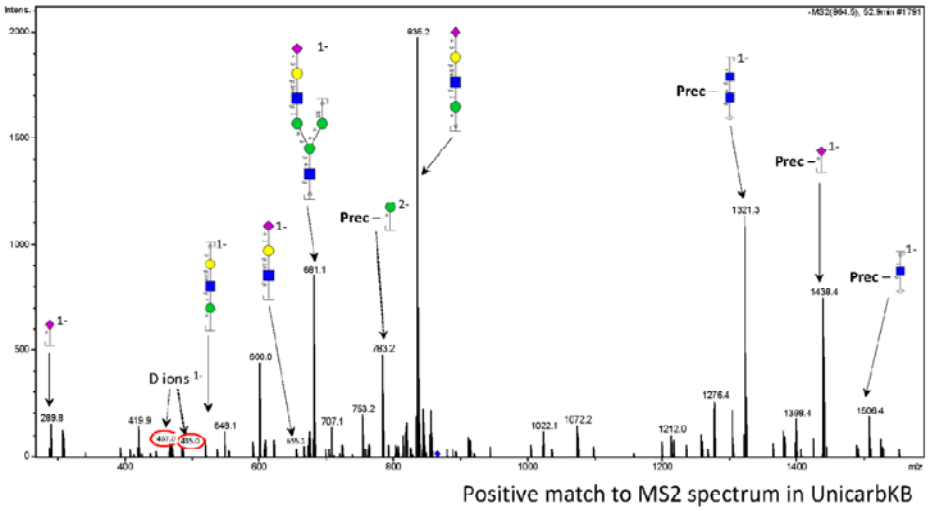


Positive match to MS2 spectrum in UnicarbKB

Glycan No 13a
Type = Hybrid
Precursor = *m/z* 864.4^{2-}
[M-H]$^{1-}$ = 1729.8 Da
LC retention time = 45 min

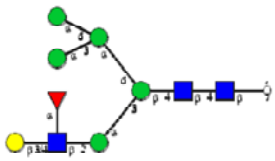


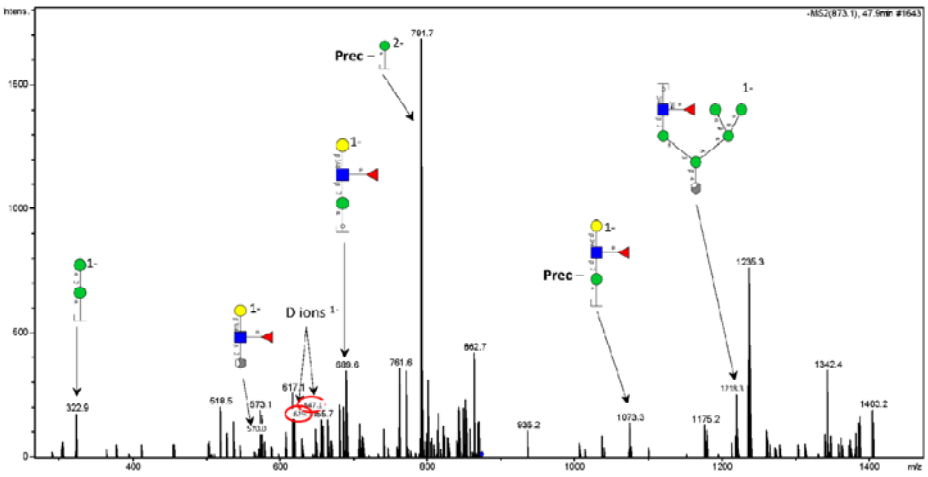
Glycan No 13b
Type = Hybrid
Precursor = *m/z* 864.4^{2-}
[M-H]$^{1-}$ = 1729.8 Da
LC retention time = 53.0 min



Positive match to MS2 spectrum in UnicarbKB

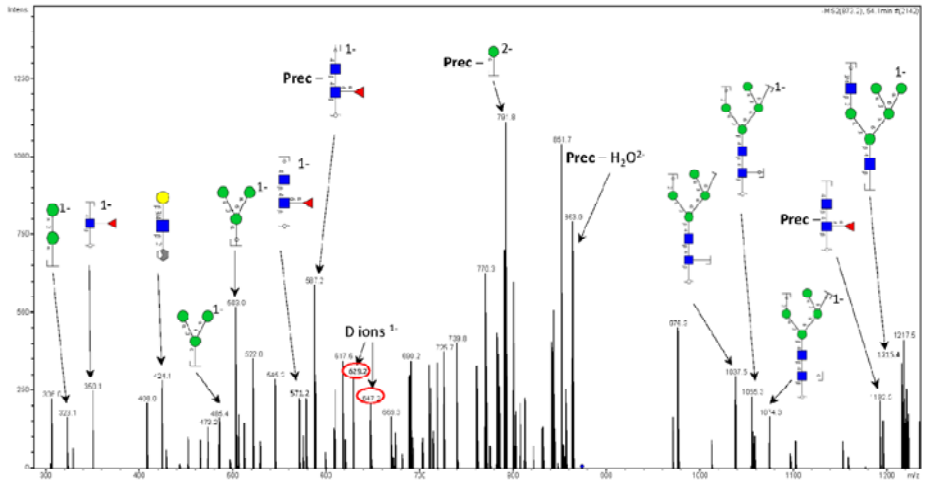
Glycan No 14a
Type = Hybrid
Precursor = *m/z* 872.8^{2-}
$[M-H]^{1-}$ = 1746.6 Da
LC retention time = 47.9 min





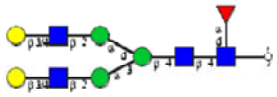
No match to MS2 spectrum in UnicarbKB

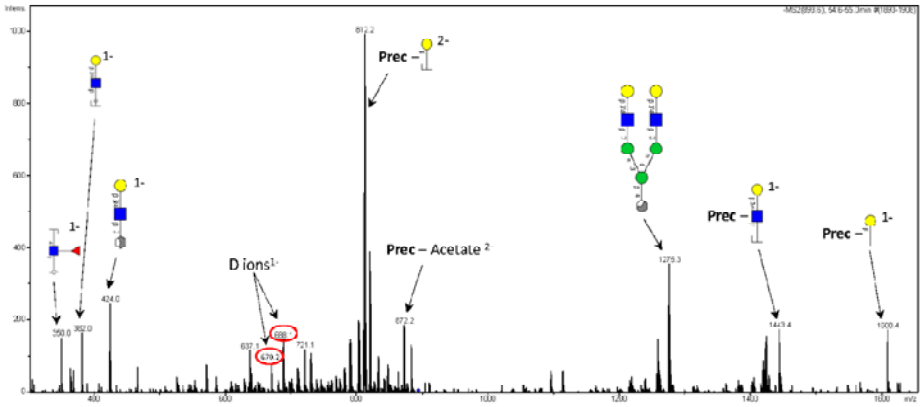
Glycan No 14b
Type = Hybrid
Precursor = *m/z* 872.9^{2-}
$[M-H]^{1-}$ = 1746.8 Da
LC retention time = 54.1 min



No match to MS2 spectrum in UnicarbKB

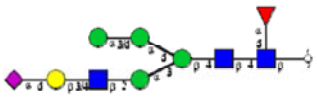
Glycan No 15
Type = Complex
Precursor = *m/z* 893.3^{2-}
[M-H]$^{1-}$ = 1787.6 Da
LC retention time = 54.8 min

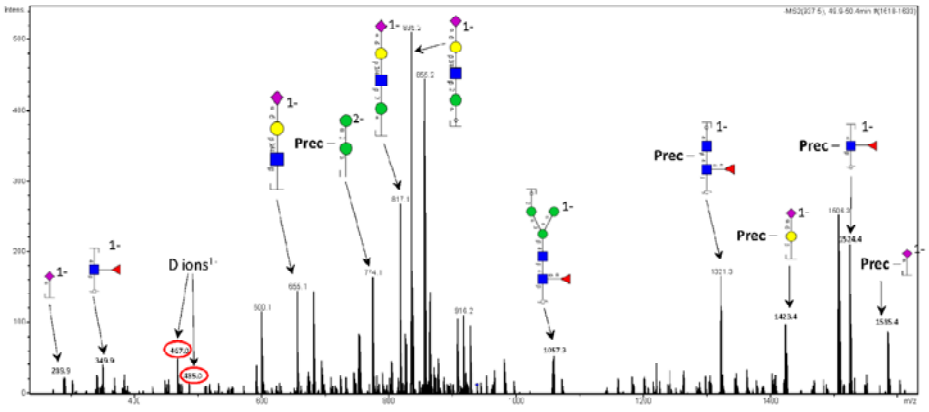




Positive match to MS2 spectrum in UnicarbKB

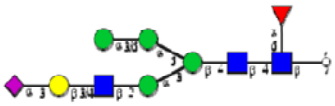
Glycan No 16a
Type = Hybrid
Precursor = *m/z* 937.3^{2-}
[M-H]$^{1-}$ = 1875.6 Da
LC retention time = 50.3 min

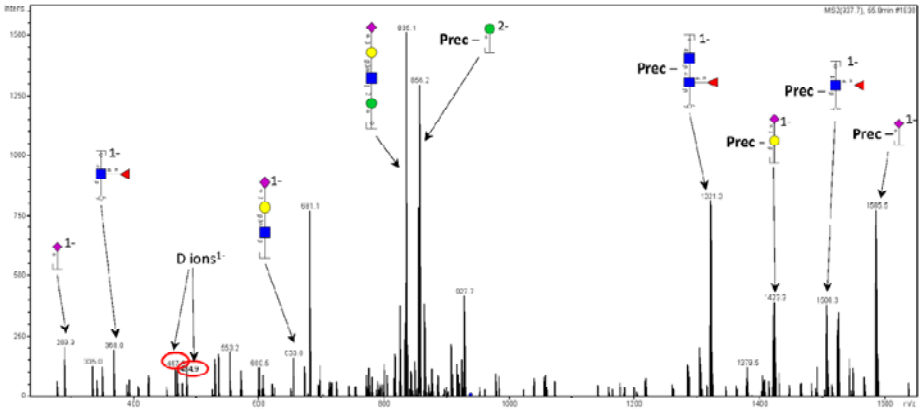




Positive match to MS2 spectrum in UnicarbKB

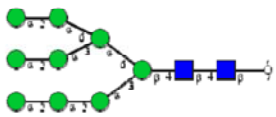
Glycan No 16b
Type = Hybrid
Precursor = *m/z* 937.3^{2-}
[M-H]$^{1-}$ = 1875.6 Da
LC retention time = 55.8 min

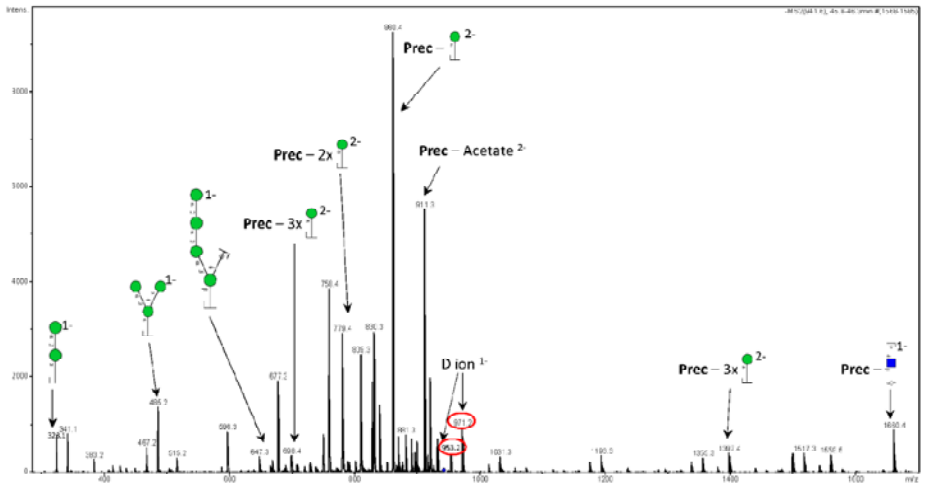




Positive match to MS2 spectrum in UnicarbKB

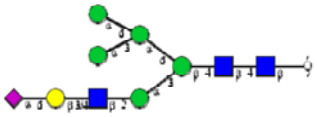
Glycan No 17
Type = High Mannose
Precursor = *m/z* 941.4^{2-}
[M-H]$^{1-}$ = 1883.8 Da
LC retention time = 46.0 min

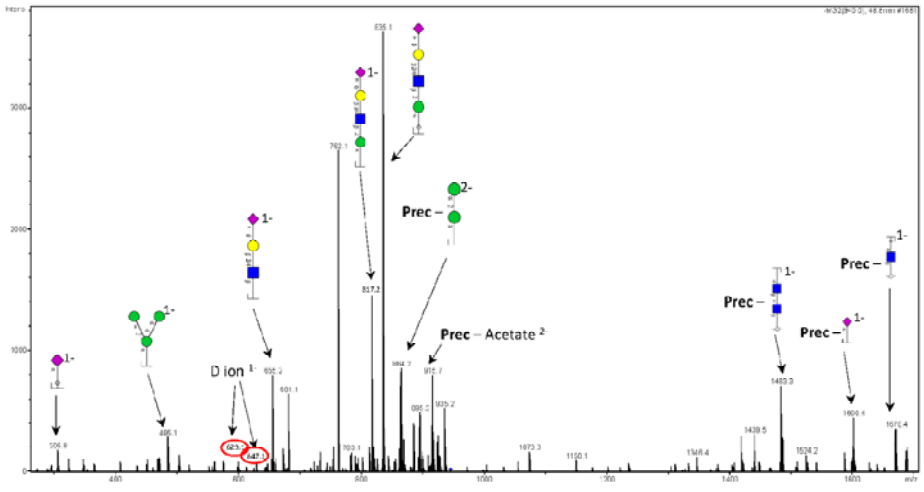




Positive match to MS2 spectrum in UnicarbKB

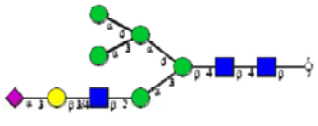
Glycan No 18a
Type = High Mannose
Precursor = *m/z* 945.4^{2-}
[M-H]$^{1-}$ = 1891.8 Da
LC retention time = 48.9 min

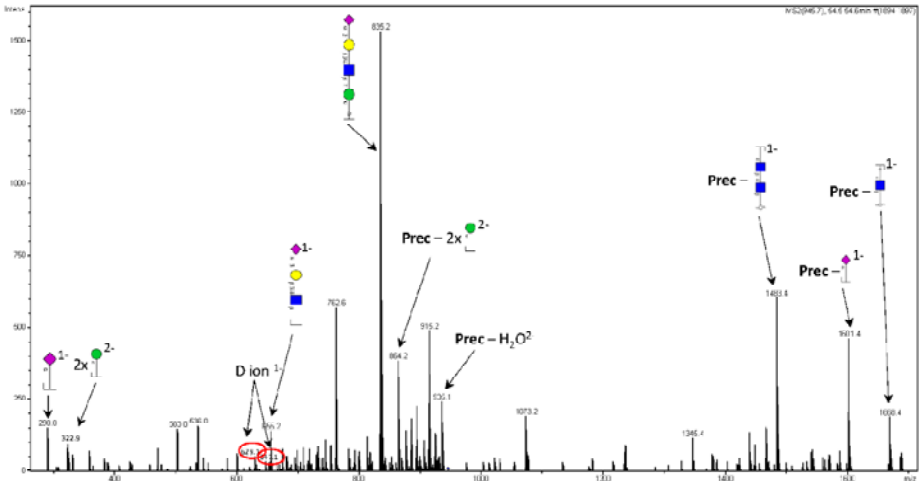




Positive match to MS2 spectrum in UnicarbKB

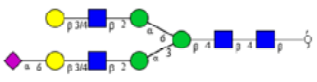
Glycan No 18b
Type = High Mannose
Precursor = *m/z* 945.4^{2-}
[M-H]$^{1-}$ = 1891.8 Da
LC retention time = 54.5 min

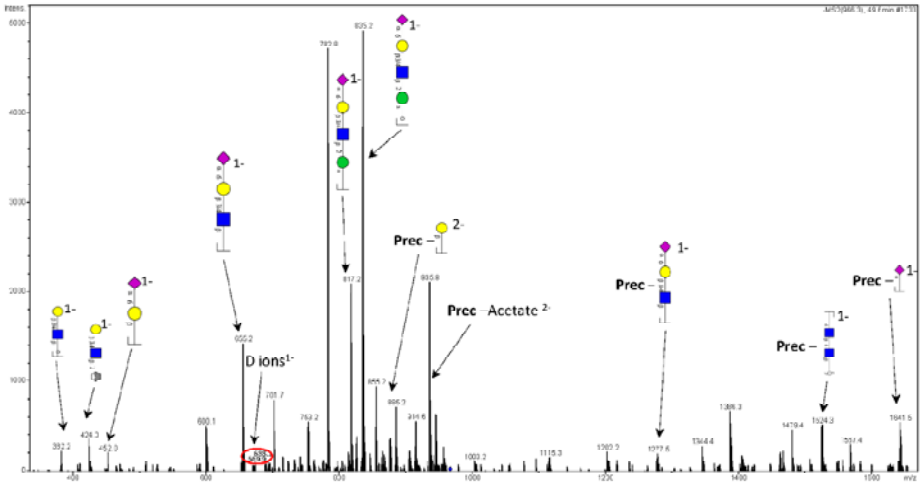




Positive match to MS2 spectrum in UnicarbKB

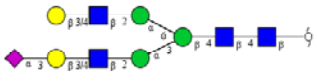
Glycan No 19a
Type = Complex
Precursor = *m/z* 965.9^{2-}
[M-H]$^{1-}$ = 1932.8Da
LC retention time = 49.6 min

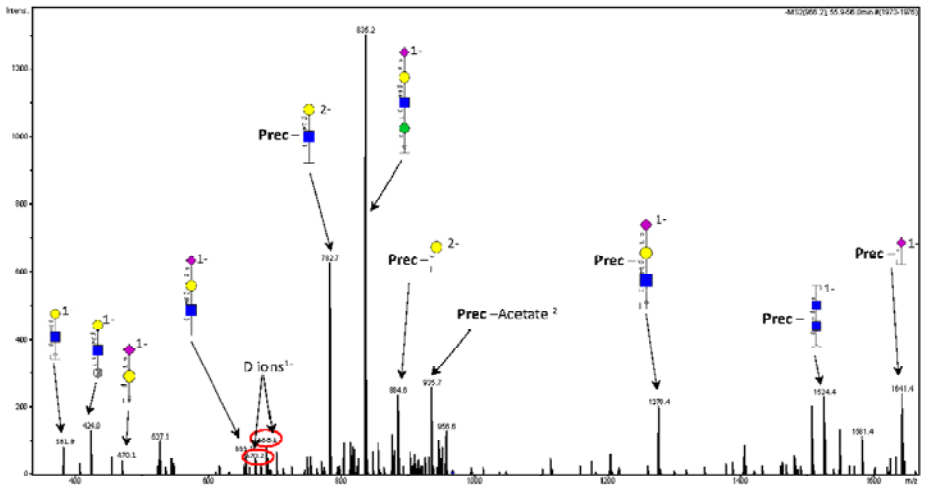




Positive match to MS2 spectrum in UnicarbKB

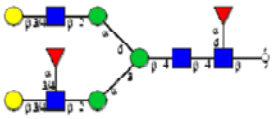
Glycan No 19b
Type = Complex
Precursor = *m/z* 965.9^{2-}
[M-H]$^{1-}$ = 1932.8Da
LC retention time = 55.9 min

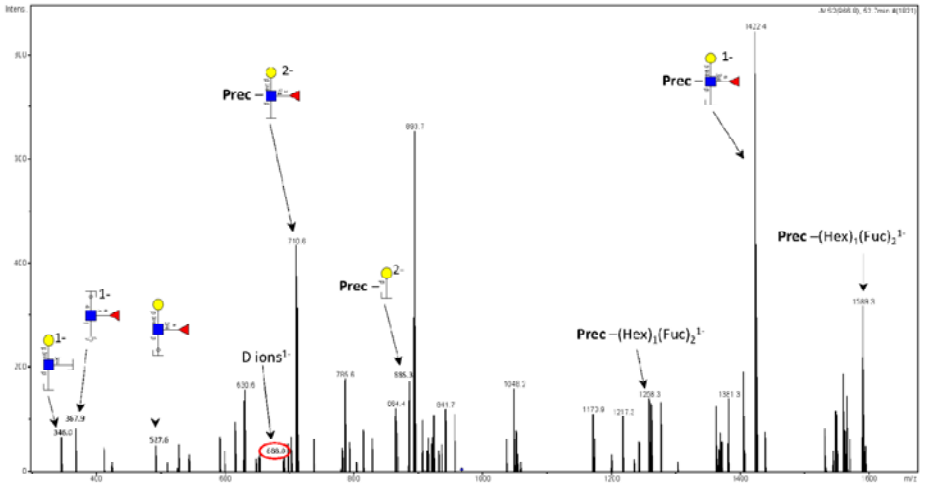




Positive match to MS2 spectrum in UnicarbKB

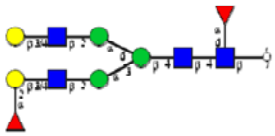
Glycan No 20a
Type = Complex
Precursor = *m/z* 966.4^{2-}
[M-H]$^{1-}$ = 1933.8Da
LC retention time = 52.2 min

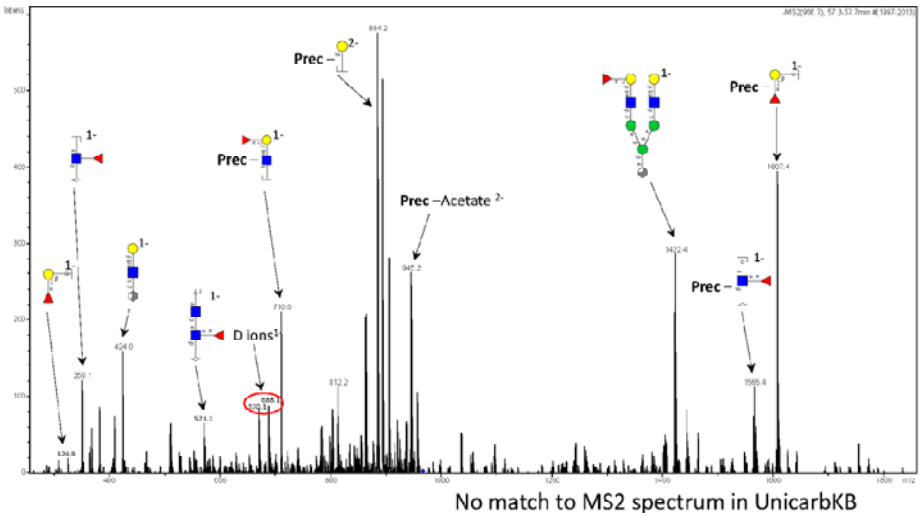




No match to MS2 spectrum in UnicarbKB

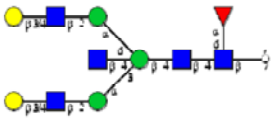
Glycan No 20b
Type = Complex
Precursor = *m/z* 966.4^{2-}
[M-H]$^{1-}$ = 1933.8Da
LC retention time = 57.5 min

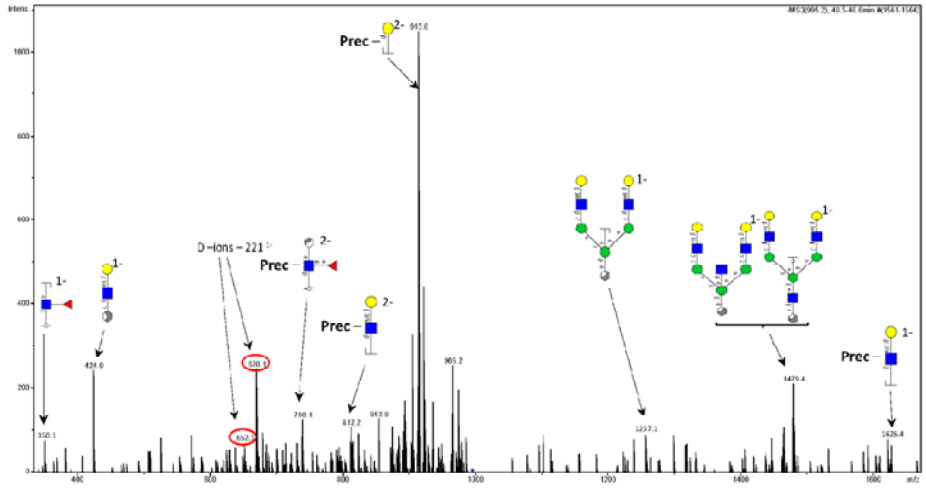




No match to MS2 spectrum in UnicarbKB

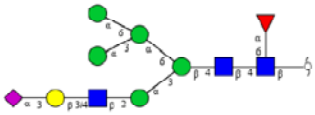
Glycan No 21
Type = Complex
Precursor = *m/z* 994.9^{2-}
[M-H]$^{1-}$ = 1990.8 Da
LC retention time = 48.4 min

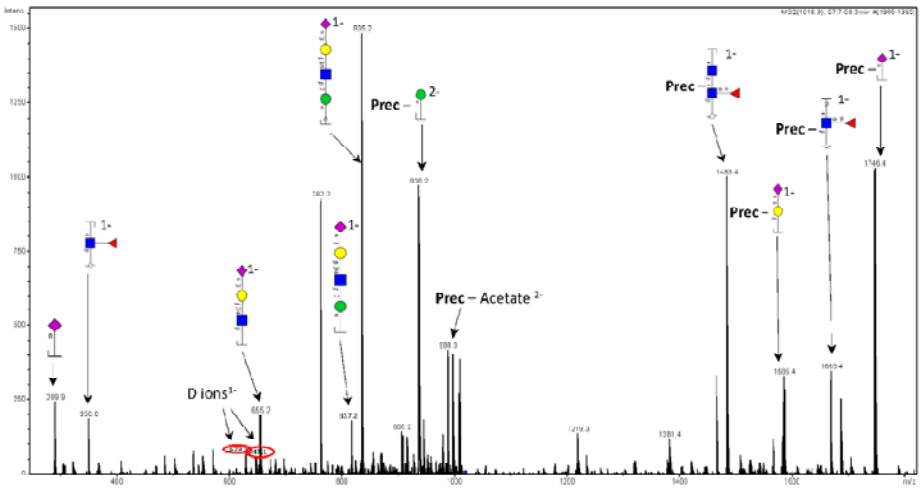




Positive match to MS2 spectrum in UnicarbKB

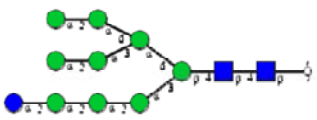
Glycan No 22
Type = Hybrid
Precursor = *m/z* 1018.4^{2-}
[M-H]$^{1-}$ = 2037.8 Da
LC retention time = 57.9 min

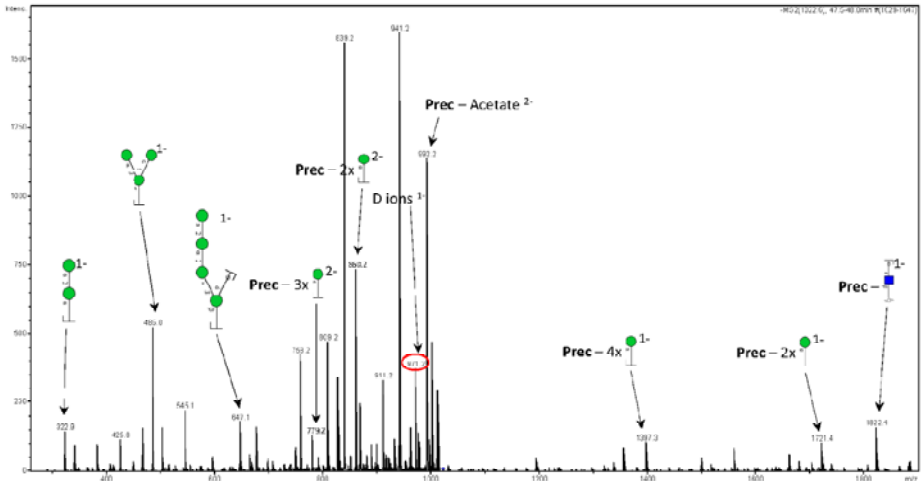




No match to MS2 spectrum in UnicarbKB

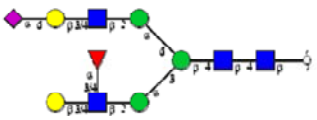
Glycan No 23
Type = High Mannose
Precursor = *m/z* 1022.4^{2-}
[M-H]$^{1-}$ = 2045.8 Da
LC retention time = 48 min

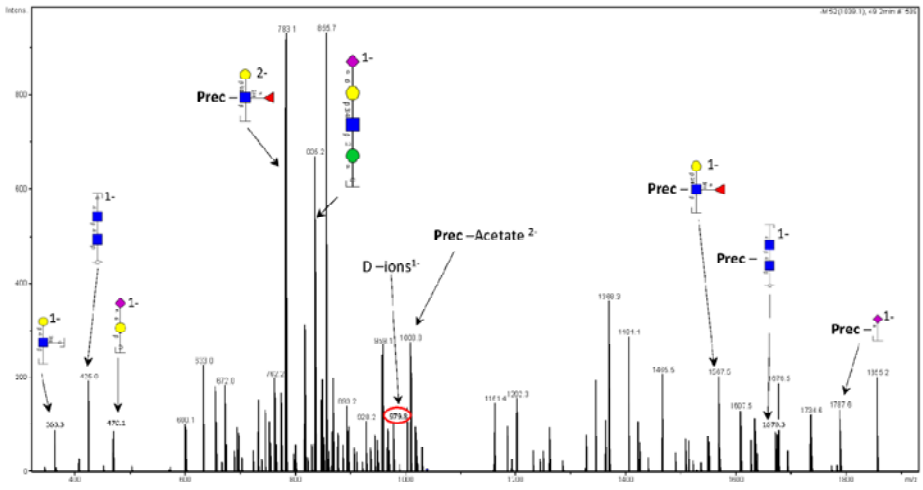




Positive match to MS2 spectrum in UnicarbKB

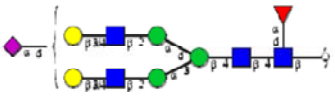
Glycan No 24a
Type = Complex
Precursor = *m/z* 1038.9^{2-}
[M-H]$^{1-}$ = 2078.8 Da
LC retention time = 49.2 min

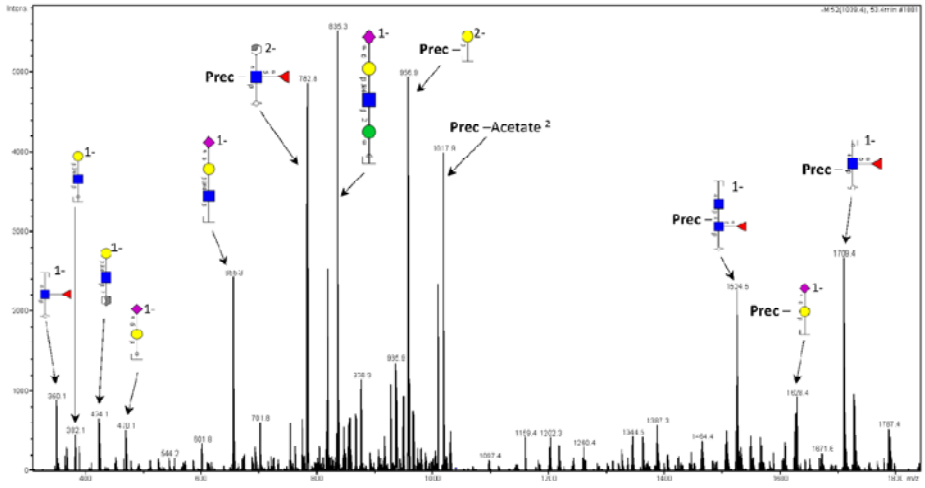




No match to MS2 spectrum in UnicarbKB

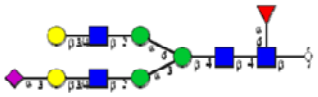
Glycan No 24b
Type = Complex
Precursor = *m/z* 1038.9^{2-}
[M-H]$^{1-}$ = 2078.8 Da
LC retention time = 53.4 min

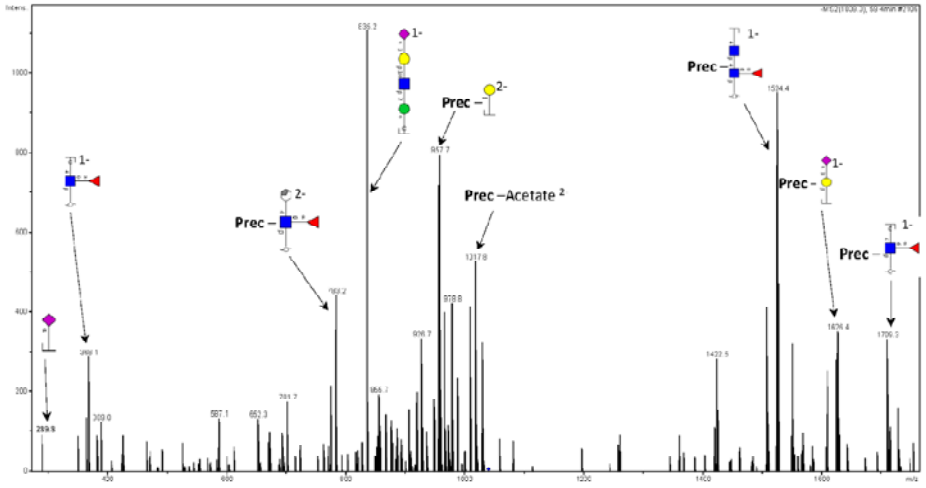




Positive match to MS2 spectrum in UnicarbKB

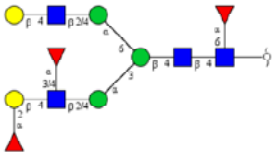
Glycan No 24c
Type = Complex
Precursor = *m/z* 1038.9^{2-}
[M-H]$^{1-}$ = 2078.8 Da
LC retention time = 59.4 min

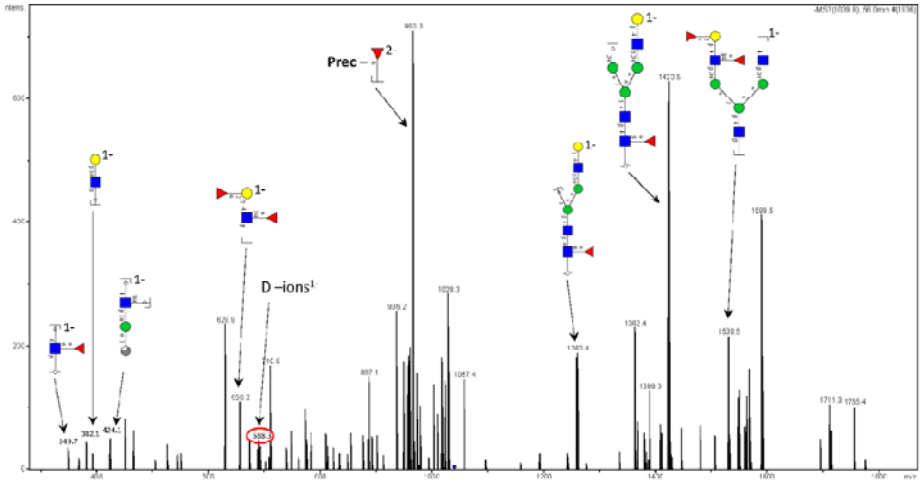




Positive match to MS2 spectrum in UnicarbKB

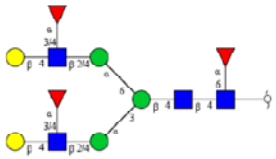
Glycan No 25a
Type = Complex
Precursor = *m/z* 1039.3^{2-}
$[M-H]^{1-}$ = 2079.6 Da
LC retention time = 56.0 min

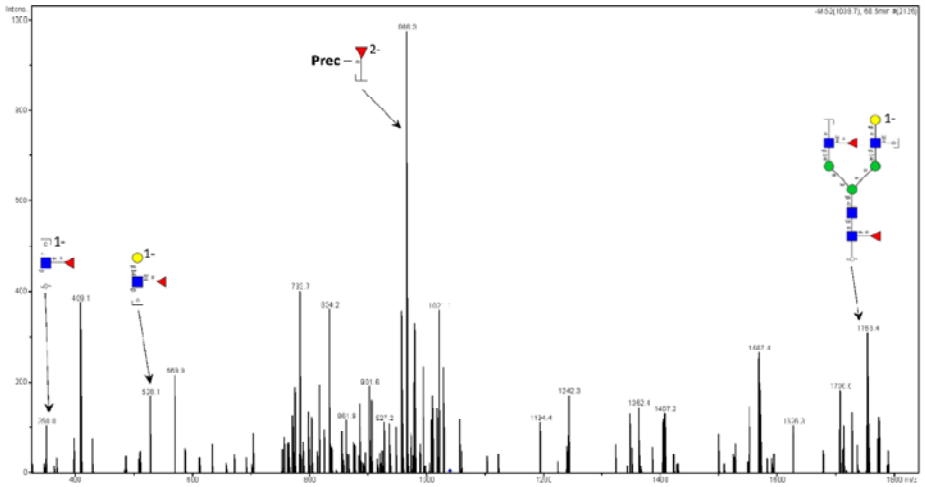




No match to MS2 spectrum in UnicarbKB

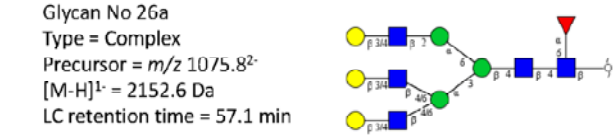
Glycan No 25b
Type = Complex
Precursor = *m/z* 1039.3^{2-}
$[M-H]^{1-}$ = 2079.6 Da
LC retention time = 60.5 min

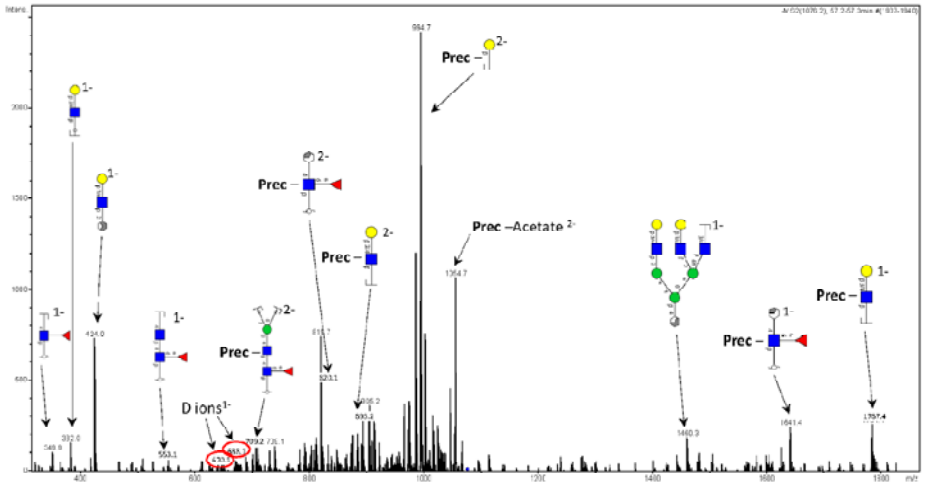




No match to MS2 spectrum in UnicarbKB

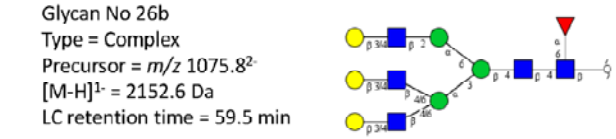
Glycan No 26a
Type = Complex
Precursor = *m/z* 1075.8^{2-}
[M-H]$^{1-}$ = 2152.6 Da
LC retention time = 57.1 min

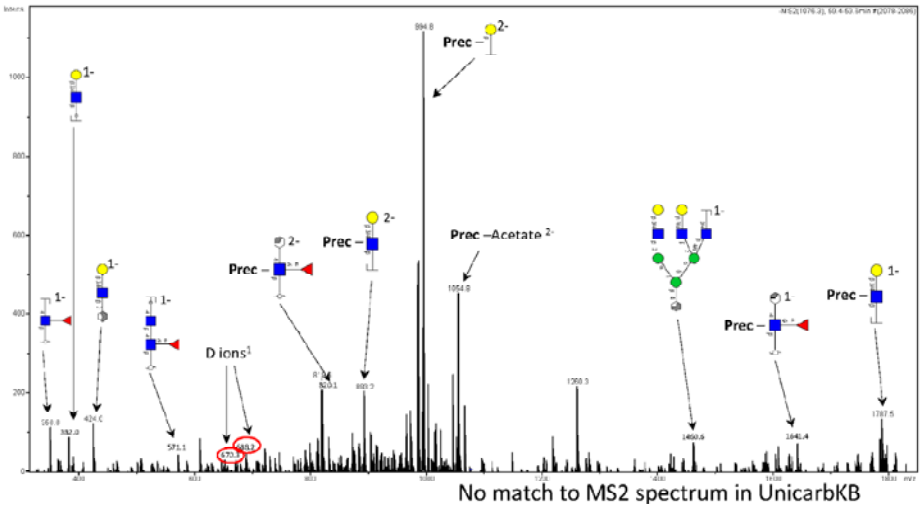




No match to MS2 spectrum in UnicarbKB

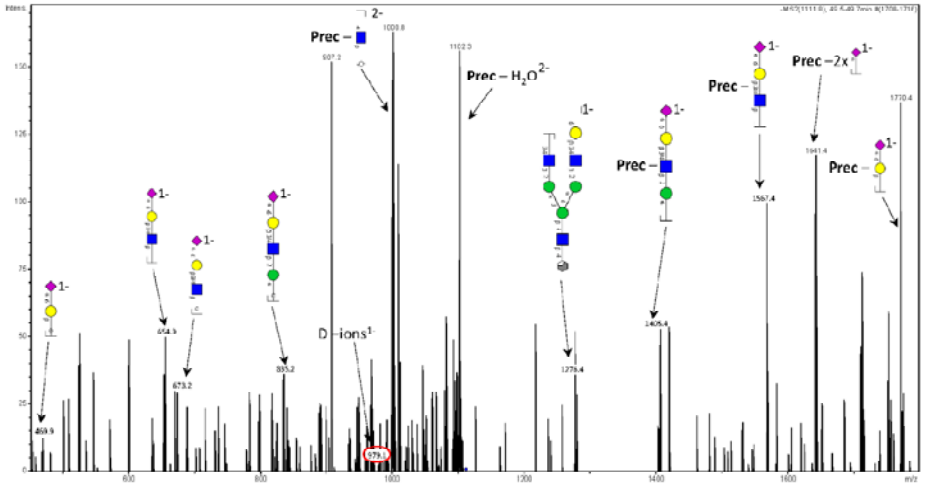
Glycan No 26b
Type = Complex
Precursor = *m/z* 1075.8^{2-}
[M-H]$^{1-}$ = 2152.6 Da
LC retention time = 59.5 min





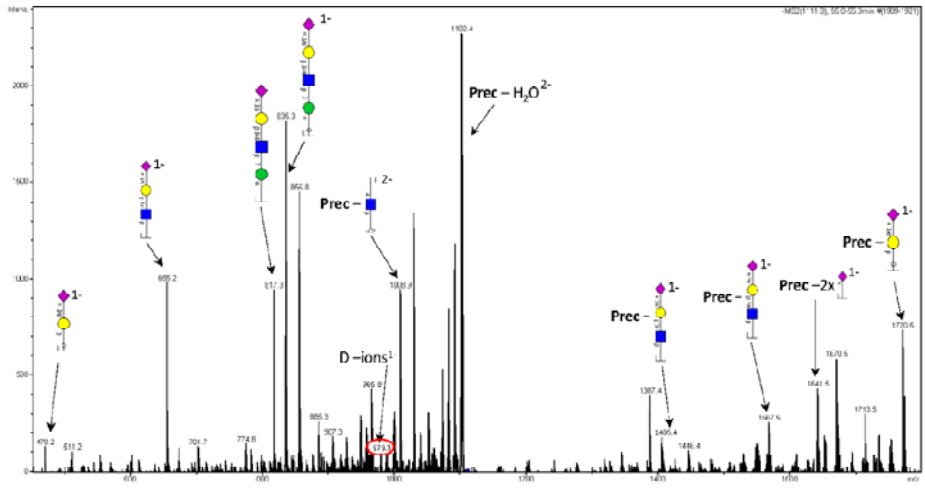
No match to MS2 spectrum in UnicarbKB

Glycan No 27a
Type = Complex
Precursor = *m/z* 1111.4^{2-}
[M-H]$^{1-}$ = 2223.8 Da
LC retention time = 49.6 min



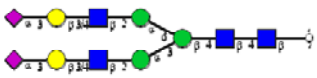
Positive match to MS2 spectrum in UnicarbKB

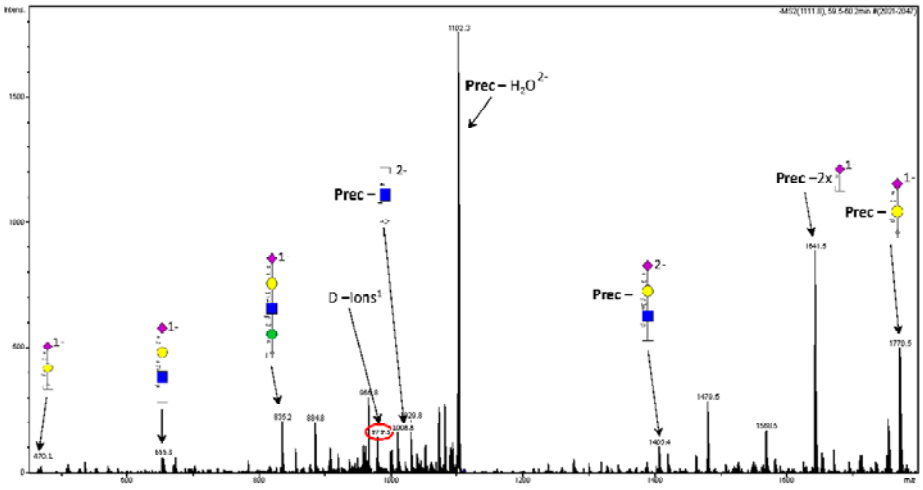
Glycan No 27b
Type = Complex
Precursor = *m/z* 1111.4^{2-}
[M-H]$^{1-}$ = 2223.8 Da
LC retention time = 55.0 min



Positive match to MS2 spectrum in UnicarbKB

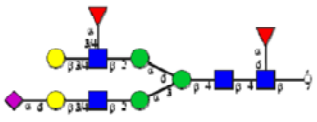
Glycan No 27c
Type = Complex
Precursor = *m/z* 1111.4^{2-}
[M-H]$^{1-}$ = 2223.8 Da
LC retention time = 59.6 min

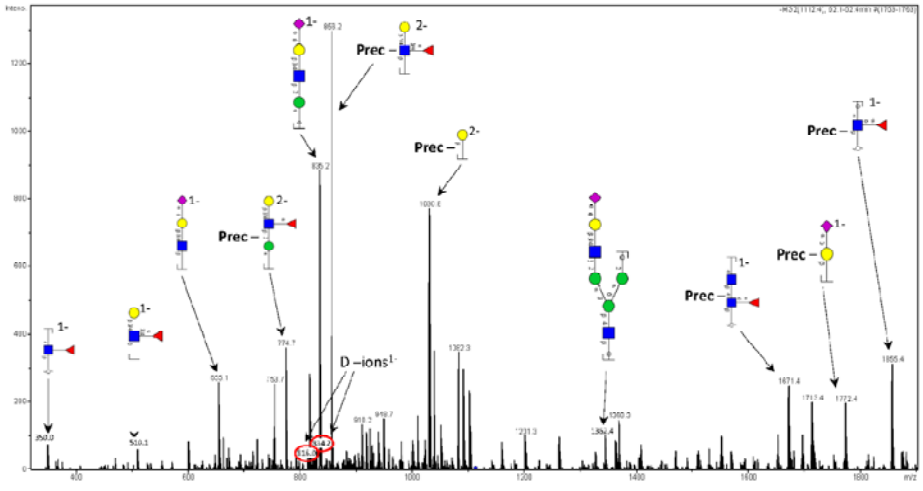




Positive match to MS2 spectrum in UnicarbKB

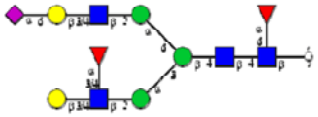
Glycan No 28a
Type = Complex
Precursor = *m/z* 1111.8^{2-}
[M-H]$^{1-}$ = 2224.6 Da
LC retention time = 52.2 min

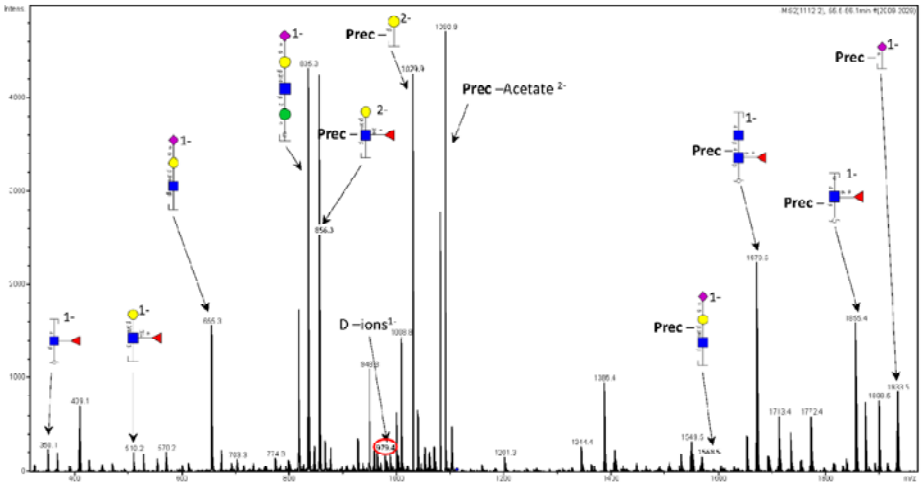




Positive match to MS2 spectrum in UnicarbKB

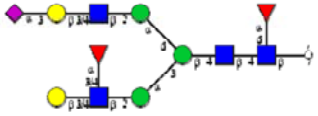
Glycan No 28b
Type = Complex
Precursor = *m/z* 1111.8^{2-}
[M-H]$^{1-}$ = 2224.6 Da
LC retention time = 55.7 min

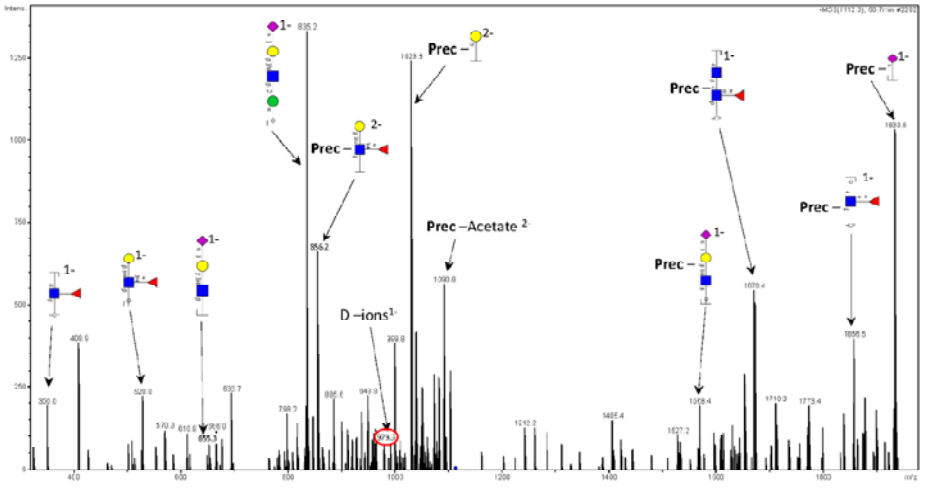




Positive match to MS2 spectrum in UnicarbKB

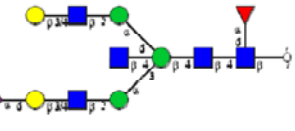
Glycan No 28c
Type = Complex
Precursor = *m/z* 1111.8^{2-}
[M-H]$^{1-}$ = 2224.6 Da
LC retention time = 60.7 min

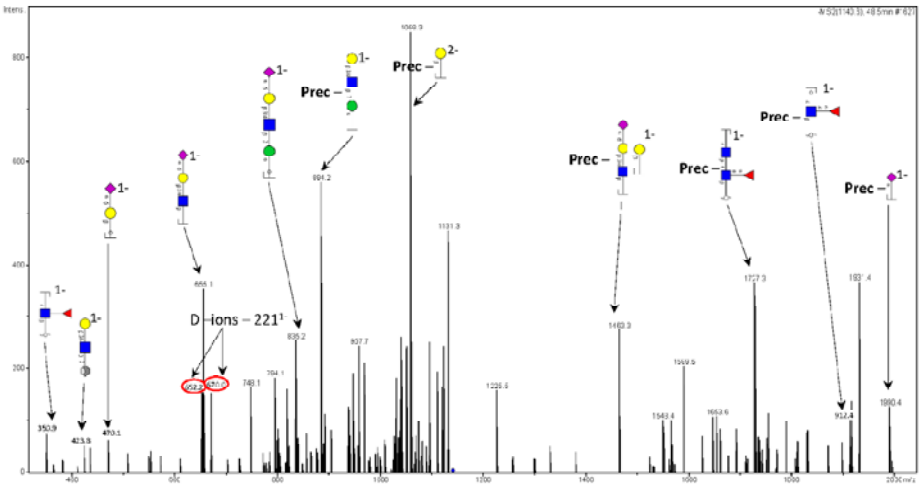




Positive match to MS2 spectrum in UnicarbKB

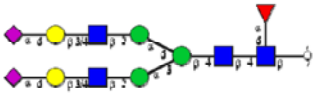
Glycan No 29
Type = Complex
Precursor = *m/z* 1140.4^{2-}
[M-H]$^{1-}$ = 2281.6 Da
LC retention time = 48.5 min

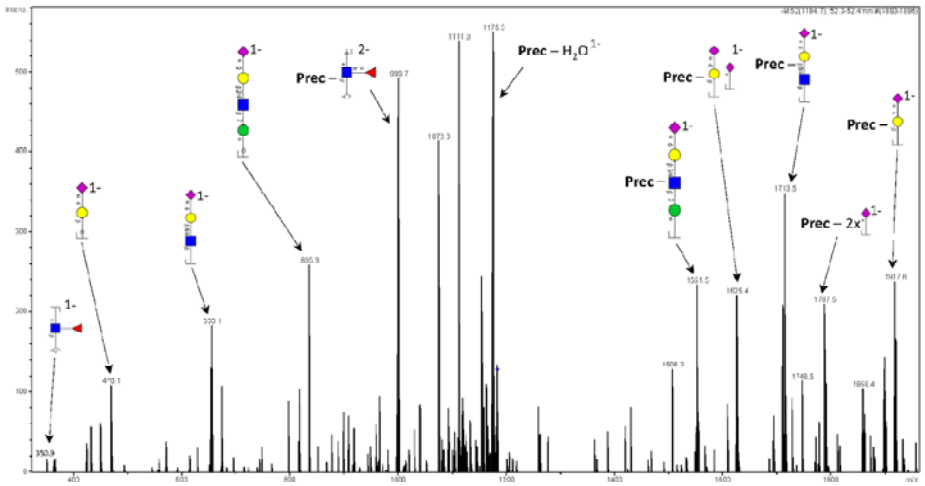




Positive match to MS2 spectrum in UnicarbKB

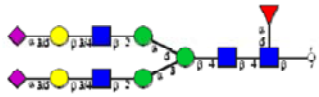
Glycan No 30a
Type = Complex
Precursor = *m/z* 1184.4^{2-}
[M-H]$^{1-}$ = 2368.8 Da
LC retention time = 52.4 min

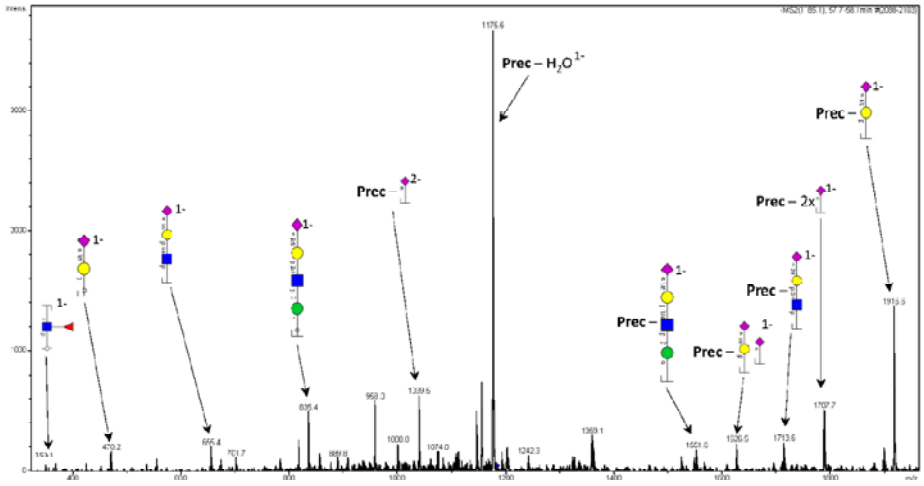




Positive match to MS2 spectrum in UnicarbKB

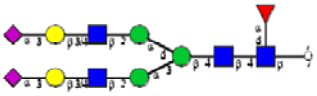
Glycan No 30b
Type = Complex
Precursor = *m/z* 1184.4^{2-}
[M-H]$^{1-}$ = 2368.8 Da
LC retention time = 57.8 min

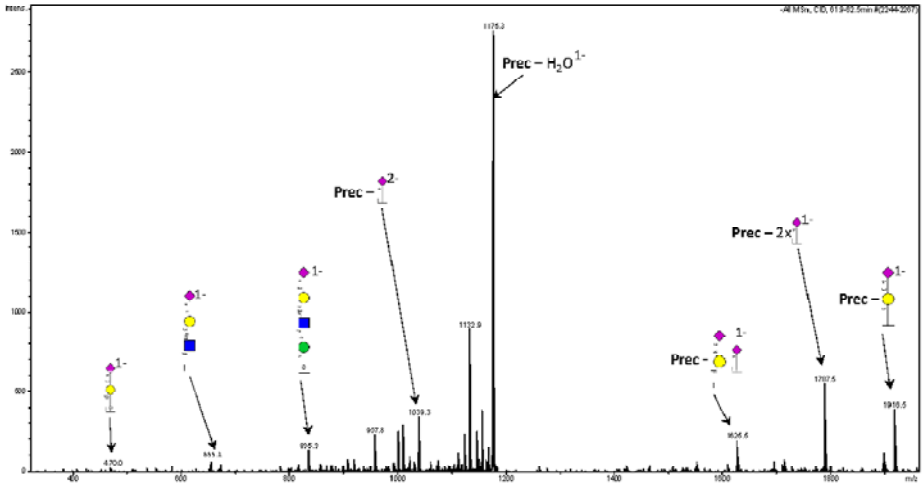




Positive match to MS2 spectrum in UnicarbKB

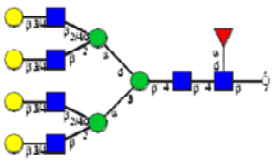
Glycan No 30c
Type = Complex
Precursor = *m/z* 1184.4^{2-}
[M-H]$^{1-}$ = 2368.8 Da
LC retention time = 62.1 min

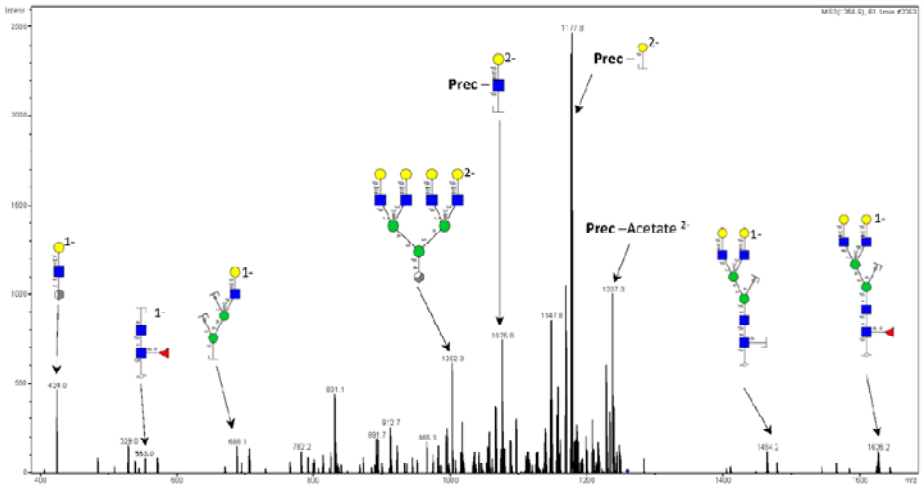




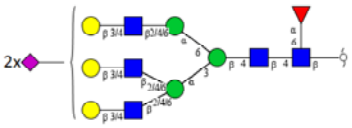
Positive match to MS2 spectrum in UnicarbKB

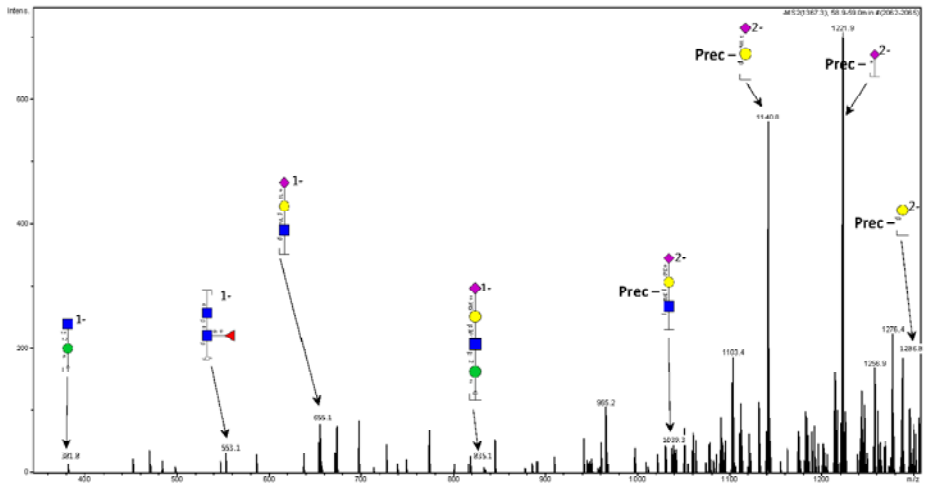
Glycan No 31
Type = Complex
Precursor = *m/z* 1258.5^{2-}
[M-H]$^{1-}$ = 2517 Da
LC retention time = 61.1 min





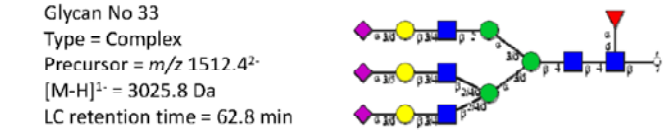
Glycan No 32
Type = Complex
Precursor = *m/z* 1366.8^{2-}
[M-H]$^{1-}$ = 2734.6Da
LC retention time = 58.8 min

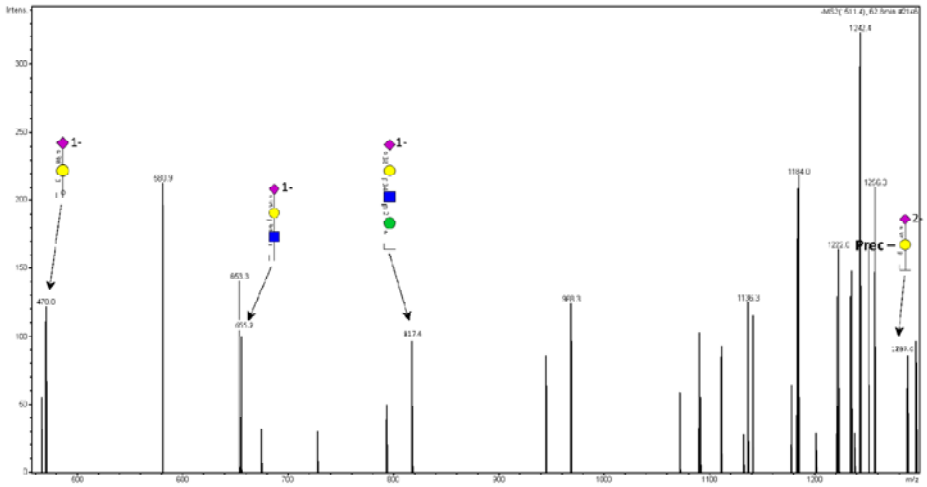




No match to MS2 spectrum in UnicarbKB

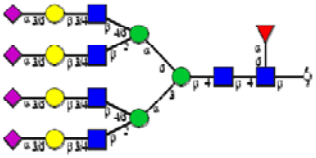
Glycan No 33
Type = Complex
Precursor = *m/z* 1512.4^{2-}
$[M-H]^{1-}$ = 3025.8 Da
LC retention time = 62.8 min

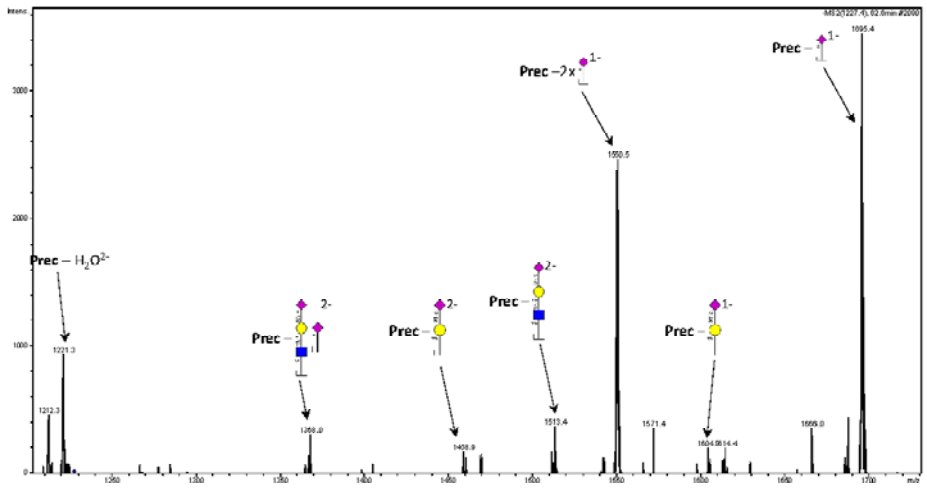




No match to MS2 spectrum in UnicarbKB

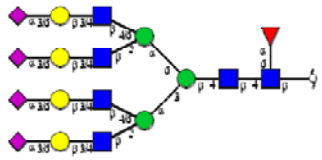
Glycan No 34a
Type = Complex
Precursor = *m/z* 1226.8^{3}
$[M-H]^{1-}$ = 3682.3 Da
LC retention time = 62.5 min

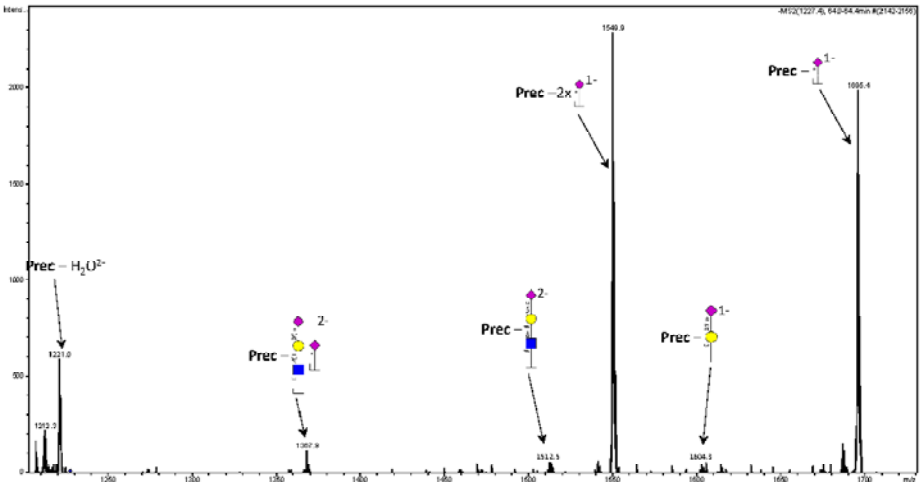




No match to MS2 spectrum in UnicarbKB

Glycan No 34b
Type = Complex
Precursor = *m/z* 1226.8[3]
$[M-H]^{1-}$ = 3682.3 Da
LC retention time = 64.4 min





No match to MS2 spectrum in UnicarbKB

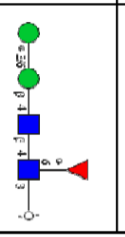
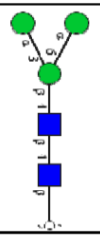
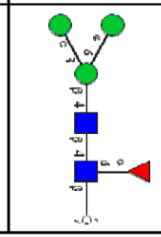
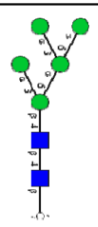
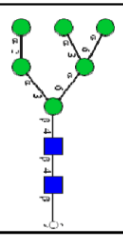
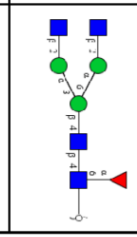
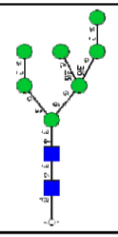
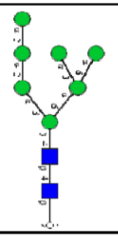
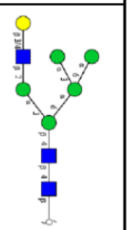
Appendix 4a - Released *N* -glycans identified from membrane proteins of six breast epithelial cell lines by PGC-LC-MS/MS (CID)

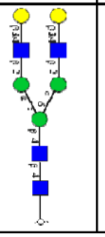
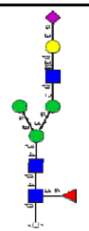
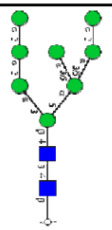
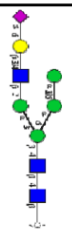
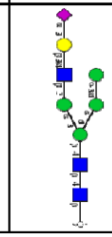
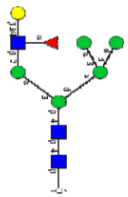
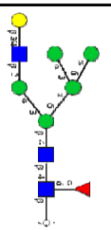
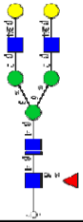
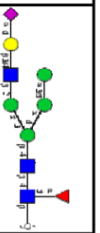
										RT=Retention time												
Glycan #	Theo m/z	charge	Hex	HexNAc	dHex	NeuAc	Isoform	MS/MS	Theo mass (M)	RT_Hmec	Obs. Mass (M)_HMEC	RT_MCF7	Obs. Mass (M)_MCF7	RT_SKBR3	Obs. Mass (M)_SKBR3	RT_MDA157	Obs. Mass (M)_MDA1	RT_MDA231	Obs. Mass (M)_MDA2	RT_HS578T	Obs. Mass (M)_HS578T	Delta mass HMEC
1	895.4	1	2	2	1		a	Y	896.3	50.8	896.407	50.4	896.407	51.0	896.407	51.0	896.407	51.0	896.407	51.8	896.407	0.1
2	911.3	1	3	2			a	Y	912.3					50.0	912.307	50.0	912.307	50.0	912.307			
3	1057.5	1	3	2	1		a	Y	1058.4	55.7	1058.507	55.4	1058.507	56.0	1058.507	55.0	1058.507	56.0	1058.507	56.7	1058.507	0.1
4	1235.5	1	5	2			a	Y	1236.4	51.9	1236.507	51.4	1236.507	52.0	1236.507	52.0	1236.507	52.0	1236.507	52.7	1236.507	0.1
5	698.3	2	6	2			a	Y	1398.5	46.7	1398.615	46.4	1398.615	47.0	1398.615	47.0	1398.615	47.0	1398.615	48.0	1398.615	0.1
6	731.2	2	3	4	1		a	Y	1464.5			52.0	1464.415	52.0	1464.415	52.0	1464.415	52.0	1464.415			
7a	779.3	2	7	2			a	Y	1560.5	45.7	1560.615	45.5	1560.615	46.0	1560.615	46.0	1560.615	45.8		47.0		0.1
7b	779.3	2	7	2			b	Y	1560.5	46.8	1560.615	46.5	1560.615	47.0	1560.615	47.0	1560.615	47.0	1560.615	48.4	1560.615	0.1
8	783.3	2	4	3		1	a	Y	1568.5					47.0	1568.615							
9	799.8	2	6	3			a	Y	1601.6	49.2	1601.615	48.9	1601.615	50.0	1601.615	50.0	1601.615	50.0	1601.615	50.5	1601.615	0.1
10	820.3	2	5	4			a	Y	1642.6	50.5	1642.615	50.3	1642.615			51.0	1642.615	51.0	1642.615	51.7	1642.615	0.0
11a	856.3	2	4	3	1	1	a	Y	1714.6	50.7	1714.615			51.0	1714.615	51.0	1714.615					0.0
11b	856.3	2	4	3	1	1	b	Y	1714.6			56.5	1714.615	57.0	1714.615	57.0	1714.615	57.0	1714.615			
12	860.3	2	8	2			a	Y	1722.6	46.0	1722.615	46.1	1722.615	46.0	1722.615	46.0	1722.615	46.0	1722.615	47.3	1722.615	0.0
13a	864.4	2	5	3		1	a	Y	1730.6	45.7	1730.815	45.7	1730.815	46.0	1730.815	46.0	1730.815	47.0	1730.815			0.2
13b	864.4	2	5	3		1	b	Y	1730.6	52.4	1730.815	52.8	1730.815	53.0	1730.815	53.0	1730.815	52.0	1730.815			0.2
14a	872.9	2	6	3	1		a	Y	1747.6			47.6	1747.815	48.0	1747.815							
14b	872.9	2	6	3	1		b	Y	1747.6			53.4	1747.815			54.0	1747.815					
15	893.3	2	5	4	1		a	Y	1788.6	55.0	1788.615	54.8	1788.615			55.0	1788.615	56.0	1788.615	56.1	1788.615	0.0
16a	937.3	2	5	3	1	1	a	Y	1876.7	49.5	1876.615			50.0	1876.615	50.0	1876.615					0.0
16b	937.3	2	5	3	1	1	b	Y	1876.7			56.5	1876.615	57.0	1876.615	56.0	1876.615	56.0	1876.615			
17	941.4	2	9	2			a	Y	1884.6	46.4	1884.815	46.5	1884.815	47.0	1884.815	47.0	1884.815	47.0	1884.815	47.9	1884.815	0.2
18a	945.4	2	6	3		1	a	Y	1892.7	48.2	1892.815			49.0	1892.815	49.0	1892.815	49.0	1892.815			0.2
18b	945.4	2	6	3		1	b	Y	1892.7	54.6	1892.815	54.1	1892.815	55.0	1892.815	55.0	1892.815	55.0	1892.815			0.2
19a	965.9	2	5	4		1	a	Y	1933.7	49.6	1933.815			50.0	1933.815	50.0	1933.815	50.0	1933.815	51.3	1933.815	0.1
19b	965.9	2	5	4		1	b	Y	1933.7	55.9	1933.815					56.0	1933.815	56.0	1933.815	57.1	1933.815	0.1
20a	966.4	2	5	4	2		a	Y	1934.7			52.7	1934.815									
20b	966.4	2	5	4	2		b	Y	1934.7	57.6	1934.815	57.5	1934.815									0.1
21	994.9	2	5	5	1		a	Y	1991.7							49.0	1991.815					
22	1018.4	2	6	3	1	1	a	Y	2038.7							58.0	2038.815					
23	1022.4	2	10	2			a	Y	2046.7	48.1	2046.815	48.4	2046.815	49.0	2046.815	49.0	2046.815	48.0	2046.815	49.5	2046.815	0.1
24a	1038.9	2	5	4	1	1	a	Y	2079.7					49.0	2079.815							
24b	1038.9	2	5	4	1	1	b	Y	2079.7	53.4	2079.815	53.8	2079.815	54.0	2079.815	54.0	2079.815	54.0	2079.815	54.8	2079.815	0.1
24c	1038.9	2	5	4	1	1	c	Y	2079.7	59.4	2079.815	59.7	2079.815			59.0	2079.815	59.0	2079.815	60.0	2079.815	0.1
25a	1039.3	2	5	4	3		a	Y	2080.8			55.0	2080.615									
25b	1039.3	2	5	4	3		b	Y	2080.8			60.5	2080.615									
26a	1075.8	2	6	5	1		a	Y	2153.8							57.0	2153.615					
26b	1075.8	2	6	5	1		b	Y	2153.8			60.4	2153.615									
27a	1111.4	2	5	4		2	a	Y	2224.8	49.2	2224.815			49.0	2224.815	49.0	2224.815			50.3	2224.815	0.0
27b	1111.4	2	5	4		2	b	Y	2224.8	55.2	2224.815			55.0	2224.815	55.0	2224.815	54.0	2224.815	56.4	2224.815	0.0
27c	1111.4	2	5	4		2	c	Y	2224.8	59.8	2224.815	59.2	2224.815	60.0	2224.815	60.0	2224.815	59.0	2224.815	60.3	2224.815	0.0
28a	1111.8	2	5	4	2	1	a	Y	2225.8			51.8	2225.615	53.0	2225.615	52.0	2225.615	52.0	2225.615			
28b	1111.8	2	5	4	2	1	b	Y	2225.8	55.9	2225.615	57.3	2225.615			56.0	2225.615	56.0	2225.615			-0.2
28c	1111.8	2	5	4	2	1	c	Y	2225.8	61.0	2225.615	61.0	2225.615			61.0	2225.615					-0.2
29	1140.4	2	5	5	1	1	a	Y	2282.8							48.0	2282.815					
30a	1184.4	2	5	4	1	2	a	Y	2370.8	52.5	2370.815			53.0	2370.815	53.0	2370.815	53.0	2370.815	53.8	2370.815	0.0
30b	1184.4	2	5	4	1	2	b	Y	2370.8	57.9	2370.815	58.1	2370.815	58.0	2370.815	58.0	2370.815	58.0	2370.815	58.6	2370.815	0.0
30c	1184.4	2	5	4	1	2	c	Y	2370.8	62.5	2370.815	62.4	2370.815	62.0	2370.815	62.0	2370.815	62.0	2370.815	62.6	2370.815	0.0
31	1258.5	2	7	6	1		a	Y	2518.9			60.5	2519.015			60.0	2519.015					
32	1366.8	2	6	5	1	2	a	Y	2736.0	59.0	2735.615	59.4	2735.615	59.0	2735.615	60.0	2735.615	60.0	2735.615			-0.3
33	1512.4	2	6	5	1	3	a	Y	3027.1	62.7	3026.815	62.9	3026.815	63.0	3026.815	63.0	3026.815	62.0	3026.815			-0.2
34a	1226.8	3	7	6	1	4	a	Y	3683.3					62.5								
34b	1226.8	3	7	6	1	4	b	Y	3683.3			63.4	3683.422	64.0	3683.422	64.0	3683.422	64.0	3683.422			

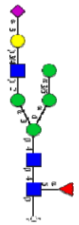
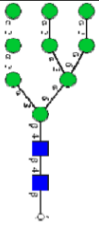
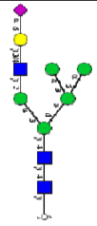
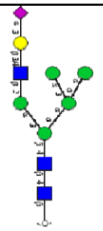
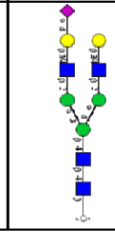
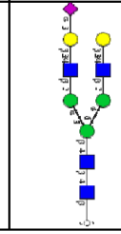
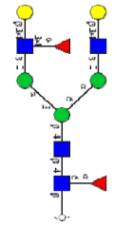
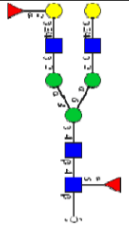
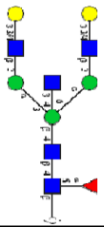
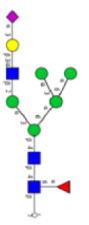
(COND'T) Appendix 4a - Released *N*-glycans identified from membrane proteins of six breast epithelial cell lines by PGC-LC-MS/MS (CID)

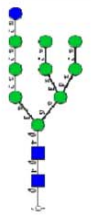
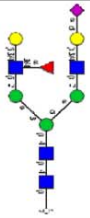
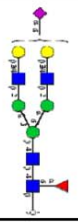
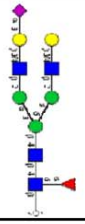
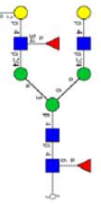
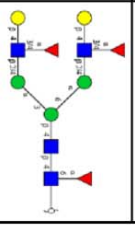
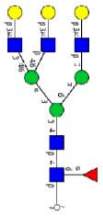
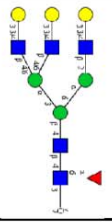
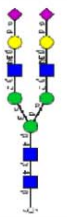
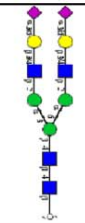
					avg AUC= averga e area under curve of three techical replicates						RA=Relative abundance					
Delta mass MCF7	Delta mass SKBR3	Delta mass MDA157	Delta mass MDA231	Delta mass HS578T	avgAUC_HMEC	avgAUC_MCF7	avgAUC_SKBR3	avgAUC_MDA157	avgAUC_MDA231	avgAUC_HS578T	RA_HMEC	RA_MCF7	RA_SKBR3	RA_MDA157	RA_MDA231	RA_HS578T
0.1	0.1	0.1	0.1	0.1	641196.67	799457.33	2453145.67	516991.67	367020.00	132308.00	0.4108	0.2546	3.3100	4.3577	0.8266	0.6160
	0.0	0.0	0.0				581334.67	1055199.00	240014.00				0.7298	0.9712	0.4749	
0.1	0.1	0.1	0.1	0.1	888054.67	2370328.67	1991144.67	5180007.00	739734.33	268305.80	0.5674	1.7816	2.6487	2.4005	1.6340	1.0779
0.1	0.1	0.1	0.1	0.1	583768.00	2732235.00	2072959.00	3067731.00	1178073.33	160561.00	0.4465	0.9506	2.8952	3.6565	2.6234	0.8112
0.1	0.1	0.1	0.1	0.1	14259788.33	11271466.67	7255415.00	6122362.00	2241549.33	1228931.00	9.8288	11.9382	9.1127	7.0507	4.8796	5.4534
-0.1	-0.1	-0.1	-0.1			1610157.33	444334.67	529635.00	1218222.33			0.7733	0.6194	0.6206	2.7685	
0.1	0.1				13600338.33	8298971.67	5004408.33	6807650.67	2762904.33	802938.30	9.5168	7.9051	6.3277	9.1156	6.2566	3.8106
0.1	0.1	0.1	0.1	0.1	2624079.33	3004596.67	1913136.00	1736352.67	1209309.00	551605.87	1.6788	3.0932	2.4051	1.9150	2.6030	2.7642
	0.1						1256038.33						1.5866			
0.1	0.1	0.1	0.1	0.1	844416.67	1052827.00	376201.00	1664158.67	547806.67	143944.67	0.5490	0.5038	0.4670	1.8297	1.2444	0.7441
0.0		0.0	0.0	0.0	1073760.33	180252.33		366154.00	524816.33	432786.00	0.6860	0.0730		0.3200	1.1595	2.0107
	0.0				534256.00		3172747.00	953478.67			0.3372		4.3528	1.2539		
0.0	0.0	0.0	0.0			520039.00	544706.67	1367142.67	198277.67			0.4528	0.8155	1.0426	0.4829	
0.0	0.0	0.0	0.0	0.0	17209323.33	30584558.00	12561848.33	12110075.00	6797472.33	3446591.00	11.5664	30.8617	15.6035	14.6431	15.1674	15.6143
0.2	0.2	0.2	0.2		1757737.67	162842.33	1192311.33	399212.67	87569.33		1.1735	0.1309	1.5030	0.4919	0.2096	
0.2	0.2	0.2	0.2		721405.33	187385.33	327593.00	1351050.67	521800.33		0.4780	0.0418	0.4669	1.7373	1.1784	
0.2	0.2					872493.67	1050504.67					0.9689	1.3568			
0.2						710716.00		1583291.00				0.4593		1.5697		
0.0		0.0	0.0	0.0	3249158.00	2522355.00		2156480.00	2057917.00	2134134.03	2.1901	1.9870		1.3361	4.5921	11.5962
	0.0				858273.67		1314722.00	1072579.00			0.5908		1.6970	1.4092		
0.0	0.0	0.0	0.0			416972.67	177082.67	3226760.67	288656.00			0.3600	0.2623	2.2734	0.6307	
0.2	0.2	0.2	0.2	0.2	14279500.00	26122909.00	9618080.33	12597121.97	6305151.91	3855330.53	9.7483	25.6105	11.7647	15.6362	13.8020	17.8352
	0.2	0.2	0.2		2056740.67		4254315.67	1054910.67	262122.33		1.4698		5.3229	1.0079	0.6055	
0.2	0.2	0.2	0.2		969251.33	444727.00	1161219.00	2864460.33	701661.67		0.5949	0.3931	1.5465	2.3840	1.6101	
	0.1	0.1	0.1	0.1	8300686.00		427173.33	1050657.33	440792.00	387712.33	5.8356		0.5660	1.4283	0.9959	2.7543
		0.1	0.1	0.1	2399725.33			1466339.67	925387.00	427631.33	1.5185			0.9449	2.0649	1.9359
0.1						555836.33						0.3154				
0.1					1721464.67	1319907.33					1.0860	0.9465				
								1470333.33						1.2041		
								1767047.33						1.7838		
0.1	0.1	0.1	0.1	0.1	2383412.67	4598088.00	2148920.47	2338855.00	1836612.33	413534.33	1.6461	5.2442	2.7243	1.9622	4.1032	1.9545
	0.1						791641.33						0.9945			
0.1	0.1	0.1	0.1	0.1	2734979.67	614896.67	1110715.00	3633465.67	2399046.33	429999.33	18.8173	0.4322	1.4740	3.4336	5.1271	1.9568
0.1		0.1	0.1	0.1	2316668.33	1260685.33		2980925.33	2846119.67	2265250.67	1.5800	0.9734		3.0828	6.0082	10.2678
-0.1						618280.67						0.4661				
-0.1						577996.33						0.2329				
								254417.67						0.1278		
-0.2						804339.33						0.4449				
	0.0			0.0	715629.33		569508.33	128403.77		1273391.33	0.4835		0.7019	0.1068		8.8332
	0.0	0.0	0.0	0.0	3436449.67		939168.33	976757.33	323251.67	151037.67	2.3356		1.2198	0.7306	0.7323	0.6913
0.0	0.0	0.0	0.0	0.0	856174.33	113544.00	249038.00	732090.00	904699.00	294593.67	0.5823	0.0843	0.3308	0.5656	1.8819	1.2416
-0.2	-0.2	-0.2	-0.2			296814.00	2356740.67	221442.33	437102.33			0.0784	3.1733	0.2739	0.9732	
-0.2		-0.2	-0.2		7594152.00	241921.33		541480.33	242356.00		5.1586	0.1826		0.4251	0.5907	
-0.2					665439.00	168108.00		280001.67			0.4710	0.0629		0.2965		
								296457.67						0.2564		
	0.0	0.0	0.0	0.0	2981142.33		1888105.33	178934.33	82203.00	60699.33	1.9615		2.5889	0.0403	0.1796	0.3568
0.0	0.0	0.0	0.0	0.0	6945517.67	297767.10	3288025.00	2474626.33	1162157.63	403978.00	4.6722	0.2742	4.6915	1.7125	2.6209	1.8691
0.0	0.0	0.0	0.0	0.0	2063720.33	1081422.00	774425.70	2222634.00	3742638.67	1323357.67	1.5088	0.7174	1.2097	3.0279	9.2134	5.8048
0.1						451668.67		140773.67				0.1820		0.1021		
-0.3	-0.3	-0.3	-0.3		425347.00	300977.33	191553.00	566253.00	226277.33		0.2802	0.1912	0.2695	0.4720	0.4978	
-0.2	-0.2	-0.2	-0.2		321280.00	401630.00	890675.00	507689.67	665680.00		0.2297	0.2742	1.3786	0.6192	1.6389	
							1298347.67						2.0495			
0.1	0.1	0.1	0.1			655457.33	1145727.00	203592.00	253674.67			0.3579	1.8335	0.3805	0.6229	

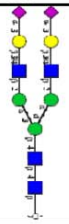
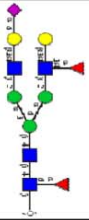
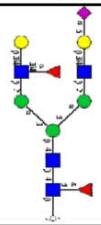
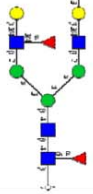
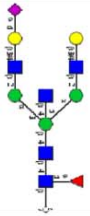
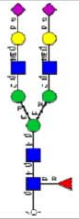
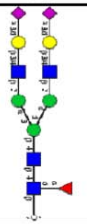
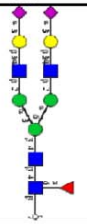
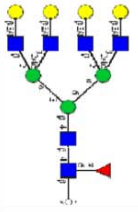
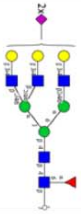
Appendix 4b - Released *N*-glycans identified from membrane proteins of six breast epithelial cell lines by PGC-LC-MS/MS (CID)

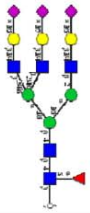
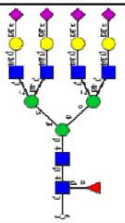
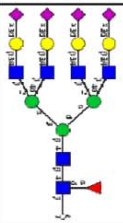
Confidence	High	High	High	High	High	High	High	High	High	High
Type	Paucimannose	Paucimannose	Paucimannose	High mannose	High mannose	Complex	High mannose	High mannose	Complex	Hybrid
Glycan no	1	2	3	4	5	6	7a	7b	8	9
										

Confidence	High	Medium	HIgh	HIgh	High	High	High	High	High	Medium/High
Type	Complex	Complex	Complex	High mannose	Hybrid	Hybrid	Hybrid	Hybrid	Complex	Hybrid
Glycan no	10	11a	11b	12	13a	13b	14a	14b	15	16a
										

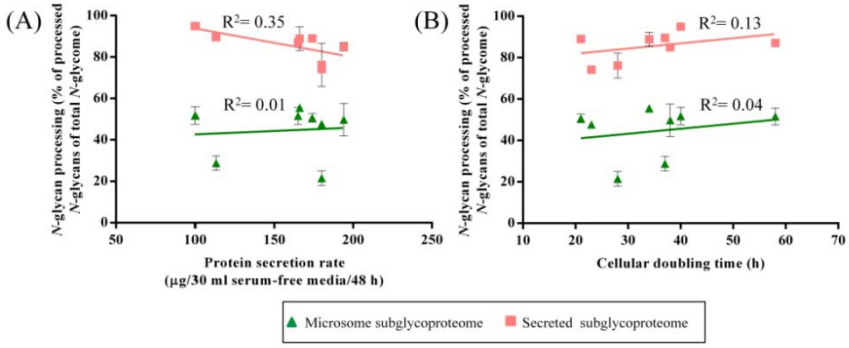
Confidence	Medium/High	High	High	High	High	High	High	High	High	High
Type	Hybrid	High mannose	Hybrid	Hybrid	Complex	Complex	Complex	Complex	Complex	Hybrid
Glycan no	16b	17	18a	18b	19a	19b	20a	20b	21	22
										

Confidence	High	High	High	High	High	High	High	High	High	High
Type	High mannose	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	23	24a	24b	24c	25a	25b	26a	26b	27a	27b
										

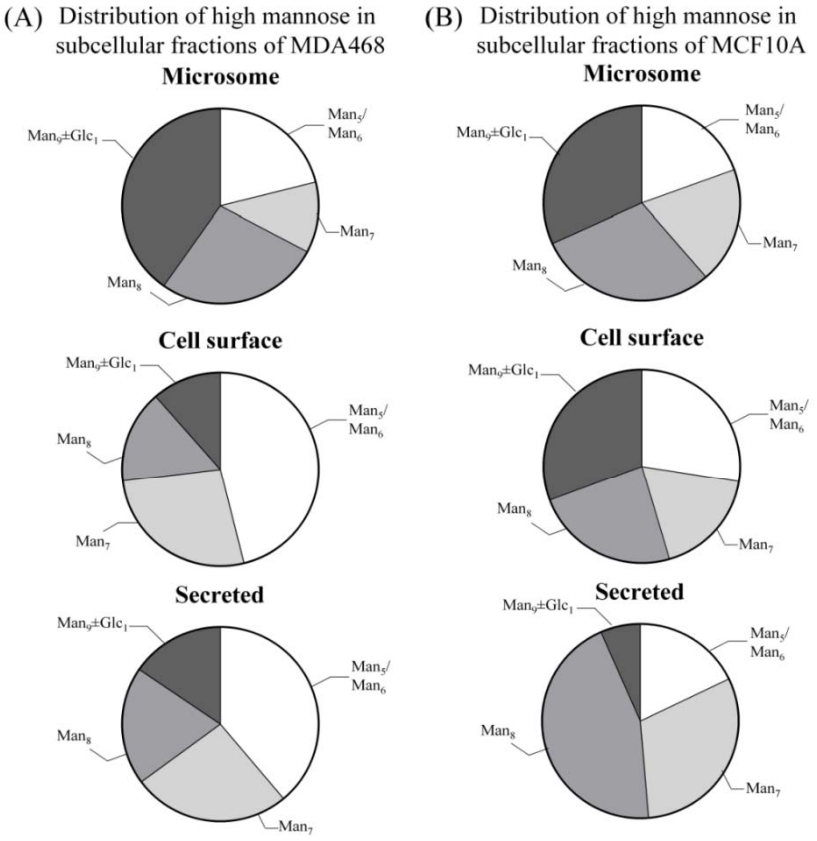
Confidence	High	High	High	High	High	High	High	High	Medium	Medium/High
Type	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex	Complex
Glycan no	27c	28a	28b	28c	29	30a	30b	30c	31	32
										

Confidence	High	Medium	Medium
Type	Complex	Complex	Complex
Glycan no	33	34a	34b
			

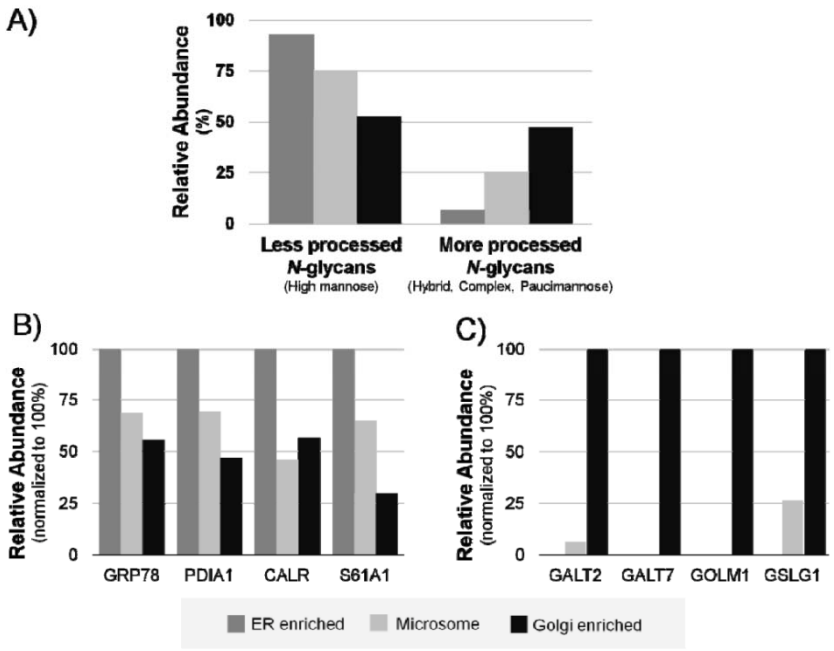
Supplementary Figures



Supplementary Figure 1. **(A)** Secretion rate as measured by the protein concentration (µg / 30 mL serum-free media / 48 h) in the culture media shows weak correlation with the *N*-glycan processing as measured by the molar proportion of more processed *N*-glycan types (i.e. hybrid, complex and paucimannose) of the total *N*-glycome for the secretome (R^2 = 0.35) (light red) but not the microsome (R^2 = 0.01) (green) derived from the investigated panel of human breast cell lines. **(B)** No/negligible correlations were observed between the cellular doubling time (hours) and *N*-glycan processing of either the secretome (R^2 = 0.13) or the microsome (R^2 = 0.04).

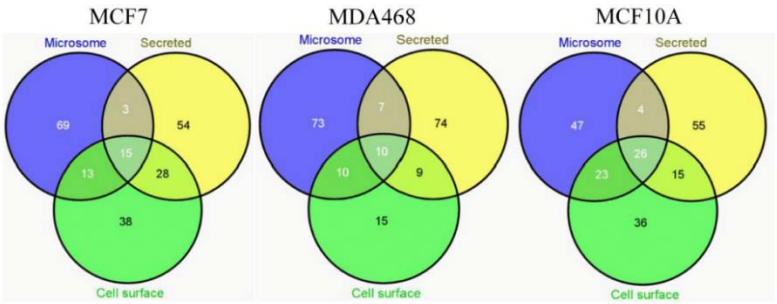


Supplementary Figure 2. The subcellular-specific distribution of the high mannose series of **(A)** MDA468 and **(B)** MCF10A into Man_5, Man_6, Man_7, Man_8, Man_9 ± Glc_1. The latter represents immature *N*-glycans.

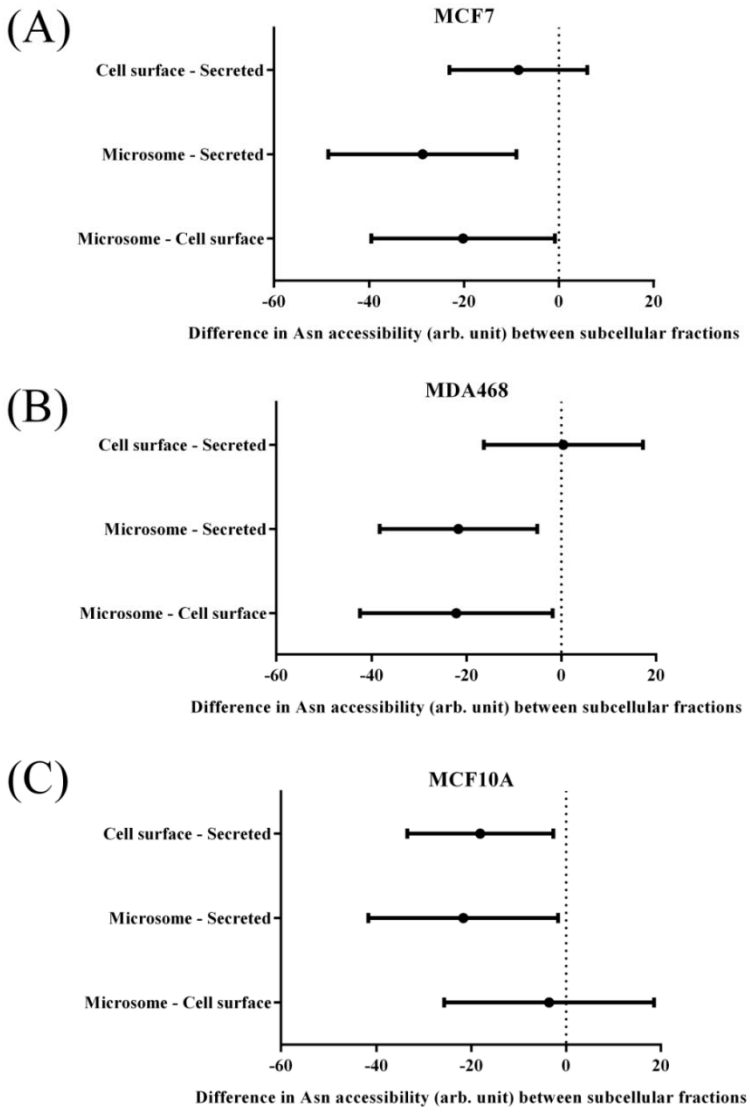


Supplementary Figure 3. **(A)** *N*-glycome mapping of the three subcellular fractions of SW480 i.e. ER-enriched (dark grey bars), microsome (light grey bars) and Golgi-enriched (black bars). *N*-glycome profiles are presented as the molar distribution of the less processed (left, high mannose type *N*-glycans) and the more processed (right, complex, hybrid, paucimannose) *N*-glycans. **(B-C)** Quantitative LC-MS/MS-based proteomics of the same subcellular fractions (bottom) confirmed the desired enrichment of proteins by mapping the relative abundances of known ER- (left) and Golgi- (right) protein markers.

Distribution of top 100 most abundant glycoproteins (identified and putative) in the three subcellular proteomes of MCF7, MDA468 and MCF10A



Supplementary Figure 4. The distribution of the top-100 most abundant putative glycoproteins in the three subcellular proteomes of MCF7, MDA468 and MCF10A. The unique proteins identified in each subcellular fraction were used for determining average glycosylation site accessibilities of the subcellular glycoproteomes.



Supplementary Figure 5. The comparison of subcellular-specific glycosylation site accessibilities (unit-less, arbitrary values) in (A) MCF7, (B) MDA468 and (C) MCF10A as measured by one-way ANOVA. The graphs show the 95% confidence intervals for the differences between the means (site accessibility).

Supplementary Tables

Supplementary Table 1.
Overview of investigated panel of human breast cells. Subtype, presence (+) / absence (-) / data not available (NA) of hormone receptors (i.e. estrogen receptor (ER), progesterone receptor (PR), and ERBB2), origin as extracted from ATCC and other supporting literature is listed. Doubling time (hours) and protein secretion rate (µg / 30 mL serum-free media / 48 h) were determined for all cells and plotted against the *N*-glycan type processing of the secreted and microsome subcellular glycoproteomes (see **Supplementary Figure 1**).

Cell line	**Sub-type**	**ER / PR/ ERBB2**	**Origin**	**Doubling time*** (h)	**Protein secretion** (µg / 30 mL media / 48 h)	**Viability** (%)
MCF7	Luminal	+/+/-	Human adenocarcinoma	28	180	93
SKBR3	Luminal	-/-/+	Human adenocarcinoma	38	194	97
MDA157	Basal B	-/-/-	Human medullary carcinoma	58	165	92
MDA231	Basal B	-/-/-	Human metastatic adenocarcinoma	21	174	98
MDA468	Basal A	-/-/-	Human metastatic adenocarcinoma	37	113	91
HS578T	Basal B	-/-/-	Human carcinoma sarcoma	40	100	98
HMEC	NA	NA / NA / NA	Human mammary epithelia cells, primary tissue	34	166	95
MCF10A	Basal B	-/-/-	Human fibrocystic disease	23	180	92

Supplementary Table 2

Summary of unique glycoproteins from the three subcellular fractions i.e. secreted, cell surface and microsome subglycoproteome and the solvent accessibility of asparagine residues residing in conserved sequons (mean ± SD).

MCF7 Microsome (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Trans-membrane	Glycosylation sites (No)	Glycoprotein	Mol Weight
P28288	ATP-binding cassette sub-family D member 3	ABCD3_HUMAN	N/A				659	N	Y	1	Y	75 kDa
Q8N2K0	Monoacylglycerol lipase ABHD12	ABD12_HUMAN	N/A				398	N	Y	1	Y	45 kDa
O00767	Acyl-CoA desaturase	ACOD_HUMAN	N/A				359	N	Y	3	P	42 kDa
Q9HDC9	Adipocyte plasma membrane-associated protein	APMAP_HUMAN	N/A				416	Y	Y	2	Y	46 kDa
Q9HD20	Probable cation-transporting ATPase 13A1	AT131_HUMAN	N/A				1204	N	Y	6	Y	133 kDa
P54709	Sodium/potassium-transporting ATPase subunit beta-3	AT1B3_HUMAN	Model on 2kdpB	40	Asn124	116.34	279	Y	Y	3	Y	32 kDa
			Model on 2kdpB		Asn240	78.57						
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	AT2A2_HUMAN	Model on 3ar4	84	Asn19	78.53	1042	N	Y	7	P	115 kDa
			Model on 3ar4		Asn421	10.53						
			Model on 3ar4		Asn589	71.33						
			Model on 3ar4		Asn738	6.43						
			Model on 3ar4		Asn918	46.98						
			Model on 3ar4		Asn962	75.61						
Q93084-2	Isoform SERCA3A of Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	AT2A3_HUMAN	Model on 3ar4a	84	Asn213	48.52	1043	N	Y	3	P	109 kDa
			Model on 3ar4a		Asn421	11.44						
			Model on 3ar4a		Asn919	36.27						
Q8NHH9	Atlastin-2	ATLA2_HUMAN	Model on 4IDO	73	Asn163	116.43	583	N	Y	2	P	66 kDa
			Model on 4IDO		Asn204	1.64						
Q6DD88	Atlastin-3	ATLA3_HUMAN	Model on 3q5e	69	Asn173	2.3	541	N	Y	5	P	61 kDa
			Model on 3q5e		Asn312	11.41						
			Model on 3q5e		Asn432	17.41						
O75110	Probable phospholipid-transporting ATPase IIA	ATP9A_HUMAN	N/A				1047	N	Y	4	P	119 kDa
P27824	Calnexin	CALX_HUMAN	N/A				592	Y	Y	1	Y	68 kDa
Q96A33	Coiled-coil domain-containing protein 47	CCD47_HUMAN	N/A				483	Y	Y	1	Y	56 kDa
P08962	CD63 antigen	CD63_HUMAN	N/A				238	Y	Y	3	Y	26 kDa
Q96D05	Cleft lip and palate transmembrane protein 1	CLPT1_HUMAN	N/A				669	N	Y	6	Y	76 kDa
P52569	Low affinity cationic amino acid transporter 2	CTR2_HUMAN	N/A				658	N	Y	4	Y	72 kDa
Q53GQ0	Estradiol 17-beta-dehydrogenase 12	DHB12_HUMAN	N/A				312	N	Y	2	P	34 kDa
Q9UBM7	7-dehydrocholesterol reductase	DHCR7_HUMAN	N/A				475	N	Y	2	P	54 kDa
Q15125	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomerase	EBP_HUMAN	N/A				230	N	Y	1	P	26 kDa
Q9BW60	Elongation of very long chain fatty acids protein 1	ELOV1_HUMAN	N/A				279	N	Y	2	P	33 kDa
Q8N766	ER membrane protein complex subunit 1	EMC1_HUMAN	N/A				993	Y	Y	3	Y	112 kDa
P33947	Cluster of ER lumen protein retaining receptor 2	ERD22_HUMAN	N/A				212	N	Y	1	P	24 kDa
Q14534	Squalene monooxygenase	ERG1_HUMAN	N/A				574	N	Y	1	P	64 kDa
Q969X5	Endoplasmic reticulum-Golgi intermediate compartment protein 1	ERGI1_HUMAN	N/A				290	Y	Y	1	Y	33 kDa
Q94905	Cluster of Erlin-2	ERLN2_HUMAN	N/A				339	Y	Y	1	Y	38 kDa
Q7Z2K6	Endoplasmic reticulum metallopeptidase 1	ERMP1_HUMAN	Model on 3PB6X	40	Asn182	52.21	904	N	Y	4	Y	100 kDa
A0FGR8-6	Isoform 6 of Extended synaptotagmin-2	ESYT2_HUMAN	4NPK		Asn433	46.4	921	N	Y	7	P	105 kDa
			4NPK		Asn511	78.6						
			4NPK		Asn587	52.68						
			4NPK		Asn643	98.77						
Q10471	Polypeptide N-acetylgalactosaminyltransferase 2	GALT2_HUMAN	2FFU		Asn516	139.38	571	Y	Y	1	P	65 kDa
B7ZAQ6	Golgi pH regulator A	GPHRA_HUMAN	N/A				455	N	Y	3	P	53 kDa
Q9P035	Very-long-chain (3R)-3-hydroxyacyl-[acyl-carrier protein] dehydratase 3	HACD3_HUMAN	N/A				362	N	Y	3	P	43 kDa
Q8TCT9	Minor histocompatibility antigen H13	HM13_HUMAN	N/A				377	Y	Y	4	Y	41 kDa
A1L0T0	Acetolactate synthase-like protein	ILVBL_HUMAN	N/A				632	N	Y	1	P	68 kDa
Q14573	Inositol 1,4,5-trisphosphate receptor type 3	ITPR3_HUMAN	Model on 3UJ0A	75	Asn104	4.01	2671	N	Y	16	P	304 kDa
			Model on 3UJ0A		Asn216	111.91						
P13473	Lysosome-associated membrane glycoprotein 2	LAMP2_HUMAN	N/A				410	Y	Y	16	Y	45 kDa
Q16706	Alpha-mannosidase 2	MA2A3_HUMAN	Model on 3bwxA	41	Asn1125	105.28	1144	N	Y	5	Y	131 kDa
			Model on 3bwxA		Asn1130	27.58						
Q96N66	Lysophospholipid acyltransferase 7	MBOA7_HUMAN	N/A				472	N	Y	1	P	53 kDa
Q14728	Major facilitator superfamily domain-containing protein 10	MFS10_HUMAN	N/A				455	N	Y	1	P	48 kDa
Q6PIU2	Neutral cholesterol ester hydrolase 1	NCEH1_HUMAN	Model on 3ainA	40	Asn270	8.01	408	Y	Y	3	Y	46 kDa
			Model on 3ainA		Asn287	42.73						
Q969V3	Nicalin	NCLN_HUMAN	N/A				563	Y	Y	2	Y	63 kDa
Q15738	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL_HUMAN	N/A				373	N	Y	1	P	42 kDa
Q8NF37	Lysophosphatidylcholine acyltransferase 1	PCAT1_HUMAN	N/A				534	N	Y	1	P	59 kDa
Q9UHG3	Prenylcysteine oxidase 1	PCYOX_HUMAN	N/A				505	Y	Y	3	Y	57 kDa
Q92508	Piezo-type mechanosensitive ion channel component 1	PIEZ1_HUMAN	N/A				2521	Y	Y	8	Y	287 kDa
Q96S52	GPI transamidase component PIG-S	PIGS_HUMAN	N/A				555	Y	Y	2	Y	62 kDa
P20340	Cluster of Ras-related protein Rab-6A	RAB6A_HUMAN	1yzq		Asn126	2.09	208	N	Y	1	P	24 kDa
P04844	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2_HUMAN	N/A				631	Y	Y	3	Y	69 kDa
Q6P1M0	Long-chain fatty acid transport protein 4	S27A4_HUMAN	N/A				643	N	Y	4	P	72 kDa
Q96K37	Solute carrier family 35 member E3	S35E3_HUMAN	N/A				430	N	Y	2	P	45 kDa
Q8N357	Solute carrier family 35 member F6	S35F6_HUMAN	N/A				371	Y	Y	3	Y	40 kDa
Q92504	Zinc transporter SLC39A7	S39A7_HUMAN	N/A				469	N	Y	1	P	50 kDa
O15126	Secretory carrier-associated membrane protein 1	SCAM1_HUMAN	N/A				338	N	Y	2	P	38 kDa
Q9UGP8	Translocation protein SEC63 homolog	SEC63_HUMAN	N/A				760	N	Y	2	P	88 kDa
P61009	Signal peptidase complex subunit 3	SPCS3_HUMAN	N/A				180	N	Y	1	Y	20 kDa
Q9UNL2	Translocon-associated protein subunit gamma	SSRG_HUMAN	N/A				185	N	Y	2	P	21 kDa
P46977	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	STT3A_HUMAN	N/A				705	Y	Y	4	Y	81 kDa
Q8TCJ2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	STT3B_HUMAN	N/A				826	N	Y	6	P	94 kDa
Q03518	Antigen peptide transporter 1	TAP1_HUMAN	N/A				808	N	Y	5	P	87 kDa
Q9NZ01	Very-long-chain enoyl-CoA reductase	TECR_HUMAN	N/A				308	N	Y	3	Y	36 kDa
O15321	Transmembrane 9 superfamily member 1	TM9S1_HUMAN	N/A				606	N	Y	4	Y	69 kDa
Q9HD45	Transmembrane 9 superfamily member 3	TM9S3_HUMAN	N/A				589	Y	Y	2	Y	68 kDa
Q92544	Transmembrane 9 superfamily member 4	TM9S4_HUMAN	N/A				642	Y	Y	1	P	75 kDa
Q9Y383	Transmembrane emp24 domain-containing protein 7	TMED7_HUMAN	N/A				224	Y	Y	1	Y	25 kDa
Q9BVK6	Transmembrane emp24 domain-containing protein 9	TMED9_HUMAN	N/A				235	N	Y	1	Y	27 kDa
P49755	Transmembrane emp24 domain-containing protein 10	TMED10_HUMAN	N/A				219	Y	Y	1	Y	25 kDa
Q9Y320	Thioredoxin-related transmembrane protein 2	TMX2_HUMAN	N/A				236	Y	Y	2	P	24 kDa
P02788	Lactotransferrin	TRFL_HUMAN	1LCF		Asn156	129.24	710	Y	N	3	Y	78 kDa
			1LCF		Asn497	129.34						
			1LCF		Asn642	144.08						
Q9P0L0	Vesicle-associated membrane protein-associated protein A	VAPA_HUMAN	N/A				249	N	Y	4	P	28 kDa
O95292	Vesicle-associated membrane protein-associated protein B/C	VAPB_HUMAN	N/A				243	N	Y	1	P	27 kDa
O76024	Wolframin	WFS1_HUMAN	N/A				890	N	Y	5	Y	100 kDa
					Mean	**59.44**						
					SD	**46.58**						

N = No
Y = Yes
P = Potential
N/A = Not available

MCF7 Cell surface (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Transmembrane	ASN SITE (NO)	Glycoprotein	Mol Weight
Q5TH69	Brefeldin A-inhibited guanine nucleotide-exchange protein 3	BIG3_HUMAN	N/A				2177	N	Y	7	P	241 kDa
P13987	CD59 glycoprotein	CD59_HUMAN	2J8B		Asn43	90.78	128	Y	Y	1	Y	14 kDa
P21926	CD9 antigen	CD9_HUMAN	N/A				228	Y	Y	2	Y	25 kDa
Q9HCU4	Cadherin EGF LAG seven-pass G-type receptor 2	CELR2_HUMAN	Model on 1x2mA		Asn81	76.1	2923	Y	Y	19	Y	317 kDa
Q8WWA5	Choline transporter-like protein 2	CTL2_HUMAN	N/A				706	Y	Y	4	Y	80 kDa
P08174	Complement decay-accelerating factor	DAF_HUMAN	1OK3		Asn95	48.58	381	Y	Y	1	Y	41 kDa
P22413	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1	ENPP1_HUMAN	Model on 4b56	80	Asn179	24.32	925	Y	Y	11	Y	105 kDa
			Model on 4b56		Asn285	84.77						
			Model on 4b56		Asn341	71.32						
			Model on 4b56		Asn477	70.12						
			Model on 4b56		Asn578	132.45						
			Model on 4b56		Asn585	37.51						
			Model on 4b56		Asn643	113.01						
			Model on 4b56		Asn700	108.15						
			Model on 4b56		Asn731	129.34						
			Model on 4b56		Asn748	149.21						
P29317	Ephrin type-A receptor 2	EPHA2_HUMAN	3CBX		Asn407	108.79	976	Y	Y	2	Y	108 kDa
			3CBX		Asn435	86.11						
P29323	Ephrin type-B receptor 2	EPHB2_HUMAN	3ZFM		Asn704	53.81	1055	Y	Y	5	Y	117 kDa
P54753	Ephrin type-B receptor 3	EPHB3_HUMAN	N/A				998	Y	Y	3	Y	110 kDa
P54760	Ephrin type-B receptor 4	EPHB4_HUMAN	2VWX		Asn768	87.93	987	Y	Y	4	Y	108 kDa
P37268	Squalene synthase	FDFT_HUMAN	3VJ8		Asn48	56.35	417	N	Y	2	P	48 kDa
			3VJ8		Asn282	22.26						
Q7L5L3	Glycerophosphodiester phosphodiesterase domain-containing protein 3	GDPD3_HUMAN	N/A				318	N	Y	1	P	37 kDa
Q08379	Golgin subfamily A member 2	GOGA2_HUMAN	N/A				1002	N	Y	2	P	113 kDa

P05362	Intercellular adhesion molecule 1	ICAM1_HUMAN	1IAM		Asn130	108.22	532	Y	Y	8	Y	58 kDa
			1IAM		Asn145	58.32						
			1IAM		Asn183	86.77						
			1IAM		Asn202	53.58						
			2D24		Asn287	145.78						
			2D24		Asn298	163.34						
			2D24		Asn335	102.35						
			2D24		Asn406	68.34						
P08069	Insulin-like growth factor 1 receptor	IGF1R_HUMAN	2GGR		Asn51	76.06	1367	Y	Y	17	Y	155 kDa
			2GGR		Asn102	44.86						
			2GGR		Asn135	35.48						
			2GGR		Asn244	60.52						
			2GGR		Asn314	92.37						
			2GGR		Asn417	99.23						
			2GGR		Asn438	99.42						
O75054	Immunoglobulin superfamily member 3	IGSF3_HUMAN	N/A				1394	Y	Y	12	Y	155 kDa
P17301	Integrin alpha-2	ITA2_HUMAN	1DZI		Asn343	52.82	1103	Y	Y	10	Y	129 kDa
P26006	Integrin alpha-3	ITA3_HUMAN	N/A				1051	Y	Y	14	Y	117 kDa
P05556	Integrin beta-1	ITB1_HUMAN	3VI3		Asn84	103.93	798	Y	Y	12	Y	88 kDa
			3VI3		Asn97	67.29						
			3VI3		Asn212	114.74						
			3VI3		Asn269	114.84						
			3VI3		Asn363	73.91						
			3VI3		Asn406	86.36						
			3VI3		Asn417	109.5						
P13612	Integrin beta-4	ITB4_HUMAN	N/A				1822	Y	Y	8	Y	202 kDa
P18084	Integrin beta-5	ITB5_HUMAN	Model on 4g1e	56	Asn347	95.88	799	Y	Y	10	Y	88 kDa
			Model on 4g1e		Asn460	54.88						
			Model on 4g1e		Asn477	89.39						
			Model on 4g1e		Asn552	59.16						
			Model on 4g1e		Asn586	84.71						
			Model on 4g1e		Asn654	100.92						
			Model on 4g1e		Asn705	25.07						
P18564	Integrin beta-6	ITB6_HUMAN	Model on 3FCS	53	Asn18	95.69	788	Y	Y	10	Y	86 kDa
			Model on 3FCS		Asn97	2.71						
			Model on 3FCS		Asn260	135.49						
			Model on 3FCS		Asn337	90.46						
			Model on 3FCS		Asn396	142.45						
			Model on 3FCS		Asn463	99.28						
			Model on 3FCS		Asn471	81.78						
			Model on 3FCS		Asn541	61.38						
			Model on 3FCS		Asn575	59.2						
Q9H0X4	Protein ITFG3	ITFG3_HUMAN	N/A				552	N	Y	8	Y	60 kDa
P15529-2	Isoform B of Membrane cofactor protein	MCP_HUMAN	3OBE		Asn83	130.71	390	Y	Y	8	Y	44 kDa
			3OBE		Asn114	53.41						
			3OBE		Asn273	130.84						
P11717	Cation-independent mannose-6-phosphate receptor	MPRI_HUMAN	2CBZ		Asn648	89.3	2491	Y	Y	21	Y	274 kDa
			2CBZ		Asn719	97.06						
			2CBZ		Asn767	140.84						
			2CBZ		Asn819	119.92						
O95297	Myelin protein zero-like protein 1	MPZL1_HUMAN	Model on 1neuA		Asn50	66.93	269	Y	Y	3	Y	29 kDa
			Model on 1neuA	43	Asn133	35.78						
Q04721	Neurogenic locus notch homolog protein 2	NOTC2_HUMAN	2F8Y		Asn1922	89.22	2471	Y	Y	9	Y	265 kDa
Q9UM47	Neurogenic locus notch homolog protein 3	NOTC3_HUMAN	Model on 2F8yA	53	Asn2066	47.74	2321	Y	Y	4	Y	244 kDa
Q9Y4D7	Plexin-D1	PLXD1_HUMAN	Model on 3ig3A	49	Asn1841	0.34	1925	Y	Y	20	Y	212 kDa
			Model on 3ig3A		Asn1559	48.75						
Q15437	Protein transport protein Sec23B	SC23B_HUMAN	Model on 2nupA	86	Asn241	60.29	767	N	Y	5	P	86 kDa
			Model on 2nupA		Asn338	68.48						
			Model on 2nupA		Asn367	30.64						
			Model on 2nupA		Asn380	41.48						
			Model on 2nupA		Asn598	146.73						
P53992	Protein transport protein Sec24C	SC24C_HUMAN	3EH2		Asn868	9.91	1094	N	Y	5	P	118 kDa
			3EH2		Asn892	71						
			3EH2		Asn774	67.42						
Q8WTV0-4	Isoform 4 of Scavenger receptor class B member 1	SCRB1_HUMAN	N/A				552	Y	Y	9	Y	54 kDa
Q15270	Serine palmitoyltransferase 2	SPTC2_HUMAN	N/A				562	N	Y	1	P	63 kDa
Q9Y5N6	Suppressor of tumorigenicity 14 protein	ST14_HUMAN	3NE1		Asn772	73.35	855	N	Y	4	Y	95 kDa
Q6NUQ8	Transmembrane protein 214	TM214_HUMAN	N/A				689	N	Y	4	Y	77 kDa
Q9H1E5	Thioredoxin-related transmembrane protein 4	TMX4_HUMAN	Model on 1x5wA	50	Asn46	47.43	349	Y	Y	1	P	39 kDa
Q8NFA0	Ubiquitin carboxyl-terminal hydrolase 32	UBP32_HUMAN	Model on 2y6eA	58	Asn748	17.61	1604	N	Y	13	P	182 kDa
			Model on 2y6eA		Asn772	9.21						
			Model on 2y6eA		Asn875	35.5						
					Mean	76.20						
					SD	38.13						

MCF7 secreted (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Transmembrane	ASN SITE (NO)	Glycoprotein	Mol Weight
P05067	Amyloid beta A4 protein	A4_HUMAN	1UMH		Asn467	74.25	770	Y	Y	2	Y	87 kDa
P01011	Alpha-1-antichymotrypsin	AACT_HUMAN	1ASH		Asn93	112.72	423	Y	N	6	Y	48 kDa
			1ASH		Asn106	88.7						
			1ASH		Asn127	26.52						
			1ASH		Asn186	123.36						
			1ASH		Asn271	86.61						
O14672	Disintegrin and metalloproteinase domain-containing protein 10	ADA10_HUMAN	Model on 2ao7A	92	Asn651	71.72	748	Y	Y	4	Y	84 kDa
Q06481	Amyloid-like protein 2	APLP2_HUMAN	Model on 3umh	69	Asn541	79.78	763	Y	Y	1	Y	87 kDa
P15291	Beta-1,4-galactosyltransferase 1	B4GT1_HUMAN	N/A				398	N	Y	1	Y	44 kDa
P10275	Bone morphogenetic protein 7	BMP7_HUMAN	1LX1		Asn372	65.32	431	Y	N	4	Y	49 kDa
P07355	Annexin A2	ANXA2_HUMAN	N/A				568	Y	N	6	Y	39 kDa
P09668	Pro-cathepsin H	CATH_HUMAN	Model on 5jchA	92	Asn230	98.87	335	Y	N	3	Y	37 kDa
P16870	Carboxypeptidase E	CBPE_HUMAN	Model on 2nsmA	54	Asn139	62.26	476	Y	N	2	Y	53 kDa
			Model on 2nsmA		Asn390	98.56						
Q6YHK3	CD109 antigen	CD109_HUMAN	N/A				1445	Y	Y	19	Y	162 kDa
P10909	Clusterin	CLUS_HUMAN	N/A				449	Y	N	7	Y	52 kDa
P02461	Collagen alpha-3(V) chain	CO5A3_HUMAN	N/A				1031	Y	N	5	Y	104 kDa
Q99715	Collagen alpha-1(XII) chain	COCA1_HUMAN	N/A				3063	Y	N	9	Y	333 kDa
P39060	Collagen alpha-1(XVIII) chain	COIA1_HUMAN	N/A				1754	Y	N	5	Y	178 kDa
O94985-2	Isoform 2 of Calsyntenin-1	CSTN1_HUMAN	N/A				981	Y	Y	4	Y	109 kDa
Q9H4D0	Calsyntenin-2	CSTN2_HUMAN	N/A				955	Y	Y	6	Y	107 kDa
Q9BQT9	Calsyntenin-3	CSTN3_HUMAN	N/A				956	Y	Y	5	Y	106 kDa
Q14118	Dystroglycan	DAG1_HUMAN	Model on 3u2cA	95	Asn141	114.54	895	Y	Y	6	Y	97 kDa
			1E0d		Asn882	196.79						
O94907	Dickkopf-related protein 1	DKK1_HUMAN	N/A				266	Y	N	1	Y	29 kDa
P30040	Endoplasmic reticulum resident protein 29	ERP29_HUMAN	2QC7		Asn187	28.6	261	Y	N	1	Y	29 kDa
P02765	Alpha-2-HS-glycoprotein	FETUA_HUMAN	N/A				367	Y	N	2	Y	39 kDa
P02751	Fibronectin	FINC_HUMAN	3M7P		Asn430	74.08	2386	Y	N	11	Y	263 kDa
			3M7P		Asn628	83.89						
			3M7P		Asn542	42.9						
			1FNFN		Asn1236	80.39						
			1FNFN		Asn1284	48.71						
			1FNFN		Asn1417	140.61						
			3RSQ		Asn1813	61.87						
Q9BTY2	Plasma alpha-L-fucosidase	FUCO2_HUMAN	N/A				467	Y	N	4	Y	54 kDa
Q99988	Growth/differentiation factor 15	GDF15_HUMAN	N/A				308	Y	N	2	Y	34 kDa
Q92820	Gamma-glutamyl hydrolase	GGH_HUMAN	1L9K		Asn116	70.56	318	Y	N	4	Y	36 kDa
			1L9K		Asn163	66.09						
			1L9K		Asn203	48.44						
			1L9K		Asn307	93.36						
P35052	Glypican-1	GPC1_HUMAN	4AOI		Asn79	103.6	558	Y	Y	2	Y	62 kDa
			4AOI		Asn116	73.08						
O75487	Glypican-4	GPC4_HUMAN	Model on 4acrB	47	Asn268	31.08	556	Y	Y	1	Y	62 kDa
			Model on 4acrB		Asn406	78.51						
P05154	Plasma serine protease inhibitor	IPSP_HUMAN	3DY0		Asn248	118.66	406	Y	N	3	Y	46 kDa
			3DY0		Asn263	107.12						
			3DY0		Asn338	133.47						
P55268	Laminin subunit beta-2	LAMB2_HUMAN	Model on 4aqsA	83	Asn248	23.47	1798	Y	N	7	Y	196 kDa
			Model on 4aqsA		Asn868	86.45						
P11047	Laminin subunit gamma-1	LAMC1_HUMAN	Model on 4aqtA	97	Asn60	115.11	1609	Y	N	14	Y	178 kDa
			Model on 4aqtA		Asn134	45.97						
Q08380	Galectin-3-binding protein	LG3BP_HUMAN	1BY2		Asn69	69.64	585	Y	N	7	Y	65 kDa
			1BY2		Asn125	133.42						
Q14766	Latent-transforming growth factor beta-binding protein 1	LTBP1_HUMAN	Model on 3lmjA	50	Asn1197	120.57	1721	Y	N	8	Y	187 kDa
			Model on 3lmjA	99	Asn1386	97.73						

N = No
Y = Yes
P = Potential
N/A = Not available

P10253	Lysosomal alpha-glucosidase	LYAG_HUMAN	Model on 3l4ya	44	Asn140	125.92	952	N	Y	7	Y	105 kDa
			Model on 3l4ya		Asn470	34.89						
			Model on 3l4ya		Asn652	69.47						
			Model on 3l4ya		Asn882	75.95						
			Model on 3l4ya		Asn925	91.02						
O95274	Ly6/PLAUR domain-containing protein 3	LYPD3_HUMAN	N/A				346	Y	Y	6	Y	36 kDa
P33908	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	MA1A1_HUMAN	Model on 1nxca	92	Asn513	71.29	653	Y	Y	1		73 kDa
P69849	Nodal modulator 3	NOMO3_HUMAN	N/A				1222	Y	Y	7	Y	134 kDa
Q8UXI9	Nephronectin	NPNT_HUMAN	N/A				565	Y	N	1	Y	62 kDa
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	PGBM_HUMAN	N/A				4391	Y	N	10	Y	469 kDa
Q60568	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3_HUMAN	N/A				738	Y	N	2	Y	85 kDa
P23470	Receptor-type tyrosine-protein phosphatase gamma	PTPRG_HUMAN	3JXH		Asn109	92.44	1445	Y	Y	13		162 kDa
			3JXH		Asn113	90.13						
			3JXH		Asn156	113.58						
			2NLK		Asn863	139.8						
			2NLK		Asn1181	49.39						
Q92626	Peroxidasin homolog	PXDN_HUMAN	N/A				1479	Y	N	11	Y	165 kDa
O00391	Sulfhydryl oxidase 1	QSOX1_HUMAN	3QJ6O		Asn130	149.4	747	Y	Y	4	Y	83 kDa
			3QJ6O		Asn243	125.74						
Q9Y6N7	Roundabout homolog 1	ROBO1_HUMAN	4HLJ		Asn820	103.57	1651	Y	Y	15	Y	181 kDa
			4HLJ		Asn827	96.61						
Q8UXD5-3	Isoform 3 of Seizure 6-like protein 2	SEZ62_HUMAN	N/A				930	Y	Y	9	Y	92 kDa
Q99985	Semaphorin-3C	SEM3C_HUMAN	Model on 1q47A	51	Asn81	10.15	751	Y	N	7	Y	85 kDa
			Model on 1q47A		Asn123	62.24						
			Model on 1q47A		Asn252	154.01						
			Model on 1q47A		Asn268	40.77						
			Model on 1q47A		Asn465	112.35						
Q13275	Cluster of Semaphorin-3F	SEM3F_HUMAN	Model on 1q47A	54	Asn53	85.05	785	Y	N	2	Y	88 kDa
			Model on 1q47A		Asn126	28.49						
Q9HAT2	Sialate O-acetylesterase	SIAE_HUMAN	N/A				523	Y	N	6	Y	58 kDa
Q92673	Sortilin-related receptor	SORL_HUMAN	Model on 1j8e	60	Asn1164	81.91	2214	Y	Y	27	Y	248 kDa
			Model on 1j8e		Asn1191	60.75						
			Model on 2dm4A	92	Asn1706	46.31						
Q99523	Sortilin	SORT_HUMAN	3F6K		Asn98	134.06	831	Y	Y	7	Y	92 kDa
			3F6K		Asn162	99.51						
			3F6K		Asn274	139.63						
			3F6K		Asn406	68.9						
			3F6K		Asn582	19.36						
			3F6K		Asn684	127.41						
P52823	Stanniocalcin-1	STC1_HUMAN	N/A				247	Y	N	2	Y	28 kDa
O76061	Stanniocalcin-2	STC2_HUMAN	N/A				302	Y	N	1	Y	33 kDa
Q8IWU5-2	Isoform 2 of Extracellular sulfatase Sulf-2	SULF2_HUMAN	N/A				870	Y	Y	12	Y	100 kDa
Q24JP5	Transmembrane protein 132A	T132A_HUMAN	N/A				1023	Y	Y	3	Y	110 kDa
P01033	Metalloproteinase inhibitor 1	TIMP1_HUMAN	3V96		Asn101	83.1	207	Y	N	2	Y	23 kDa
					Mean	85.65						
					SD	35.47						

MDA468 Microsome (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Trans-membrane	ASN SITE (NO)	Glycoprotein	Mol Weight
P28288	ATP-binding cassette sub-family D member 3	ABCD3_HUMAN	N/A				659	Y	Y	1	N	75 kDa
O00767	Acyl-CoA desaturase	ACOD_HUMAN	N/A				359	Y	P	3	N	42 kDa
P15144	Aminopeptidase N	AMPN_HUMAN	N/A				967	Y	Y	11	Y	110 kDa
Q9HD20	Probable cation-transporting ATPase 13A1	AT131_HUMAN	N/A				1204	Y	Y	6	N	133 kDa
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	AT2A2_HUMAN	Model on 3ard	84	Asn19	78.53	1042	Y	P	7	N	115 kDa
			Model on 3ard		Asn421	10.53						
			Model on 3ard		Asn589	71.33						
			Model on 3ard		Asn738	6.43						
			Model on 3ard		Asn918	46.98						
			Model on 3ard		Asn962	75.61						
P23634	Cluster of Plasma membrane calcium-transporting ATPase 4	AT2B4_HUMAN	N/A				1241	Y	P	3	N	138 kDa
Q8NHH9	Atlastin-2	ATLA2_HUMAN	Model on 4IDO	73	Asn163	116.43	585	Y	P	2	N	66 kDa
			Model on 4IDO		Asn204	1.64						
Q6DD88	Atlastin-3	ATLA3_HUMAN	Model on 3q5E	69	Asn173	2.3	541	Y	P	5	N	61 kDa
			Model on 3q5E		Asn312	11.41						
P35613-2	Isoform 2 of Basigin	BASI_HUMAN	3B8I		Asn160	82.87	385	N	Y	5	Y	29 kDa
Q9NR09	Baculoviral IAP repeat-containing protein 6	BIRC6_HUMAN	N/A				4857	Y	Y	28	Y	530 kDa
P27824	Calnexin	CALX_HUMAN	N/A				592	Y	Y	1	Y	68 kDa
Q96A33	Coiled-coil domain-containing protein 47	CCD47_HUMAN	N/A				488	Y	Y	1	Y	56 kDa
Q96S66	Chloride channel CLIC-like protein 1	CLCC1_HUMAN	N/A				551	Y	Y	1	Y	62 kDa
Q96005	Cleft lip and palate transmembrane protein 1	CLPT1_HUMAN	N/A				669	Y	Y	6	N	76 kDa
Q16850	Lanosterol 14-alpha demethylase	CP51A_HUMAN	3ld6		Asn183	32.36	503	Y	Y	1	Y	57 kDa
Q9UBM7	7-dehydrocholesterol reductase	DHCR7_HUMAN	N/A				475	Y	P	2	N	54 kDa
Q9BW60	Elongation of very long chain fatty acids protein 1	ELOV1_HUMAN	N/A				279	Y	P	2	N	33 kDa
Q9NYP7	Elongation of very long chain fatty acids protein 5	ELOV5_HUMAN	N/A				299	Y	P	1	N	35 kDa
Q8N766	ER membrane protein complex subunit 1	EMC1_HUMAN	N/A				993	Y	Y	3	Y	112 kDa
P33947	Cluster of ER lumen protein retaining receptor 2	ERD22_HUMAN	N/A				212	Y	P	1	N	24 kDa
Q9UKR5	Probable ergosterol biosynthetic protein 28	ERG28_HUMAN	N/A				140	Y	P	1	N	16 kDa
Q7Z2K6	Endoplasmic reticulum metallopeptidase 1	ERMP1_HUMAN	Model on 3PB6X	40	Asn182	52.21	904	Y	Y	4	N	100 kDa
Q96SI9	Extended synaptotagmin-1	ESYT1_HUMAN	Model on 2dmga	46	Asn1062	54.85	1104	Y	P	3	N	123 kDa
A0FGR8-6	Isoform 6 of Extended synaptotagmin-2	ESYT2_HUMAN	N/A				921	Y	P	7	N	105 kDa
P37268	Squalene synthase	FDFT_HUMAN	3VJ8		Asn48	56.35	417	Y	P	2	N	48 kDa
			3VJ8		Asn282	22.26						
Q10471	Polypeptide N-acetylgalactosaminyltransferase 2	GALT2_HUMAN	2FFU		Asn616	139.38	571	Y	P	1	Y	65 kDa
P14314	Glucosidase 2 subunit beta	GLU2B_HUMAN	N/A				528	N	Y	2	Y	59 kDa
Q92896	Golgi apparatus protein 1	GSLG1_HUMAN	N/A				1179	Y	Y	5	Y	135 kDa
Q8TCT9	Minor histocompatibility antigen H13	HM13_HUMAN	N/A				377	Y	Y	4	Y	41 kDa
P07099	Epoxide hydrolase 1	HYEP_HUMAN	N/A				455	Y	P	2	Y	53 kDa
P23229-2	Isoform Alpha-6X1A of Integrin alpha-6	ITA6_HUMAN	N/A				1130	Y	Y	10	Y	119 kDa
Q14571	Inositol 1,4,5-trisphosphate receptor type 2	ITPR2_HUMAN	Model on 3jrra	82	Asn82	154.25	2701	Y	P	18	N	308 kDa
			Model on 3jrra		Asn103	10.58						
Q14573	Inositol 1,4,5-trisphosphate receptor type 3	ITPR3_HUMAN	Model on 3jrra	82	Asn215	133.79	2671	Y	P	16	N	304 kDa
			Model on 3uj0a	75	Asn104	4.01						
			Model on 3uj0a		Asn216	111.91						
Q92608	Acyl-CoA:lysophosphatidylglycerol acyltransferase 1	LGAT1_HUMAN	N/A				370	Y	P	2	N	43 kDa
Q96N66	Lysophospholipid acyltransferase 7	MBOA7_HUMAN	N/A				472	Y	P	1	N	53 kDa
Q33724	Mannosyl-oligosaccharide glucosidase	MOG1_HUMAN	N/A				837	Y	Y	2	Y	92 kDa
Q6PIU2	Neutral cholesterol ester hydrolase 1	NCEH1_HUMAN	Model on 3ainA	40	Asn270	8.01	408	Y	Y	3	Y	46 kDa
			Model on 3ainA		Asn287	42.73						
Q969V3	Nicalin	NCLN_HUMAN	N/A				563	Y	Y	2	Y	63 kDa
P08473	Neprilysin	NEP_HUMAN	1r1H		Asn145	64.42	750	Y	Y	6	Y	86 kDa
			1r1H		Asn285	113.85						
			1r1H		Asn311	96.09						
			1r1H		Asn325	57.51						
			1r1H		Asn335	147.36						
			1r1H		Asn628	106.17						
Q15738	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL_HUMAN	N/A				373	Y	P	1	N	42 kDa
Q9UHG3	Prenylcysteine oxidase 1	PCYOX_HUMAN	N/A				505	Y	Y	3	Y	57 kDa
Q8TEQ8	GPI ethanolamine phosphate transferase 3	PIGO_HUMAN	N/A				1089	Y	Y	2	Y	119 kDa
Q8IY17	Neuropathy target esterase	PLPL6_HUMAN	N/A				1366	Y	Y	4	Y	150 kDa
P05187	Alkaline phosphatase, placental type	PPB1_HUMAN	12ED		Asn144	118.69	535	N	Y	2	Y	58 kDa
					Asn271	19.37						
P20340	Cluster of Ras-related protein Rab-6A	RAB6A_HUMAN	1VZQ		Asn126	2.09	208	Y	P	1	N	24 kDa
Q8TC12	Retinol dehydrogenase 11	RDH11_HUMAN	N/A				318	Y	P	2	Y	35 kDa
Q8NUM9	All-trans-retinol 13,14-reductase	RETST_HUMAN	N/A				610	Y	Y	1	Y	67 kDa
P04843	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1_HUMAN	N/A				607	Y	Y	2	Y	69 kDa
P04844	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2_HUMAN	N/A				631	Y	Y	3	Y	69 kDa
Q9P2E9	Ribosome-binding protein 1	RRBP1_HUMAN	N/A				1410	Y	P	3	N	152 kDa
Q9NQC3-2	Isoform 2 of Reticulon-4	RTN4_HUMAN	N/A				1192	Y	P	7	N	40 kDa
Q9BXP2	Solute carrier family 12 member 9	S12A9_HUMAN	N/A				914	Y	Y	3	N	96 kDa
Q92504	Zinc transporter SLC39A7	S39A7_HUMAN	N/A				469	Y	P	1	N	50 kDa
P43007	Neutral amino acid transporter A	SATT_HUMAN	N/A				532	Y	Y	3	Y	56 kDa
O75396	Vesicle-trafficking protein SEC22b	SC22B_HUMAN	N/A				215	Y	P	1	N	25 kDa
O15126	Secretory carrier-associated membrane protein 1	SCAM1_HUMAN	N/A				338	Y	P	2	N	38 kDa
Q9UGP8	Translocation protein SEC63 homolog	SEC63_HUMAN	N/A				760	Y	P	2	N	88 kDa
P50454	Serpin H1	SERPH_HUMAN	Model on 3zhaA	99	Asn120	113.05	418	N	Y	2	Y	46 kDa
			Model on 3zhaA		Asn125	64.4						
O43291	Kunitz-type protease inhibitor 2	SPIT2_HUMAN	Model on 4isoB	59	Asn57	31.32	252	Y	Y	2	Y	28 kDa
P43307	Translocon-associated protein subunit alpha	SSRA_HUMAN	N/A				286	Y	Y	2	Y	32 kDa
Q9UNL2	Translocon-associated protein subunit gamma	SSRG_HUMAN	N/A				185	Y	P	2	Y	21 kDa

N = No
Y = Yes
P = Potential
N/A = Not available

P46977	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	STT3A_HUMAN	N/A				705	Y	Y	4	Y	81 kDa
Q8TCJ2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	STT3B_HUMAN	N/A				826	Y	P	6	N	94 kDa
Q8NBJ7	Sulfatase-modifying factor 2	SUMF2_HUMAN	1VMJ		Asn291	82.99	301	N	Y	1	Y	34 kDa
Q16563	Synaptophysin-like protein 1	SYPL1_HUMAN	N/A				299	Y	Y	3	N	29 kDa
Q9NX61	Transmembrane protein 161A	T161A_HUMAN	N/A				479	Y	Y	2	Y	54 kDa
Q05519	Antigen peptide transporter 1	TAP1_HUMAN	N/A				808	Y	P	5	N	87 kDa
Q9NZ01	Very-long-chain enoyl-CoA reductase	TECR_HUMAN	N/A				308	Y	Y	3	N	36 kDa
Q9Z2S4	Transmembrane 9 superfamily member 4	TM9S4_HUMAN	N/A				642	Y	P	3	Y	75 kDa
P49755	Transmembrane emp24 domain-containing protein 10	TMED8_HUMAN	N/A				233	Y	Y	3	Y	25 kDa
Q9Y320	Thioredoxin-related transmembrane protein 2	TMX2_HUMAN	Model on 2dj0A	99	Asn165	132.71	296	Y	P	2	Y	34 kDa
			Model on 2dj0A		Asn185	29.89						
O95292	Vesicle-associated membrane protein-associated protein B/C	VAPB_HUMAN	3RKK		Asn68	98.09	243	Y	P	1	N	27 kDa
P38435	Vitamin K-dependent gamma-carboxylase	VKGC_HUMAN	N/A				758	Y	Y	8	N	88 kDa
					Mean	64.98						
					SD	46.99						

MDA468 Cell surface (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Transmembrane	ASN SITE (NO)	Glycoprotein	Mol Weight
P21589	5'-nucleotidase	5NTD_HUMAN	4H2G		Asn63	72.34	574	Y	Y	4	Y	63 kDa
					Asn311	41.82						
					Asn333	92.24						
					Asn403	128.39						
Q92485	Acid sphingomyelinase-like phosphodiesterase 3b	ASM3B_HUMAN	N/A				455	Y	N	3	Y	51 kDa
P50895	Basal cell adhesion molecule	BCAM_HUMAN	N/A				628	Y	Y	5	Y	67 kDa
P12830	Cadherin-1	CADH1_HUMAN	N/A				882	Y	Y	6	Y	97 kDa
P08571	Monocyte differentiation antigen CD14	CD14_HUMAN	4GLP		Asn87	77.49	375	Y	N	4	Y	40 kDa
					Asn151	61.66						
					Asn282	35.57						
					Asn329	73.13						
Q9H5V8	CUB domain-containing protein 1	CDCP1_HUMAN	N/A				836	Y	Y	35	Y	93 kDa
P14625	Endoplasmin	ENPL_HUMAN	4NH9		Asn307	51.8	803	Y	N	6	Y	92 kDa
			4NH9		Asn217	151.6						
P54753	Cluster of Ephrin type-B receptor 3	EPHB3_HUMAN	Model on 2X10	40	Asn351	119.4	998	Y	Y	3	Y	110 kDa
			Model on 2X10		Asn145	80.76						
P05362	Intercellular adhesion molecule 1	ICAM1_HUMAN	1IAM		Asn130	108.22	532	Y	Y	8	Y	58 kDa
			1IAM		Asn145	58.32						
			1IAM		Asn183	86.77						
			1IAM		Asn202	51.58						
			2OZ4		Asn267	145.78						
			2OZ4		Asn296	163.34						
			2OZ4		Asn385	102.35						
			2OZ4		Asn406	48.16						
P05556	Integrin beta-1	ITB1_HUMAN	3VI3		Asn154	103.93	798	Y	Y	12	Y	88 kDa
			3VI3		Asn197	67.29						
			3VI3		Asn232	114.74						
			3VI3		Asn269	114.84						
			3VI3		Asn363	78.91						
			3VI3		Asn406	86.16						
			3VI3		Asn437	109.5						
P18084	Integrin beta-5	ITB5_HUMAN	Model on 4g1e	56	Asn347	96.88	799	Y	Y	10	Y	88 kDa
			Model on 4g1e		Asn460	54.83						
			Model on 4g1e		Asn477	89.89						
			Model on 4g1e		Asn652	99.16						
			Model on 4g1e		Asn686	84.71						
			Model on 4g1e		Asn694	100.92						
			Model on 4g1e		Asn705	25.07						
Q9Y624	Junctional adhesion molecule A	JAM1_HUMAN	1NBQ		Asn185	114.68	299	Y	Y	1	Y	33 kDa
P15529	Cluster of Membrane cofactor protein	MCP_HUMAN	3OBE		Asn63	130.73	392	Y	Y	3	Y	44 kDa
			3OBE		Asn114	58.45						
			3OBE		Asn279	110.84						
Q96297	Myelin protein zero-like protein 3	MPZL3_HUMAN	Model on 1neuA	43	Asn60	66.98	289	Y	Y	3	Y	29 kDa
			Model on 1neuA		Asn181	15.78						
Q7L5N7	Lysophosphatidylcholine acyltransferase 2	PCAT2_HUMAN	N/A				544	N	Y	1	P	60 kDa
					Mean	85.95						
					SD	34.08						

MDA468 secreted (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Transmembrane	ASN SITE (NO)	Glycoprotein	Mol Weight
P01023	Cluster of Alpha-2-macroglobulin	A2MG_HUMAN	4ACQ		Asn65	146.06	1474	Y	N	8	Y	163 kDa
			4ACQ		Asn70	95.14						
			4ACQ		Asn747	98.85						
			4ACQ		Asn936	108.13						
			4ACQ		Asn10	91.42						
			4ACQ		Asn889	101.89						
			4ACQ		Asn991	72.29						
			4ACQ		Asn1424	114.32						
P05067	Amyloid beta A4 protein	A4_HUMAN	1UMH		Asn467	74.25	770	Y	Y	2	Y	87 kDa
O14672	Disintegrin and metalloproteinase domain-containing protein 10	ADA10_HUMAN	Model on 2ao7A	92	Asn553	71.72	748	Y	Y	4	Y	84 kDa
O00468-6	Isoform 6 of Agrin	AGRIN_HUMAN	Model on 3B83	90	Asn335	54.9	2067	Y	Y	4	Y	215 kDa
Q06481	Amyloid-like protein 2	APLP2_HUMAN	Model on 3umh	69	Asn541	79.78	763	Y	Y	3	Y	87 kDa
P20933	N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	ASPG_HUMAN	1APY		Asn88	75.77	346	Y	N	2	Y	37 kDa
			1APY		Asn308	29.62						
Q9Y2A9	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 3	B3GN3_HUMAN	N/A				372	Y	N	6	Y	43 kDa
P13497	Bone morphogenetic protein 1	BMP1_HUMAN	1EDH		Asn91	109.32	986	Y	N	5	Y	111 kDa
P53634	Dipeptidyl peptidase 1	CATC_HUMAN	1PPF		Asn79	101.78	463	Y	N	4	P	52 kDa
			3PDF		Asn63	87.11						
			3PDF		Asn119	85.83						
			3PDF		Asn276	112.51						
P09668	Pro-cathepsin H	CATH_HUMAN	Model on 3pcha	92	Asn280	98.87	335	Y	N	3	Y	37 kDa
P00450	Ceruloplasmin	CERU_HUMAN	4ENZ		Asn138	141.96	1065	Y	N	7	Y	122 kDa
			4ENZ		Asn227	26.13						
			4ENZ		Asn358	117.74						
			4ENZ		Asn397	92.06						
			4ENZ		Asn688	26.12						
			4ENZ		Asn762	157.07						
			4ENZ		Asn926	31.51						
P10909	Clusterin	CLUS_HUMAN	N/A				449	Y	N	7	Y	52 kDa
P01024	Complement C3	CO3_HUMAN	2A74		Asn65	89.77	1663	Y	N	3	Y	187 kDa
			2WII		Asn939	106.05						
			2A74		Asn1617	119.8						
P0C0L4	Cluster of Complement C4-A	CO4A_HUMAN	N/A				1744	Y	N	1	Y	193 kDa
Q99715	Collagen alpha-1(XII) chain	COCA1_HUMAN	N/A				3063	Y	N	9	Y	333 kDa
Q9H3G5	Probable serine carboxypeptidase CPVL	CPVL_HUMAN	N/A				476	Y	N	4	Y	54 kDa
Q96RW5	Calsyntenin-1	CSTN1_HUMAN	N/A				981	Y	Y	4	Y	110 kDa
P29279	Connective tissue growth factor	CTGF_HUMAN	N/A				349	Y	N	2	Y	38 kDa
Q08345	Epithelial discoidin domain-containing receptor 1	DDR1_HUMAN	4agd		Asn213	52.72	913	Y	Y	7	Y	101 kDa
			4agd		Asn260	52.09						
			4bkj		Asn777	134.16						
O00115	Deoxyribonuclease-2-alpha	DNS2A_HUMAN	N/A				360	Y	N	4	Y	40 kDa
Q9NZ08	Endoplasmic reticulum aminopeptidase 1	ERAP1_HUMAN	2YD0		Asn70	39.31	941	Y	Y	5	Y	107 kDa
			2YD0		Asn154	116.56						
			2YD0		Asn416	81.81						
			2YD0		Asn790	154.79						
			2YD0		Asn901	58.45						
			3AHQ		Asn280	50.3						
Q14517	Protocadherin Fat 1	FAT1_HUMAN	N/A				4588	Y	Y	29	Y	506 kDa
Q9NYQ8	Protocadherin Fat 2	FAT2_HUMAN	N/A				4349	Y	Y	41	Y	479 kDa
P23142	Cluster of Fibulin-1	FBLN1_HUMAN	N/A				703	Y	N	3	Y	77 kDa
Q12805	EGF-containing fibulin-like extracellular matrix protein 1	FBLN3_HUMAN	N/A				493	Y	N	1	Y	55 kDa
P02765	Alpha-2-HS-glycoprotein	FETUA_HUMAN	N/A				367	Y	N	2	Y	39 kDa
P02751	Fibronectin	FINC_HUMAN	3M7P		Asn830	74.08	2386	Y	N	33	Y	263 kDa
			3M7P		Asn928	88.89						
			3M7P		Asn642	42.9						
			1FNFN		Asn1286	80.39						
			1FNFN		Asn1244	40.71						
			1FNFN		Asn1617	140.61						
			5HBL		Asn1013	63.87						
Q86XX4-2	Isoform 2 of Extracellular matrix protein FRAS1	FRAS1_HUMAN	N/A				4008	Y	Y	17	Y	444 kDa
Q92820	Gamma-glutamyl hydrolase	GGH_HUMAN	1L9X		Asn116	70.56	318	Y	N	4	Y	36 kDa

N = No
Y = Yes
P = Potential
N/A = Not available

			1L9K		Asn163	66.09						
			1L9K		Asn203	49.44						
			1L9K		Asn307	93.36						
P35052	Glypican-1	GPC1_HUMAN	4AOI		Asn79	103.6	558	Y	Y	2	Y	62 kDa
			4AOI		Asn116	73.03						
P06865	Beta-hexosaminidase subunit alpha	HEXA_HUMAN	2GJX		Asn115	62.89	529	Y	Y	3	Y	61 kDa
			2GJX		Asn157	41.28						
			2GJX		Asn295	120.2						
P07686	Beta-hexosaminidase subunit beta	HEXB_HUMAN	1NOW		Asn84	122.62	556	Y	Y	5	Y	63 kDa
			1NOW		Asn142	48.29						
			1NOW		Asn190	25.5						
			1NOW		Asn327	130.22						
			1NOW		Asn497	57.09						
Q9Y4L1	Hypoxia up-regulated protein 1	HYOU1_HUMAN	N/A				999	Y	N	9	Y	111 kDa
P17936	Insulin-like growth factor-binding protein 3	IBP3_HUMAN	Model on 1h59b	66	Asn116	102.3	291	Y	N	3	Y	32 kDa
Q969P0	Immunoglobulin superfamily member 8	IGSF8_HUMAN	N/A				813	Y	Y	3	Y	89 kDa
Q9Y337	Kallikrein-5	KLK5_HUMAN	2PSX		Asn69	77.48	293	Y	N	4	Y	32 kDa
			2PSX		Asn173	115.47						
			2PSX		Asn208	77.12						
			2PSX		Asn252	133.98						
Q16363	Laminin subunit alpha-4	LAMA4_HUMAN	N/A				1823	Y	N	20	Y	203 kDa
O15230	Laminin subunit alpha-5	LAMA5_HUMAN	N/A				3695	Y	N	25	Y	400 kDa
P07942	Laminin subunit beta-1	LAMB1_HUMAN	Model on 4aqs	93	Asn320	134.53	1786	Y	N	12	Y	198 kDa
			Model on 1klbA	44	Asn356	89.99						
P55268	Laminin subunit beta-2	LAMB2_HUMAN	Model on 4aqsA	68	Asn248	23.47	1798	Y	N	7	Y	196 kDa
			Model on 4aqsA		Asn168	86.45						
P11047	Laminin subunit gamma-1	LAMC1_HUMAN	4AQT		Asn60	122.81	1609	Y	N	14	Y	178 kDa
			4AQT		Asn184	29.91						
Q13753	Laminin subunit gamma-2	LAMC2_HUMAN	N/A				1193	Y	N	6	Y	131 kDa
P01130	Low-density lipoprotein receptor	LDLR_HUMAN	1IJQ		Asn615	136.09	860	Y	Y	5	Y	95 kDa
			1IJQ		Asn657	97.5						
Q99538	Legumain	LGMN_HUMAN	N/A				433	Y	N	5	Y	49 kDa
Q9NR60	Lysyl oxidase homolog 2	LOXL2_HUMAN	N/A				778	Y	N	3	Y	87 kDa
Q8N152	Ly6/PLAUR domain-containing protein 6B	LY6D_HUMAN	N/A				183	Y	N	1	Y	21 kDa
Q9NS15	Cluster of Latent-transforming growth factor beta-binding protein 3	LTBP3_HUMAN	N/A				1303	Y	N	5	Y	139 kDa
P10253	Lysosomal alpha-glucosidase	LYAG_HUMAN	Model on 3lkya	44	Asn140	125.92	952	N	Y	7	Y	105 kDa
			Model on 3lkya		Asn233	17.49						
			Model on 3lkya		Asn890	24.16						
			Model on 3lkya		Asn670	34.89						
			Model on 3lkya		Asn652	69.47						
			Model on 3lkya		Asn802	25.95						
			Model on 3lkya		Asn925	90.02						
Q16706	Alpha-mannosidase 2	MA2A1_HUMAN	Model on 3bmA	43	Asn1125	105.28	1164	N	Y	5	Y	133 kDa
			Model on 3bmA		Asn1190	27.58						
P08581	Hepatocyte growth factor receptor	MET_HUMAN	2ZCL		Asn1171	59.14	1390	Y	Y	17	Y	156 kDa
			2ZCL		Asn1239	29.83						
P11717	Cation-independent mannose-6-phosphate receptor	MPRI_HUMAN	2V5O		Asn1656	66.28	2491	Y	Y	21	Y	274 kDa
			2V5O		Asn1757	100.52						
			2V5O		Asn1816	86.86						
			2V5O		Asn2005	156.36						
Q8WXI7	Mucin-16	MUC16_HUMAN	N/A				22152	N	Y	46	Y	2353 kDa
P80188	Neutrophil gelatinase-associated lipocalin	NGAL_HUMAN	1X71		Asn65	45.97	198	Y	N	1	Y	23 kDa
Q8IUX6	Olfactomedin-4	OLFM4_HUMAN	N/A				510	Y	N	6	Y	57 kDa
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	PGBM_HUMAN	N/A				4391	Y	N	10	Y	469 kDa
Q02809	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	PLOD1_HUMAN	N/A				727	Y	Y	4	Y	84 kDa
Q96NY8	Poliovirus receptor-related protein 4	PVRL4_HUMAN	N/A				530	Y	Y	3	Y	55 kDa
Q92626	Peroxidasin homolog	PXDN_HUMAN	N/A				1479	Y	N	11	Y	165 kDa
O00391	Sulfhydryl oxidase 1	QSOX1_HUMAN	3QJO		Asn310	149.4	747	Y	Y	4	Y	83 kDa
			3QJO		Asn243	125.74						
Q13275	Semaphorin-3F	SEM3F_HUMAN	Model on 1q47	54	Asn653	85.05	705	Y	N	7	Y	80 kDa
			Model on 1q47		Asn126	28.49						
Q92673	Sortilin-related receptor	SORL_HUMAN	Model on 1j8e	60	Asn1164	81.91	2214	Y	Y	27	Y	248 kDa
			Model on 1j8e		Asn1191	60.75						
			Model on 2dm6a	92	Asn1706	46.31						
Q99523	Sortilin	SORT_HUMAN	3F6K		Asn93	134.06	831	Y	Y	7	Y	92 kDa
			3F6K		Asn162	99.51						
			3F6K		Asn274	139.63						
			3F6K		Asn606	68.9						
			3F6K		Asn634	127.41						
O43278	Kunitz-type protease inhibitor 1	SPIT1_HUMAN	N/A				529	Y	Y	3	Y	58 kDa
Q9Y5Y6	Suppressor of tumorigenicity 14 protein	ST14_HUMAN	N/A				855	N	Y	4	Y	95 kDa
P52823	Stanniocalcin-1	STC1_HUMAN	N/A				247	Y	N	2	Y	28 kDa
O76061	Stanniocalcin-2	STC2_HUMAN	N/A				302	Y	N	1	Y	33 kDa
Q6UB95	Transmembrane protein 132A	T132A_HUMAN	N/A				1023	Y	Y	3	Y	110 kDa
P24821	Tenascin	TENA_HUMAN	Model on 3tes	58	Asn1079	133.33	2203	Y	N	23	Y	241 kDa
			Model on 3tes		Asn1093	47.51						
			Model on 3tes		Asn1119	98.76						
			Model on 3tes	50	Asn1134	60.08						
			Model on 3tes		Asn1230	109.67						
P61812	Transforming growth factor beta-2	TGFB2_HUMAN	N/A				414	Y	N	3	Y	48 kDa
P49788	Retinoic acid receptor responder protein 1	TIG1_HUMAN	N/A				298	Y	N	1	P	33 kDa
Q9ZDM7	Tubulointerstitial nephritis antigen-like	TINAL_HUMAN	N/A				467	Y	N	2	Y	52 kDa
P02787	Serotransferrin	TRFE_HUMAN	4H0W		Asn432	107.27	698	Y	N	2	Y	77 kDa
			4H0W		Asn630	130.6						
Q6EMK4	Vasorin	VASN_HUMAN	N/A				673	Y	Y	5	Y	72 kDa
P25311	Zinc-alpha-2-glycoprotein	ZA2G_HUMAN	1ZTV		Asn109	182.94	298	Y	N	4	Y	34 kDa
			1ZTV		Asn132	123.92						
			1ZTV		Asn228	78.27						
			1ZTV		Asn259	88.93						
					Mean	85.44						
					SD	36.85						

MCF10A Microsome [Unique]

Acc No	Identified Proteins	UniprotID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Trans-membrane	ASN SITE (NG)	Glycoprotein	Mol Weight
P33527	ATP-binding cassette sub-family C member 1	ABCC1_HUMAN	N/A				1531	N	Y	3	Y	171 kDa
O00767	Acyl-CoA desaturase	ACOD_HUMAN	N/A				359	N	Y	3	P	42 kDa
Q9HD20	Probable cation-transporting ATPase 13A1	AT131_HUMAN	N/A				1204	N	Y	6	Y	133 kDa
Q8WXF7	Atlastin-1	ATLA1_HUMAN	4IDQ		Asn177	3.53	558	N	Y	3	P	64 kDa
			4IDQ		Asn283	74.38						
			4IDQ		Asn436	0.8						
Q96G23	Ceramide synthase 2	CERS2_HUMAN	Model on 1a2mA	45	Asn61	76.1	380	N	Y	2	Y	45 kDa
P53569	Low affinity cationic amino acid transporter 2	CTR2_HUMAN	N/A				650	N	Y	4	Y	72 kDa
Q53GQ0	Estradiol 17-beta-dehydrogenase 12	DHB12_HUMAN	N/A				312	N	Y	2	P	34 kDa
P37059	Estradiol 17-beta-dehydrogenase 2	DHB2_HUMAN	N/A				387	Y	Y	1	P	43 kDa
Q9BWG0	Elongation of very long chain fatty acids protein 3	ELOV1_HUMAN	N/A				279	N	Y	2	P	33 kDa
Q8N766	ER membrane protein complex subunit 1	EMC1_HUMAN	N/A				998	Y	Y	3	Y	112 kDa
P33947	Cluster of ER lumen protein retaining receptor 2	ERD22_HUMAN	N/A				212	N	Y	1	P	24 kDa
Q75477	Erlin-1	ERLN1_HUMAN	N/A				346	Y	N	1	Y	39 kDa
Q96905	Erlin-2	ERLN2_HUMAN	N/A				339	Y	Y	1	Y	38 kDa
Q72696	Endoplasmic reticulum metallopeptidase 1	ERMP1_HUMAN	Model on 3P8BX	40	Asn182	52.21	908	N	Y	4	Y	100 kDa
A0FGR8-2	Cluster of Isoform 2 of Extended synaptotagmin-2	ESYT2_HUMAN	4NPK		Asn433	46.4	923	N	Y	7	P	99 kDa
			4NPK		Asn533	78.6						
			4NPK		Asn637	52.83						
			4NPK		Asn643	98.77						
Q53EP0	Fibronectin type III domain-containing protein 3B	FND3B_HUMAN	Model on 1x9dA	62	Asn109	62.52	1204	N	Y	5	P	133 kDa
P11169	Solute carrier family 2, facilitated glucose transporter member 3	GTR3_HUMAN	N/A				496	N	Y	2	Y	54 kDa
Q9P035	Very-long-chain (3R)-3-hydroxyacyl-[acyl-carrier protein] dehydratase 3	HACD3_HUMAN	N/A				362	N	Y	3	P	43 kDa
Q14571	Cluster of Inositol 1,4,5-trisphosphate receptor type 2	ITPR2_HUMAN	Model on 3jmA	82	Asn62	154.25	2701	N	Y	18	P	308 kDa
			Model on 3jmA		Asn1003	10.53						
			Model on 3jmA		Asn2215	133.79						
P13473	Lysosome-associated membrane glycoprotein 2	LAMP2_HUMAN	N/A				410	Y	Y	16	Y	45 kDa
Q14728	Major facilitator superfamily domain-containing protein 10	MFS10_HUMAN	N/A				455	N	Y	1	P	48 kDa
O14880	Microsomal glutathione S-transferase 3	MGST3_HUMAN	N/A				152	N	Y	1	P	17 kDa
Q13724	Mannosyl-oligosaccharide glucosidase	MOG1_HUMAN	N/A				837	Y	Y	2	Y	92 kDa
Q969V3	Nicalin	NCLN_HUMAN	N/A				563	Y	Y	2	Y	69 kDa
P69849	Nodal modulator 3	NOMO3_HUMAN	N/A				1222	Y	Y	7	Y	134 kDa
Q9Y639	Neuroplastin	NPTN_HUMAN	Model on 2wv3A	92	Asn171	46.49	398	Y	Y	7	Y	44 kDa
			Model on 2wv3A		Asn197	89.36						

N = No
Y = Yes
P = Potential
N/A = Not available

			Model on 2wv3A		Asn229	72.58						
			Model on 2wv3A		Asn284	77.98						
			Model on 2wv3A		Asn296	71.46						
			Model on 2wv3A		Asn317	82.16						
Q9H0X9	Oxysterol-binding protein-related protein 5	OSBL5_HUMAN	N/A				879	N	Y	5	P	99 kDa
Q92508	Piezo-type mecha nosensitive ion channel component 1	PIEZ1_HUMAN	N/A				2521	Y	Y	8	Y	287 kDa
Q8IY17	Neuropathy target esterase	PLPL6_HUMAN	N/A				1366	Y	Y	4	Y	150 kDa
Q96AA3	Protein RFT1 homolog	RFT1_HUMAN	N/A				541	N	Y	1	P	60 kDa
Q8N357	Solute carrier family 35 member F6	S35F6_HUMAN	N/A				371	Y	Y	3	Y	40 kDa
Q92504	Zinc transporter SLC39A7	S39A7_HUMAN	N/A				469	N	Y	1	P	50 kDa
Q9H2J7	Sodium-dependent neutral amino acid transporter B(0)AT2	S6A15_HUMAN	N/A				730	N	Y	5	Y	82 kDa
O15126	Secretory carrier-associated membrane protein 1	SCAM1_HUMAN	N/A				338	N	Y	2	P	38 kDa
Q9UGP8	Translocation protein SEC63 homolog	SEC63_HUMAN	N/A				760	N	Y	2	P	88 kDa
P35610	Sterol O-acyltransferase 1	SOAT1_HUMAN	N/A				550	N	Y	3	P	65 kDa
O15269	Serine palmitoyltransferase 1	SPTC1_HUMAN	N/A				473	N	Y	1	P	53 kDa
O43291	Kunitz-type protease inhibitor 2	SPIT2_HUMAN	Model on 4iss8	59	Asn67	33.32	252	Y	Y	2	Y	28 kDa
P43307	Translocon-associated protein subunit alpha	SSRA_HUMAN	N/A				296	Y	Y	2	Y	32 kDa
Q8TCJ2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	STT3B_HUMAN	N/A				826	N	Y	6	P	94 kDa
Q05519	Antigen peptide transporter 1	TAP1_HUMAN	N/A				808	N	Y	5	P	87 kDa
Q9NZD1	Very-long-chain enoyl-CoA reductase	TECR_HUMAN	N/A				308	N	Y	3	Y	36 kDa
Q8NBQ4	Transmembrane protein 214	TM214_HUMAN	N/A				699	N	Y	4	Y	77 kDa
Q9H045	Transmembrane 9 superfamily member 3	TM9S3_HUMAN	N/A				599	Y	Y	2	Y	68 kDa
Q92544	Transmembrane 9 superfamily member 4	TM9S4_HUMAN	N/A				642	Y	Y	1	P	75 kDa
Q9Y320	Thioredoxin-related transmembrane protein 2	TMX2_HUMAN	Model on 2kj0A	99	Asn165	132.73	296	Y	Y	2	P	34 kDa
			Model on 2kj0A		Asn185	29.89						
Q99050	V-type proton ATPase 116 kDa subunit a isoform 1	VPP1_HUMAN	N/A				837	N	Y	4	P	96 kDa
					Mean	64.84						
					SD	40.97						

MCF10A Cell surface (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Transmembrane	ASN SITE (NO)	Glycoprotein	Mol Weight
Q96S86	Chloride channel CLIC-like protein 1	CLCC1_HUMAN	N/A				591	Y	Y	1	Y	62 kDa
Q9UBS4	Dnaj homolog subfamily B member 11	DJB11_HUMAN	N/A				358	Y	N	2	Y	41 kDa
O14756	Squalene monooxygenase	ERG1_HUMAN	N/A				574	N	Y	1	P	64 kDa
P78302	Peptidyl-prolyl cis-trans isomerase FKBP9	FKBP9_HUMAN	Model on 2pbcA	53	Asn286	38%	570	Y	N	4	Y	63 kDa
			Model on 2pbcA		Asn302	103.63						
Q8NBJ5	Procollagen galactosyltransferase 1	GT251_HUMAN	N/A				622	Y	N	3	Y	72 kDa
P07099	Epoxide hydrolase 1	HYEP_HUMAN	N/A				455	Y	Y	2	P	53 kDa
Q9NPH3	Interleukin-1 receptor accessory protein	IL1AP_HUMAN	4DEP		Asn67	78.52	570	Y	Y	7	Y	65 kDa
			4DEP		Asn307	116.34						
			4DEP		Asn313	61.35						
			4DEP		Asn338	55.44						
			4DEP		Asn196	58.58						
			4DEP		Asn209	70.53						
			4DEP		Asn299	102.99						
P56199	Integrin alpha-1	ITA1_HUMAN	1PT6		Asn217	111.94	1179	Y	Y	26	Y	131 kDa
			1PT6		Asn837	139.22						
			1PT6		Asn843	82.09						
Q9UIQ6	Leucyl-cystinyl aminopeptidase	LCAP_HUMAN	N/A				1025	Y	Y	18	Y	117 kDa
P15529	Cluster of Membrane cofactor protein	MCP_HUMAN	3OBE		Asn83	130.73	392	Y	Y	3	Y	44 kDa
			3OBE		Asn134	53.45						
			3OBE		Asn273	110.84						
Q9H8H3	Methyltransferase-like protein 7A	MET7A_HUMAN	N/A				244	Y	N	1	P	28 kDa
P08581	Hepatocyte growth factor receptor	MET_HUMAN	3ZCL		Asn1171	99.16	1390	Y	Y	17	Y	156 kDa
			3ZCL		Asn1239	29.01						
Q9UIW1	Multiple inositol polyphosphate phosphatase 1	MINP1_HUMAN	N/A				487	Y	Y	2	Y	55 kDa
O95297	Myelin protein zero-like protein 1	MPZL1_HUMAN	Model on 1neuA	43	Asn60	66.93	269	Y	Y	3	Y	29 kDa
			Model on 1neuA		Asn131	15.78						
Q6PIU2	Neutral cholesterol ester hydrolase 1	NCEH1_HUMAN	Model on 5ainA	40	Asn270	8.01	24.005	72.015	408	Y	Y	3
			Model on 5ainA		Asn287	42.78						
O15118	Niemann-Pick C1 protein	NPC1_HUMAN	3GKH		Asn70	132.02	1278	Y	Y	19	Y	142 kDa
			3GKH		Asn323	81.99						
			3GKH		Asn335	103.42						
			3GKH		Asn385	66.56						
			3GKH		Asn222	66.36						
P32823	Neuronal cell adhesion molecule	NRCAM_HUMAN	Model on 3p3yA	56	Asn83	78.89	1304	Y	Y	20	Y	144 kDa
			Model on 3p3yA		Asn223	51.84						
			Model on 3p3yA		Asn245	62.92						
			Model on 3p3yA		Asn251	54.56						
			Model on 3p3yA		Asn276	112.44						
			Model on 3p3yA		Asn334	145.96						
			Model on 3p3yA		Asn439	48.17						
			Model on 3ueY	98	Asn673	40.25						
			Model on 3ueY		Asn716	5.13						
Q32P28	Prolyl 3-hydroxylase 1	P3H1_HUMAN	N/A				736	Y	N	3	Y	88 kDa
P13674	Prolyl 4-hydroxylase subunit alpha-1	P4HA1_HUMAN	4BTB		Asn233	42.35	534	Y	N	2	Y	61 kDa
O15460	Prolyl 4-hydroxylase subunit alpha-2	P4HA2_HUMAN	Model on 4bt9	63	Asn235	45.81	535	Y	N	2	Y	61 kDa
Q9H7Z7	Prostaglandin E synthase 2	PGES2_HUMAN	Model on 1z9hA	99	Asn363	84.64	377	N	Y	1	P	42 kDa
O00469	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	PLOD2_HUMAN	N/A				737	Y	Y	7	Y	85 kDa
Q9UIW2	Plexin-A1	PLXA1_HUMAN	Model on 3ig1A	49	Asn1341	0.16	1896	Y	Y	16	Y	211 kDa
			Model on 3ig1A		Asn1593	48.75						
Q15293	Reticulocalbin-1	RCN1_HUMAN	N/A				331	Y	N	1	Y	39 kDa
Q8TC12	Retinol dehydrogenase 11	RDH11_HUMAN	Model on 3rd5A	47	Asn174	1.62	318	Y	Y	2	P	35 kDa
			Model on 3rd5A		Asn298	52.4						
Q00765	Receptor expression-enhancing protein 5	REEP5_HUMAN	N/A				189	Y	Y	1	Y	21 kDa
Q9NTJ5	Phosphatidylinositide phosphatase SAC1	SAC1_HUMAN	N/A				587	N	Y	5	P	67 kDa
Q8WTV0-2	Isoform 2 of Scavenger receptor class B member 1	SCRB1_HUMAN	N/A				552	Y	Y	9	Y	57 kDa
P50454	Serpin H1	SERPH_HUMAN	Model on 5zhaA	99	Asn320	113.05	418	Y	N	2	Y	46 kDa
			Model on 5zhaA		Asn325	64.4						
Q9H173	Nucleotide exchange factor SIL1	SIL1_HUMAN	N/A				461	Y	N	2	Y	52 kDa
O43278	Kunitz-type protease inhibitor 1	SPIT1_HUMAN	N/A				529	Y	Y	3	Y	58 kDa
Q8NFQ8	Torsin-1A-interacting protein 2	TOIP2_HUMAN	N/A				470	N	Y	1	Y	51 kDa
O14763	Tumor necrosis factor receptor superfamily member 10B	TR10B_HUMAN	N/A				440	Y	Y	1	P	48 kDa
Q9Y6R9	Testis-expressed sequence 264 protein	TX264_HUMAN	N/A				333	Y	N	1	Y	34 kDa
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	UGGG1_HUMAN	N/A				1555	Y	Y	3	Y	177 kDa
Q9Y6M5	Zinc transporter 1	ZNT1_HUMAN	N/A				507	N	Y	2	Y	55 kDa
					Mean	67.70						
					SD	37.66						

MCF10A secreted (Unique)

Acc No	Identified Proteins	Uniprot ID	PDB or Model	%matches	Asn Site	Accessibility	Protein Length	Signal Peptide	Transmembrane	ASN SITE (NO)	Glycoprotein	Mol Weight
P01009	Alpha-1-antitrypsin	A1AT_HUMAN	2QUG		Asn70	114.59	418	Y	N	3	Y	47 kDa
			2QUG		Asn207	70.94						
			2QUG		Asn271	20.88						
P01023	Cluster of Alpha-2-macroglobulin	A2MG_HUMAN	4ACQ		Asn65	146.06	1474	Y	N	8	Y	163 kDa
			4ACQ		Asn70	95.14						
			4ACQ		Asn247	96.85						
			4ACQ		Asn396	103.31						
			4ACQ		Asn430	93.42						
			4ACQ		Asn869	101.89						
			4ACQ		Asn913	72.29						
			4ACQ		Asn1424	114.32						
P01011	Alpha-1-antichymotrypsin	AACT_HUMAN	1ASH		Asn99	112.72	423	Y	N	6	Y	48 kDa
			1ASH		Asn106	88.7						
			1ASH		Asn127	76.52						
			1ASH		Asn186	123.36						
			1ASH		Asn271	86.61						
O14672	Disintegrin and metalloproteinase domain-containing protein 10	ADA10_HUMAN	Model on 2ao7A	92	Asn655	71.72	748	Y	Y	4	Y	84 kDa
O00468-6	Isoform 6 of Agrin	AGRIN_HUMAN	Model on 3jb3	90	Asn1135	54.9	2067	Y	Y	4	Y	215 kDa
O75882	Attractin	ATRN_HUMAN	Model on 1kkbA	40	Asn1073	162.04	1429	N	Y	26	Y	159 kDa
			Model on 1kkbA		Asn1082	127.44						
P13497	Bone morphogenetic protein 1	BMP1_HUMAN	3EDH		Asn91	109.12	986	Y	N	5	Y	111 kDa
P00736	Complement C1r subcomponent	C1R_HUMAN	1GPZ		Asn534	128.19	705	Y	N	4	Y	80 kDa
			1GPZ		Asn581	139.62						
P09871	Complement C1s subcomponent	C1S_HUMAN	1ELV		Asn106	124.37	688	Y	N	3	Y	77 kDa
			1ELV		Asn682	58.82						
Q9BRK5	45 kDa calcium-binding protein	CAB45_HUMAN	N/A				382	Y	Y	1	Y	42 kDa
P19022	Cadherin-2	CADH2_HUMAN	Model on 3q2w	98	Asn390	103.74	906	Y	Y	9	Y	100 kDa
			Model on 3q2w		Asn273	125.53						
			Model on 3q2w		Asn325	77.68						
			Model on 3q2w		Asn402	82.43						

N = No
Y = Yes
P = Potential
N/A = Not available

			Model on 3q2w		Asn572	73.26						
			Model on 3q2w		Asn622	71.7						
			Model on 3q2w		Asn651	46.13						
			Model on 3q2w		Asn692	52.53						
P53634	Dipeptidyl peptidase 1	CATC_HUMAN	3PDF		Asn29	101.28	463	Y	N	4	P	52 kDa
			3PDF		Asn53	87.11						
			3PDF		Asn119	85.83						
			3PDF		Asn276	112.51						
P07711	Cathepsin L1	CATL1_HUMAN	1CS8		Asn221	89.15	333	Y	N	1	Y	38 kDa
P08603	Complement factor H	CFAH_HUMAN	2WII		Asn217	56.56	1231	Y	N	9	Y	139 kDa
P02462	Collagen alpha-1(IV) chain	CO4A1_HUMAN	N/A				1669	Y	N	4	Y	161 kDa
Q02388	Collagen alpha-1(VII) chain	CO7A1_HUMAN	N/A				2944	Y	N	4	Y	295 kDa
Q99715	Collagen alpha-1(XII) chain	COCA1_HUMAN	N/A				3063	Y	N	9	Y	333 kDa
Q9H3G5	Probable serine carboxypeptidase CPVL	CPVL_HUMAN	N/A				476	Y	N	4	Y	54 kDa
P09603	Macrophage colony-stimulating factor 1	CSF1_HUMAN	1HMC		Asn154	56.76	554	Y	Y	4	Y	60 kDa
			1HMC		Asn172	57.61						
P13611	Versican core protein	CSPG2_HUMAN	N/A				3396	Y	N	24	Y	373 kDa
O94985-2	Isoform 2 of Calsyntenin-1	CSTN1_HUMAN	N/A				981	Y	Y	4	Y	109 kDa
O94907	Dickkopf-related protein 1	DKK1_HUMAN	N/A				266	Y	N	1	Y	29 kDa
Q14517	Protocadherin Fat 1	FAT1_HUMAN	N/A				4588	Y	Y	29	Y	506 kDa
Q9NYQ8	Protocadherin Fat 2	FAT2_HUMAN	N/A				4349	Y	Y	41	Y	479 kDa
P23142	Cluster of Fibulin-1	FBLN1_HUMAN	N/A				703	Y	N	3	Y	77 kDa
Q12805	EGF-containing fibulin-like extracellular matrix protein 1	FBLN3_HUMAN	N/A				493	Y	N	1	Y	55 kDa
P35556	Fibrillin-2	FBN2_HUMAN	N/A				2912	Y	N	12	Y	315 kDa
Q12841	Follistatin-related protein 1	FSTL1_HUMAN	N/A				308	Y	N	3	Y	35 kDa
Q99988	Growth/differentiation factor 15	GDF15_HUMAN	N/A				308	Y	N	2	Y	34 kDa
P07093	Glia-derived nexin	GDN_HUMAN	4DY0		Asn118	54.02	398	Y	Y	2	Y	44 kDa
			4DY0		Asn159	115.07						
Q92820	Gamma-glutamyl hydrolase	GGH_HUMAN	1L9X		Asn116	70.56	318	Y	N	4	Y	36 kDa
			1L9X		Asn163	66.09						
			1L9X		Asn203	43.44						
			1L9X		Asn307	93.36						
P35052	Glypican-1	GPC1_HUMAN	4ACR		Asn79	103.6	558	Y	Y	2	Y	62 kDa
			4ACR		Asn116	73.03						
P24592	Insulin-like growth factor-binding protein 6	IBP6_HUMAN	Model on 3fcs	53	Asn48	95.63	240	Y	N	1	Y	25 kDa
			Model on 3fcs		Asn97	2.72						
			Model on 3fcs		Asn260	135.49						
			Model on 3fcs		Asn387	80.46						
			Model on 3fcs		Asn396	142.45						
			Model on 3fcs		Asn463	99.28						
			Model on 3fcs		Asn471	31.78						
			Model on 3fcs		Asn541	61.38						
			Model on 3fcs		Asn575	59.2						
			Model on 2kv9	66	Asn771	87.37						
Q13007	Interleukin-24	IL24_HUMAN	N/A				206	Y	N	3	Y	24 kDa
P08476	Inhibin beta A chain	INHBA_HUMAN	N/A				426	Y	N	1	Y	47 kDa
P32004	Neural cell adhesion molecule L1	L1CAM_HUMAN	N/A				1257	Y	Y	22	Y	140 kDa
P50851	Lipopolysaccharide-responsive and beige-like anchor protein	LRBA_HUMAN	1t77		Asn2210	53.27	2863	N	Y	15	P	319 kDa
Q86X29	Lipolysis-stimulated lipoprotein receptor	LSR_HUMAN	N/A				649	N	Y	2	P	71 kDa
Q14766	Latent-transforming growth factor beta-binding protein 1	LTBP1_HUMAN	Model on 1lmjA	50	Asn1197	120.57	1721	Y	N	8	Y	187 kDa
			Model on 1ksqA	99	Asn1366	87.73						
P14780	Matrix metalloproteinase-9	MMP9_HUMAN	1L6J		Asn38	128.01	707	Y	N	3	Y	78 kDa
			1L6J		Asn120	68.13						
			1L6J		Asn127	50.26						
Q13421-2	Isoform 3 of Mesothelin	MSLN_HUMAN	N/A				630	Y	N	4	Y	71 kDa
Q04721	Neurogenic locus notch homolog protein 2	NOTC2_HUMAN	2F8Y		Asn1922	89.22	2471	Y	Y	9	Y	265 kDa
O14786	Neuropilin-1	NRP1_HUMAN	2QQM		Asn150	83.36	923	Y	Y	6	Y	103 kDa
			2QQM		Asn261	60.27						
			2QQM		Asn300	84.55						
			2QQM		Asn522	95.01						
P05121	Plasminogen activator inhibitor 1	PAI1_HUMAN	1LJ5		Asn232	26.89	402	Y	N	3	Y	45 kDa
			1LJ5		Asn288	95.3						
			1LJ5		Asn352	60.77						
P07585	Decorin	PGS2_HUMAN	Model on 1xku	92	Asn211	48.49	359	Y	N	3	Y	40 kDa
			Model on 1xku		Asn262	131.33						
			Model on 1xku		Asn303	80.32						
Q92626	Peroxidasin homolog	PXDN_HUMAN	N/A				1479	Y	N	11	Y	165 kDa
O00391	Sulfhydryl oxidase 1	QSOX1_HUMAN	3Q6O		Asn130	149.4	747	Y	Y	4	Y	83 kDa
			3Q6O		Asn243	125.74						
P07602	Proactivator polypeptide	SAP_HUMAN	2D0B		Asn80	134.59	524	Y	N	5	Y	58 kDa
			2D0B		Asn101	107.63						
			1N69		Asn215	68.44						
			2GTG		Asn332	135.24						
			3BQP		Asn425	94.32						
O95486	Protein transport protein Sec24A	SC24A_HUMAN	2NUT		Asn485	46.04	1093	N	Y	8	P	120 kDa
			2NUT		Asn585	84.17						
			2NUT		Asn951	96.79						
			2NUT		Asn959	120.28						
			2NUT		Asn999	100.14						
Q9H4F8	SPARC-related modular calcium-binding protein 1	SMOC1_HUMAN	N/A				434	Y	N	3	Y	48 kDa
P38952	Serpin B5	SPB5_HUMAN	1WZ9		Asn99	118.16	375	Y	N	4	Y	42 kDa
			1WZ9		Asn133	18.77						
			1WZ9		Asn188	6.71						
			1WZ9		Asn361	48.95						
Q24JP5	Transmembrane protein 132A	T132A_HUMAN	N/A				1023	Y	Y	3	Y	110 kDa
P01033	Metalloproteinase inhibitor 1	TIMP1_HUMAN	3V96		Asn101	83.1	207	Y	N	2	Y	23 kDa
O95407	Tumor necrosis factor receptor superfamily member 6B	TNF6B_HUMAN	3K51		Asn173	96.89	300	Y	N	1	Y	33 kDa
P00749	Urokinase-type plasminogen activator	UROK_HUMAN	2O8T		Asn322	110.37	431	Y	N	1	Y	49 kDa
					Mean	**86.56**						
					SD	**33.54**						

Supplementary Table 3A

Subcellular fraction	Proportion of putative *N*-glycoproteins in subcellular proteomes (Number of putative *N*-glycoproteins / Number of total proteins)		
	MCF7	MDA468	MCF10A
	Total proteins = 2297	Total proteins = 2636	Total proteins = 2042
Microsome	22.7% (346/1525)	15.7% (313/1988)	24.1% (337/1399)
Cell surface	15.7% (94/599)	31.0% (45/145)	25.7% (105/490)
Secreted	24.7% (302/1222)	22.4% (302/1348)	19.6% (187/952)

Supplementary Table 3B

Subcellular fraction	Sequon-weighted normalized spectral count of top-100 most abundant putative *N*-glycoproteins / Total putative *N*-glycoproteins		
	MCF7	MDA468	MCF10A
Microsome	26.0/36.9 (70.5%)	17.0/23.6 (72.0%)	20.8/29.4 (70.7%)
Cell surface	7.2/7.2 (100%)	8.9/8.9 (100%)	6.7/6.8 (99.0%)
Secreted	41.3/51.1 (77.5%)	21.8/28.6 (76.2%)	22.1/24.8 (89.1%)

Supplementary Table 4

Cell line	Subcellular fraction	Degree of terminal *N*-glycan determinants / total complex/hybrid type *N*-glycans (%)		
		β-galactosylation	α1,6-fucosylation	α-sialylation
MCF7	Microsome	66.30 ± 2.06	86.47 ± 2.10	38.14 ± 1.41
	Cell surface	91.23 ± 1.66	85.02 ± 4.91	38.15 ± 5.01
	Secreted	68.43 ± 0.25	78.24 ± 1.15	57.55 ± 2.37
	R^2	0.05	(-) 0.61	0.64
MDA468	Microsome	48.36 ± 0.70	60.50 ± 7.59	67.28 ± 3.54
	Cell surface	78.01 ± 0.32	68.62 ± 0.92	69.85 ± 1.92
	Secreted	26.89 ± 2.33	58.59 ± 1.37	94.99 ± 0.30
	R^2	0.00	0.04	0.41
MCF10A	Microsome	46.49 ± 0.61	68.98 ± 1.68	83.77 ± 0.78
	Cell surface	49.12 ± 1.68	75.81 ± 1.07	85.02 ± 1.91
	Secreted	55.98 ± 5.54	82.23 ± 1.89	77.98 ± 3.78
	R^2	0.64	0.81	(-) 0.65

APPENDIX 6: Supplementary data from Publication IV

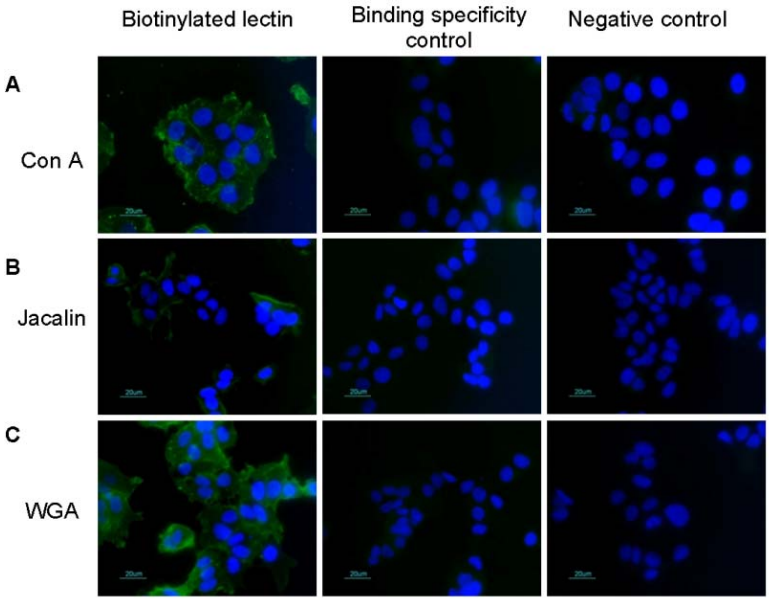


Figure S1. Lectin immunofluorescence staining. MCF7 cells were grown on coverslips, fixed, and incubated with biotinylated lectins: (A) Con A (20 µg/ml), (B) Jac (20 µg/ml) and (C) WGA (5 µg/ml), followed by fluorescence staining using FITC-conjugated Strepatividin and nuclei staining with Hoechst 33342. Lectin binding specificity controls are shown in the middle panel where 10 µg/ml of each lectin was preincubated with its corresponding inhibitory saccharide prior to its addition to the cells. Negative controls were performed without any lectin added to the cells.

REFERENCES

1. *About Breast Cancer*. 2014 [cited 2014 Sep 8]; Available from: http://www.nbcf.org.au/research/about-breast-cancer.aspx.
2. *Breast cancer statistics, Cancer Australia*. 2014 [cited 2014 July 9]; Available from: http://canceraustralia.gov.au/affected-cancer/cancer-types/breast-cancer/breast-cancer-statistics.
3. Ali, S. and R.C. Coombes, *Endocrine-responsive breast cancer and strategies for combating resistance.* Nat Rev Cancer, 2002. **2**(2): p. 101-112.
4. Polyak, K. and R. Kalluri, *The role of the microenvironment in mammary gland development and cancer.* Cold Spring Harb Perspect Biol, 2010. **2**(11): p. a003244.
5. Lu, P., V.M. Weaver, and Z. Werb, *The extracellular matrix: A dynamic niche in cancer progression.* The Journal of Cell Biology, 2012. **196**(4): p. 395-406.
6. Lakhani, S.R., *WHO classification of tumours of the breast*. 4th ed. World Health Organization classification of tumours, ed. S.R. Lakhani. 2012, France: Lyon : International Agency for Research on Cancer, 2012.
7. Perou, C., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747 - 752.
8. Arpino, G., et al., *Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome.* Breast Cancer Res, 2004. **6**(3): p. R149-56.
9. Yerushalmi, R., M.M. Hayes, and K.A. Gelmon, *Breast carcinoma—rare types: review of the literature.* Annals of Oncology, 2009. **20**(11): p. 1763-1770.
10. Viale, G., *The current state of breast cancer classification.* Annals of Oncology, 2012. **23**(suppl 10): p. x207-x210.
11. Sørlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proceedings of the National Academy of Sciences, 2001. **98**(19): p. 10869-10874.
12. Sørlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proceedings of the National Academy of Sciences, 2003. **100**(14): p. 8418-8423.
13. Parker, J.S., et al., *Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes.* Journal of Clinical Oncology, 2009. **27**(8): p. 1160-1167.
14. Falck, A.K., et al., *St Gallen molecular subtypes in primary breast cancer and matched lymph node metastases--aspects on distribution and prognosis for patients with luminal A tumours: results from a prospective randomised trial.* BMC Cancer, 2013. **13**: p. 558.
15. Prat, A., et al., *Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer.* Breast Cancer Research, 2010. **12**(5): p. 1-18.
16. Wilkins, M.R., et al., *Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It.* Biotechnology and Genetic Engineering Reviews, 1996. **13**(1): p. 19-50.
17. Doerr, A., *Targeted proteomics.* Nat Meth, 2010. **7**(1): p. 34-34.
18. Menon, R. and G.S. Omenn, *Proteomic characterization of novel alternative splice variant proteins in human epidermal growth factor receptor 2/neu-induced breast cancers.* Cancer Res, 2010. **70**(9): p. 3440-9.
19. Qian, W.J., et al., *Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications.* Mol Cell Proteomics, 2006. **5**(10): p. 1727-44.
20. Bjorhall, K., T. Miliotis, and P. Davidsson, *Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples.* Proteomics, 2005. **5**(1): p. 307-17.
21. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database.* Biochimica et Biophysica Acta, 1999. **1473**(1): p. 4-8.

22. Thaysen-Andersen, M. and N.H. Packer, *Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome.* Biochim Biophys Acta, 2014. **1844**(9): p. 1437-52.
23. Fanayan, S., M. Hincapie, and W.S. Hancock, *Using lectins to harvest the plasma/serum glycoproteome.* Electrophoresis, 2012. **33**(12): p. 1746-1754.
24. Xu, Y., L. Zhang, and H. Lu, *Use of boronic acid nanoparticles in glycoprotein enrichment.* Methods Mol Biol, 2013. **951**: p. 45-55.
25. Zhang, H., et al., *Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.* Nat Biotechnol, 2003. **21**(6): p. 660-6.
26. Plavina, T., et al., *Combination of Abundant Protein Depletion and Multi-Lectin Affinity Chromatography (M-LAC) for Plasma Protein Biomarker Discovery.* Journal of Proteome Research, 2006. **6**(2): p. 662-671.
27. Sparbier, K., et al., *Analysis of Glycoproteins in Human Serum by Means of Glycospecific Magnetic Bead Separation and LC-MALDI-TOF/TOF Analysis with Automated Glycopeptide Detection.* Journal of Biomolecular Techniques : JBT, 2007. **18**(4): p. 252-258.
28. Zhang, S., et al., *Boronic acid functionalized magnetic nanoparticles via thiol-ene click chemistry for selective enrichment of glycoproteins.* New Journal of Chemistry, 2014. **38**(9): p. 4212-4218.
29. Lee, A., et al., *Rat Liver Membrane Glycoproteome: Enrichment by Phase Partitioning and Glycoprotein Capture.* Journal of Proteome Research, 2009. **8**(2): p. 770-781.
30. Liu, T., et al., *Human Plasma N-Glycoproteome Analysis by Immunoaffinity Subtraction, Hydrazide Chemistry, and Mass Spectrometry.* Journal of proteome research, 2005. **4**(6): p. 2070-2080.
31. Ahn, Y.H., J.Y. Kim, and J.S. Yoo, *Quantitative mass spectrometric analysis of glycoproteins combined with enrichment methods.* Mass Spectrom Rev, 2014.
32. Yates, J.R., *MASS SPECTRAL ANALYSIS IN PROTEOMICS.* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**(1): p. 297-316.
33. Compton, P.D. and N.L. Kelleher, *Spinning up mass spectrometry for whole protein complexes.* Nat Meth, 2012. **9**(11): p. 1065-1066.
34. Rabilloud, T., et al., *Two-dimensional gel electrophoresis in proteomics: Past, present and future.* J Proteomics, 2010. **73**(11): p. 2064-77.
35. Tonge, R., et al., *Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology.* Proteomics, 2001. **1**(3): p. 377-96.
36. Huang, H.-L., et al., *Biomarker discovery in breast cancer serum using 2-D differential gel electrophoresis/ MALDI-TOF/TOF and data validation by routine clinical assays.* ELECTROPHORESIS, 2006. **27**(8): p. 1641-1650.
37. Rui, Z., et al., *Use of serological proteomic methods to find biomarkers associated with breast cancer.* PROTEOMICS, 2003. **3**(4): p. 433-439.
38. Zhang, L., et al., *Discovery and Preclinical Validation of Salivary Transcriptomic and Proteomic Biomarkers for the Non-Invasive Detection of Breast Cancer.* PLoS ONE, 2010. **5**(12): p. e15573.
39. Schulz, D.M., et al., *Identification of Differentially Expressed Proteins in Triple-Negative Breast Carcinomas Using DIGE and Mass Spectrometry.* Journal of Proteome Research, 2009. **8**(7): p. 3430-3438.
40. Fang, Y., D.P. Robinson, and L.J. Foster, *Quantitative Analysis of Proteome Coverage and Recovery Rates for Upstream Fractionation Methods in Proteomics.* Journal of Proteome Research, 2010. **9**(4): p. 1902-1912.
41. Piersma, S.R., et al., *Workflow Comparison for Label-Free, Quantitative Secretome Proteomics for Cancer Biomarker Discovery: Method Evaluation, Differential Analysis, and Verification in Serum.* Journal of Proteome Research, 2010. **9**(4): p. 1913-1922.
42. Washburn, M.P., D. Wolters, and J.R. Yates, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotech, 2001. **19**(3): p. 242-247.

43. Chen, E.I., et al., *Large Scale Protein Profiling by Combination of Protein Fractionation and Multidimensional Protein Identification Technology (MudPIT).* Molecular & Cellular Proteomics, 2006. **5**(1): p. 53-56.
44. Gonzalez-Begne, M., et al., *Proteomic Analysis of Human Parotid Gland Exosomes by Multidimensional Protein Identification Technology (MudPIT).* Journal of Proteome Research, 2009. **8**(3): p. 1304-1314.
45. Ziegler, Y.S., et al., *Plasma Membrane Proteomics of Human Breast Cancer Cell Lines Identifies Potential Targets for Breast Cancer Diagnosis and Treatment.* PLoS ONE, 2014. **9**(7): p. e102341.
46. Yang, F., et al., *High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis.* Expert Review of Proteomics, 2012. **9**(2): p. 129-134.
47. Siu, S.O., et al., *Fully automatable two-dimensional reversed-phase capillary liquid chromatography with online tandem mass spectrometry for shotgun proteomics.* PROTEOMICS, 2011. **11**(11): p. 2308-2319.
48. Jung, K., W. Cho, and F.E. Regnier, *Glycoproteomics of Plasma Based on Narrow Selectivity Lectin Affinity Chromatography.* Journal of Proteome Research, 2008. **8**(2): p. 643-650.
49. Durham, M. and F.E. Regnier, *Targeted glycoproteomics: Serial lectin affinity chromatography in the selection of O-glycosylation sites on proteins from the human blood proteome.* Journal of Chromatography A, 2006. **1132**(1-2): p. 165-173.
50. Yang, Z. and W.S. Hancock, *Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column.* Journal of Chromatography A, 2004. **1053**(1-2): p. 79-88.
51. Adjo Aka, J. and S.-X. Lin, *Comparison of Functional Proteomic Analyses of Human Breast Cancer Cell Lines T47D and MCF7.* PLoS One, 2012. **7**(2): p. e31532.
52. BRAUN, M., et al., *Down-regulation of Microfilamental Network-associated Proteins in Leukocytes of Breast Cancer Patients: Potential Application to Predictive Diagnosis.* Cancer Genomics - Proteomics, 2009. **6**(1): p. 31-40.
53. Lai, T.-C., et al., *Secretomic and Proteomic Analysis of Potential Breast Cancer Markers by Two-Dimensional Differential Gel Electrophoresis.* Journal of Proteome Research, 2010. **9**(3): p. 1302-1322.
54. *The Nobel Prize in Chemistry 2002*. [cited 2014 August 19]; Available from: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/.
55. Liyanage, R. and J.O. Lay, *An Introduction to MALDI-TOF MS*, in *Identification of Microorganisms by Mass Spectrometry*. 2006, John Wiley & Sons, Inc. p. 39-60.
56. Ho, C.S., et al., *Electrospray ionisation mass spectrometry: principles and clinical applications.* Clin Biochem Rev, 2003. **24**(1): p. 3-12.
57. Banerjee, S. and S. Mazumdar, *Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte.* International Journal of Analytical Chemistry, 2012. **2012**: p. 40.
58. Lucio, V., et al., *Diagnostics Methods in Ocular Infections−From Microorganism Culture to Molecular Methods, Common Eye Infections*. 2013.
59. Zhu, W., J.W. Smith, and C.-M. Huang, *Mass Spectrometry-Based Label-Free Quantitative Proteomics.* Journal of Biomedicine and Biotechnology, 2010. **2010**: p. 6.
60. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
61. Xiao, Z. and T.D. Veenstra, *Comparison of protein expression by isotope-coded affinity tag labeling.* Methods Mol Biol, 2008. **428**: p. 181-92.
62. Ye, X., et al., *18O Stable Isotope Labeling in MS-based Proteomics.* Briefings in Functional Genomics & Proteomics, 2009. **8**(2): p. 136-144.
63. Ross, P.L., et al., *Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents.* Molecular & Cellular Proteomics, 2004. **3**(12): p. 1154-1169.

64. Thompson, A., et al., *Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS*. Analytical Chemistry, 2003. **75**(8): p. 1895-1904.
65. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue*. Nat Meth, 2010. **7**(5): p. 383-385.
66. Tonack, S., et al., *iTRAQ reveals candidate pancreatic cancer serum biomarkers: influence of obstructive jaundice on their performance*. Br J Cancer, 2013. **108**(9): p. 1846-1853.
67. Rehman, I., et al., *iTRAQ Identification of Candidate Serum Biomarkers Associated with Metastatic Progression of Human Prostate Cancer*. PLoS ONE, 2012. **7**(2): p. e30885.
68. Nie, S., et al., *Glycoprotein Biomarker Panel for Pancreatic Cancer Discovered by Quantitative Proteomics Analysis*. Journal of Proteome Research, 2014. **13**(4): p. 1873-1884.
69. Savitski, M.M., et al., *Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification*. Journal of Proteome Research, 2013. **12**(8): p. 3586-3598.
70. Ow, S.Y., et al., *Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation*. Proteomics, 2011. **11**(11): p. 2341-6.
71. Wenger, C.D., et al., *Gas-phase purification enables accurate, large-scale, multiplexed proteome quantification with isobaric tagging*. Nature Methods, 2011. **8**(11): p. 933-935.
72. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics*. Anal Chem, 2004. **76**(14): p. 4193-201.
73. Bondarenko, P.V., D. Chelius, and T.A. Shaler, *Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-Phase Liquid Chromatography−Tandem Mass Spectrometry*. Analytical Chemistry, 2002. **74**(18): p. 4741-4749.
74. Chelius, D. and P.V. Bondarenko, *Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry*. Journal of Proteome Research, 2002. **1**(4): p. 317-323.
75. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics*. Mol Cell Proteomics, 2005. **4**(10): p. 1487-502.
76. Zybailov, B., et al., *Correlation of Relative Abundance Ratios Derived from Peptide Ion Chromatograms and Spectrum Counting for Quantitative Proteomic Analysis Using Stable Isotope Labeling*. Analytical Chemistry, 2005. **77**(19): p. 6218-6224.
77. Lai, X., L. Wang, and F.A. Witzmann, *Issues and Applications in Label-Free Quantitative Mass Spectrometry*. International Journal of Proteomics, 2013. **2013**: p. 13.
78. Zybailov, B., et al., *Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces cerevisiae*. Journal of Proteome Research, 2006. **5**(9): p. 2339-2347.
79. Callister, S.J., et al., *Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics*. Journal of Proteome Research, 2006. **5**(2): p. 277-286.
80. Patel, V.J., et al., *A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches*. Journal of Proteome Research, 2009. **8**(7): p. 3752-3759.
81. Bairoch, A. and R. Apweiler, *The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000*. Nucleic Acids Res, 2000. **28**(1): p. 45-8.
82. Lam, H., et al., *Development and validation of a spectral library searching method for peptide identification from MS/MS*. Proteomics, 2007. **7**(5): p. 655-67.
83. Elias, J.E., et al., *Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations*. Nat Meth, 2005. **2**(9): p. 667-675.
84. Duncan, M.W., R. Aebersold, and R.M. Caprioli, *The pros and cons of peptide-centric proteomics*. Nat Biotech, 2010. **28**(7): p. 659-664.
85. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem*. Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.

86. Eng, J., A. McCormack, and J. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.
87. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
88. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-7.
89. Ma, K., O. Vitek, and A. Nesvizhskii, *A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet.* BMC Bioinformatics, 2012. **13**(Suppl 16): p. S1.
90. Wenger, C.D. and J.J. Coon, *A proteomics search algorithm specifically designed for high-resolution tandem mass spectra.* J Proteome Res, 2013. **12**(3): p. 1377-86.
91. Zhang, B., et al., *DeMix Workflow for Efficient Identification of Co-fragmented Peptides in High Resolution Data-dependent Tandem Mass Spectrometry.* Mol Cell Proteomics, 2014.
92. Searle, B.C., *Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies.* Proteomics, 2010. **10**(6): p. 1265-9.
93. Bern, M., Y.J. Kil, and C. Becker, *Byonic: advanced peptide and protein identification software.* Curr Protoc Bioinformatics, 2012. **Chapter 13**: p. Unit13.20.
94. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-968.
95. Deutsch, E.W., et al., *A guided tour of the Trans-Proteomic Pipeline.* Proteomics, 2010. **10**(6): p. 1150-9.
96. Cox, J., et al., *A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.* Nat Protoc, 2009. **4**(5): p. 698-705.
97. Linge, A., et al., *Identification and Functional Validation of RAD23B as a Potential Protein in Human Breast Cancer Progression.* Journal of Proteome Research, 2014. **13**(7): p. 3212-3222.
98. Geiger, T., et al., *Proteomic Portrait of Human Breast Cancer Progression Identifies Novel Prognostic Markers.* Cancer Research, 2012. **72**(9): p. 2428-2439.
99. Leth-Larsen, R., et al., *Metastasis-related Plasma Membrane Proteins of Human Breast Cancer Cells Identified by Comparative Quantitative Mass Spectrometry.* Molecular & Cellular Proteomics, 2009. **8**(6): p. 1436-1449.
100. Klinke, D.J., et al., *Inferring alterations in cell-to-cell communication in HER2+ breast cancer using secretome profiling of three cell models.* Biotechnology and Bioengineering, 2014: p. n/a-n/a.
101. Pavlou, M.P., A. Dimitromanolakis, and E.P. Diamandis, *Coupling proteomics and transcriptomics in the quest of subtype-specific proteins in breast cancer.* Proteomics, 2013. **13**(7): p. 1083-95.
102. Swa, H.L.F., et al., *Mass spectrometry-based quantitative proteomics and integrative network analysis accentuates modulating roles of Annexin-1 in mammary tumorigenesis.* PROTEOMICS, 2014: p. n/a-n/a.
103. Warmoes, M., et al., *Proteomics of Mouse BRCA1-deficient Mammary Tumors Identifies DNA Repair Proteins with Potential Diagnostic and Prognostic Value in Human Breast Cancer.* Molecular & Cellular Proteomics, 2012. **11**(7).
104. Xu, B.J., et al., *Microdialysis combined with proteomics for protein identification in breast tumor microenvironment in vivo.* Cancer Microenviron, 2010. **4**(1): p. 61-71.
105. Pitteri, S.J., et al., *Tumor Microenvironment–Derived Proteins Dominate the Plasma Proteome Response during Breast Cancer Induction and Progression.* Cancer Research, 2011. **71**(15): p. 5090-5100.
106. Choong, L.Y., et al., *Proteome-Wide Profiling of the MCF10AT Breast Cancer Progression Model.* PLoS ONE, 2010. **5**(6): p. e11030.
107. Kim, S.H., et al., *Proteomic and phosphoproteomic alterations in benign, premalignant and tumor human breast epithelial cells and xenograft lesions: biomarkers of progression.* Int J Cancer, 2009. **124**(12): p. 2813-28.

108. Orazine, C.I., et al., *A proteomic analysis of the plasma glycoproteins of a MCF-7 mouse xenograft: a model system for the detection of tumor markers.* J Proteome Res, 2008. **7**(4): p. 1542-54.
109. Chung, L., et al., *Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer.* Breast Cancer Research, 2014. **16**(3): p. R63.
110. Opstal-van Winden, A., et al., *Searching for early breast cancer biomarkers by serum protein profiling of pre-diagnostic serum; a nested case-control study.* BMC Cancer, 2011. **11**(1): p. 381.
111. Kang, U.-B., et al., *Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker.* BMC Cancer, 2010. **10**(1): p. 114.
112. Zeidan, B.A., et al., *Proteomic analysis of archival breast cancer serum.* Cancer Genomics Proteomics, 2009. **6**(3): p. 141-7.
113. Schaub, N.P., et al., *Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients.* J Am Coll Surg, 2009. **208**(5): p. 970-8; discussion 978-80.
114. Pietrowska, M., et al., *Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer.* J Transl Med, 2009. **7**: p. 60.
115. Hu, X., et al., *Comparative serum proteome analysis of human lymph node negative/positive invasive ductal carcinoma of the breast and benign breast disease controls via label-free semiquantitative shotgun technology.* Omics, 2009. **13**(4): p. 291-300.
116. Kadowaki, M., et al., *Identification of vitronectin as a novel serum marker for early breast cancer detection using a new proteomic approach.* J Cancer Res Clin Oncol, 2011. **137**(7): p. 1105-15.
117. Liu, N.Q., et al., *Comparative Proteome Analysis Revealing an 11-Protein Signature for Aggressive Triple-Negative Breast Cancer.* Journal of the National Cancer Institute, 2014. **106**(2).
118. Cabezón, T., et al., *Proteomic Profiling of Triple-negative Breast Carcinomas in Combination With a Three-tier Orthogonal Technology Approach Identifies Mage-A4 as Potential Therapeutic Target in Estrogen Receptor Negative Breast Cancer.* Molecular & Cellular Proteomics, 2013. **12**(2): p. 381-394.
119. Rower, C., et al., *Towards a proteome signature for invasive ductal breast carcinoma derived from label-free nanoscale LC-MS protein expression profiling of tumorous and glandular tissue.* Anal Bioanal Chem, 2009. **395**(8): p. 2443-56.
120. Gromov, P., et al., *Up-regulated proteins in the fluid bathing the tumour cell microenvironment as potential serological markers for early detection of cancer of the breast.* Mol Oncol, 2010. **4**(1): p. 65-89.
121. Cabezon, T., et al., *Expression of S100A4 by a variety of cell types present in the tumor microenvironment of human breast cancer.* Int J Cancer, 2007. **121**(7): p. 1433-44.
122. Pavlou, M.P., et al., *Nipple aspirate fluid proteome of healthy females and patients with breast cancer.* Clin Chem, 2010. **56**(5): p. 848-55.
123. Li, J., et al., *A Targeted Proteomics Approach for Biomarker Discovery Using Bilateral Matched Nipple Aspiration Fluids.* Clinical Proteomics, 2010. **6**(3): p. 57-64.
124. Sauter, E.R., et al., *Identification of a beta-casein-like peptide in breast nipple aspirate fluid that is associated with breast cancer.* Biomark Med, 2009. **3**(5): p. 577-88.
125. Rompp, A., et al., *Identification of leptomeningeal metastasis-related proteins in cerebrospinal fluid of patients with breast cancer by a combination of MALDI-TOF, MALDI-FTICR and nanoLC-FTICR MS.* Proteomics, 2007. **7**(3): p. 474-81.
126. Streckfus, C.F., et al., *Salivary Protein Profiles among HER2/neu-Receptor-Positive and -Negative Breast Cancer Patients: Support for Using Salivary Protein Profiles for Modeling Breast Cancer Progression.* Journal of Oncology, 2012. **2012**: p. 9.
127. Bohm, D., et al., *Comparison of tear protein levels in breast cancer patients and healthy controls using a de novo proteomic approach.* Oncol Rep, 2012. **28**(2): p. 429-38.
128. Hondermarck, H., *Breast cancer: when proteomics challenges biological complexity.* Mol Cell Proteomics, 2003. **2**(5): p. 281-91.
129. Neve, R.M., et al., *A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.* Cancer Cell, 2006. **10**(6): p. 515-527.

130. Sarvaiya, H.A., J.H. Yoon, and I.M. Lazar, *Proteome profile of the MCF7 cancer cell line: a mass spectrometric evaluation.* Rapid Communications in Mass Spectrometry, 2006. **20**(20): p. 3039-3055.
131. Wu, S.L., et al., *An approach to the proteomic analysis of a breast cancer cell line (SKBR-3).* Proteomics, 2003. **3**(6): p. 1037-46.
132. Strande, V., et al., *The proteome of the human breast cancer cell line MDA-MB-231: Analysis by LTQ-Orbitrap mass spectrometry.* PROTEOMICS – Clinical Applications, 2009. **3**(1): p. 41-50.
133. Kao, J., et al., *Molecular Profiling of Breast Cancer Cell Lines Defines Relevant Tumor Models and Provides a Resource for Cancer Gene Discovery.* PLoS One, 2009. **4**(7): p. e6146.
134. Prat, A., et al., *Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes.* Breast Cancer Res Treat, 2013. **142**(2): p. 237-55.
135. Burdall, S., et al., *Breast cancer cell lines: friend or foe?* Breast Cancer Res, 2003. **5**(2): p. 1-7.
136. Stastny, J., R. Prasad, and E. Fosslien, *Tissue proteins in breast cancer, as studied by use of two-dimensional electrophoresis.* Clinical Chemistry, 1984. **30**(12): p. 1914-1918.
137. Blanco, M.A., et al., *Global secretome analysis identifies novel mediators of bone metastasis.* Cell Res, 2012. **22**(9): p. 1339-55.
138. Choi, D.-S., et al., *Proteomics of extracellular vesicles: Exosomes and ectosomes.* Mass Spectrometry Reviews, 2014: p. n/a-n/a.
139. Gonzales, P.A., et al., *Large-Scale Proteomics and Phosphoproteomics of Urinary Exosomes.* Journal of the American Society of Nephrology, 2009. **20**(2): p. 363-379.
140. Street, J., et al., *Identification and proteomic profiling of exosomes in human cerebrospinal fluid.* Journal of Translational Medicine, 2012. **10**(1): p. 5.
141. Kruger, S., et al., *Molecular characterization of exosome-like vesicles from breast cancer cells.* BMC Cancer, 2014. **14**(1): p. 44.
142. Chang, J.W., et al., *Identification of circulating endorepellin LG3 fragment: Potential use as a serological biomarker for breast cancer.* Proteomics Clin Appl, 2008. **2**(1): p. 23-32.
143. Moreira, J.M., et al., *Down-regulation of the tumor suppressor protein 14-3-3sigma is a sporadic event in cancer of the breast.* Mol Cell Proteomics, 2005. **4**(4): p. 555-69.
144. Hondermarck, H., et al., *Proteomics of breast cancer: the quest for markers and therapeutic targets.* J Proteome Res, 2008. **7**(4): p. 1403-11.
145. Mitchell, P., *Proteomics retrenches.* Nat Biotech, 2010. **28**(7): p. 665-670.
146. Hortin, G.L. and D. Sviridov, *The dynamic range problem in the analysis of the plasma proteome.* Journal of Proteomics, 2010. **73**(3): p. 629-636.
147. Omenn, G.S., *Exploring the Human Plasma Proteome.* PROTEOMICS, 2005. **5**(13): p. 3223-3225.
148. Diamandis, E.P., *Analysis of Serum Proteomic Patterns for Early Cancer Diagnosis: Drawing Attention to Potential Problems.* Journal of the National Cancer Institute, 2004. **96**(5): p. 353-356.
149. Zhou, M., et al., *An investigation into the human serum "interactome".* ELECTROPHORESIS, 2004. **25**(9): p. 1289-1298.
150. Tuck, M.K., et al., *Standard Operating Procedures for Serum and Plasma Collection: Early Detection Research Network Consensus Statement Standard Operating Procedure Integration Working Group.* Journal of Proteome Research, 2008. **8**(1): p. 113-117.
151. Farrah, T., et al., *A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas.* Mol Cell Proteomics, 2011. **10**(9): p. M110.006353.
152. Nanjappa, V., et al., *Plasma Proteome Database as a resource for proteomics research: 2014 update.* Nucleic Acids Research, 2014. **42**(D1): p. D959-D965.
153. Carriero, M.V., et al., *Vitronectin binding to urokinase receptor in human breast cancer.* Clin Cancer Res, 1997. **3**(8): p. 1299-308.

154. Aaboe, M., et al., *Vitronectin in human breast carcinomas.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2003. **1638**(1): p. 72-82.
155. Zhou, A., et al., *How vitronectin binds PAI-1 to modulate fibrinolysis and cell migration.* Nat Struct Biol, 2003. **10**(7): p. 541-4.
156. Hornberg, J.J., et al., *Cancer: a Systems Biology disease.* Biosystems, 2006. **83**(2-3): p. 81-90.
157. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Res, 2004. **32**(Database issue): p. D115-9.
158. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nat Genet, 2000. **25**(1): p. 25-29.
159. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
160. Jupe, S., et al., *A controlled vocabulary for pathway entities and events.* Database, 2014. **2014**.
161. Khatri, P., M. Sirota, and A.J. Butte, *Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges.* PLoS Comput Biol, 2012. **8**(2): p. e1002375.
162. Franceschini, A., et al., *STRING v9.1: protein-protein interaction networks, with increased coverage and integration.* Nucleic Acids Res, 2013. **41**(Database issue): p. D808-15.
163. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
164. Mi, H., A. Muruganujan, and P.D. Thomas, *PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees.* Nucleic Acids Res, 2013. **41**(Database issue): p. D377-86.
165. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nat. Protocols, 2007. **2**(10): p. 2366-2382.
166. Hu, Z.-Z., et al., *Proteomic Analysis of Pathways Involved in Estrogen-Induced Growth and Apoptosis of Breast Cancer Cells.* PLoS ONE, 2011. **6**(6): p. e20410.
167. Wu, G., X. Feng, and L. Stein, *A human functional protein interaction network and its application to cancer data analysis.* Genome Biology, 2010. **11**(5): p. R53.
168. Gujral, T.S. and G. MacBeath, *A System-Wide Investigation of the Dynamics of Wnt Signaling Reveals Novel Phases of Transcriptional Regulation.* PLoS ONE, 2010. **5**(4): p. e10024.
169. Lochter, A. and M.J. Bissell, *Involvement of extracellular matrix constituents in breast cancer.* Seminars in Cancer Biology, 1995. **6**(3): p. 165-173.
170. Yurchenco, P.D., *Basement membranes: cell scaffoldings and signaling platforms.* Cold Spring Harb Perspect Biol, 2011. **3**(2).
171. Daley, W.P., et al., *ROCK1-directed basement membrane positioning coordinates epithelial tissue polarity.* Development, 2012. **139**(2): p. 411-422.
172. Bhowmick, N.A., E.G. Neilson, and H.L. Moses, *Stromal fibroblasts in cancer initiation and progression.* Nature, 2004. **432**(7015): p. 332-337.
173. Martin, K.J., et al., *Down-regulation of laminin-5 in breast carcinoma cells.* Mol Med, 1998. **4**(9): p. 602-13.
174. Akhavan, A., et al., *Loss of Cell-Surface Laminin Anchoring Promotes Tumor Growth and Is Associated with Poor Clinical Outcomes.* Cancer Research, 2012. **72**(10): p. 2578-2588.
175. Nakano, S., et al., *Differential tissular expression and localization of type IV collagen alpha1(IV), alpha2(IV), alpha5(IV), and alpha6(IV) chains and their m*RNA *in normal breast and in benign and malignant breast tumors.* Lab Invest, 1999. **79**(3): p. 281-92.
176. Xu, D., et al., *Matrix metalloproteinase-9 regulates tumor cell invasion through cleavage of protease nexin-1.* Cancer Res, 2010. **70**(17): p. 6988-98.
177. Benson, C.S., et al., *Expression of Matrix Metalloproteinases in Human Breast Cancer Tissues.* Disease Markers, 2013. **34**(6).
178. Lamouille, S., J. Xu, and R. Derynck, *Molecular mechanisms of epithelial–mesenchymal transition.* Nat Rev Mol Cell Biol, 2014. **15**(3): p. 178-196.

179. Hazan, R.B., et al., *Cadherin switch in tumor progression.* Ann N Y Acad Sci, 2004. **1014**: p. 155-63.
180. Rimm, D.L., J.H. Sinard, and J.S. Morrow, *Reduced alpha-catenin and E-cadherin expression in breast cancer.* Lab Invest, 1995. **72**(5): p. 506-12.
181. Nakopoulou, L., et al., *Abnormal a-catenin expression in invasive breast cancer correlates with poor patient survival.* Histopathology, 2002. **40**(6): p. 536-546.
182. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* The Journal of Clinical Investigation, 2009. **119**(6): p. 1420-1428.
183. Spiro, R.G., *Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds.* Glycobiology, 2002. **12**(4): p. 43R-56R.
184. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 448-462.
185. Bause, E., *Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes.* Biochem. J., 1983. **209**(2): p. 331-336.
186. Elliott, S., et al., *Structural Requirements for Additional N-Linked Carbohydrate on Recombinant Human Erythropoietin.* Journal of Biological Chemistry, 2004. **279**(16): p. 16854-16862.
187. Bano-Polo, M., et al., *N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon.* Protein Sci, 2011. **20**(1): p. 179-86.
188. Strous, G.J. and J. Dekker, *Mucin-type glycoproteins.* Crit Rev Biochem Mol Biol, 1992. **27**(1-2): p. 57-92.
189. Stanley, P., H. Schachter, and N. Taniguchi, *N-Glycans*, in *Essentials of Glycobiology*, A. Varki, R. Cummings, and J. Esko, Editors. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
190. Schachter, H., *Paucimannose N-glycans in Caenorhabditis elegans and Drosophila melanogaster.* Carbohydrate Research, 2009. **344**(12): p. 1391-1396.
191. Samyn-Petit, B., et al., *Comparative analysis of the site-specific N-glycosylation of human lactoferrin produced in maize and tobacco plants.* European Journal of Biochemistry, 2003. **270**(15): p. 3235-3242.
192. Zipser, B., et al., *Mannitou Monoclonal Antibody Uniquely Recognizes Paucimannose, a Marker for Human Cancer, Stemness, and Inflammation.* Journal of Carbohydrate Chemistry, 2012. **31**(4-6): p. 504-518.
193. Owens, R.J. and J.E. Nettleship, *Functional and Structural Proteomics of Glycoproteins*. 2010: Springer.
194. Varki, A., et al., *Essentials of Glycobiology*. null. Vol. null. 2008. 784.
195. Sanyal, S. and A.K. Menon, *Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase.* Proceedings of the National Academy of Sciences, 2010. **107**(25): p. 11289-11294.
196. Sakaguchi, M., et al., *Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge.* Proc Natl Acad Sci U S A, 1992. **89**(1): p. 16-9.
197. Sumer-Bayraktar, Z., et al., *Micro- and macroheterogeneity of N-glycosylation yields size and charge isoforms of human sex hormone binding globulin circulating in serum.* Proteomics, 2012. **12**(22): p. 3315-27.
198. Thaysen-Andersen, M. and N.H. Packer, *Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching.* Glycobiology, 2012. **22**(11): p. 1440-1452.
199. Parekh, R.B., et al., *Cell-type-specific and site-specific N-glycosylation of type I and type II human tissue plasminogen activator.* Biochemistry, 1989. **28**(19): p. 7644-62.
200. Lee, L.Y., et al., *Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants.* Frontiers in Immunology, 2014. **5**.

201. Christiansen, M.N., et al., *Cell surface protein glycosylation in cancer.* Proteomics, 2013: p. n/a-n/a.
202. Eklund, E. and H. Freeze, *The congenital disorders of glycosylation: A multifaceted group of syndromes.* NeuroRX, 2006. **3**(2): p. 254-263.
203. Rabinovich, G.A., Y. van Kooyk, and B.A. Cobb, *Glycobiology of immune responses.* Ann N Y Acad Sci, 2012. **1253**: p. 1-15.
204. Carlsson, J., et al., *Detection of global glycosylation changes of serum proteins in type 1 diabetes using a lectin panel and multivariate data analysis.* Talanta, 2008. **76**(2): p. 333-7.
205. Hwang, H., et al., *Glycoproteomics in neurodegenerative diseases.* Mass Spectrometry Reviews, 2010. **29**(1): p. 79-125.
206. Turnbull, J.E. and R.A. Field, *Emerging glycomics technologies.* Nat Chem Biol, 2007. **3**(2): p. 74-77.
207. Rakus, J.F. and L.K. Mahal, *New Technologies for Glycomic Analysis: Toward a Systematic Understanding of the Glycome.* Annual Review of Analytical Chemistry, 2011. **4**(1): p. 367-392.
208. Takahashi, N. and H. Nishibe, *Some characteristics of a new glycopeptidase acting on aspartylglycosylamine linkages.* J Biochem, 1978. **84**(6): p. 1467-73.
209. Jensen, P.H., et al., *Structural analysis of N- and O-glycans released from glycoproteins.* Nat. Protocols, 2012. **7**(7): p. 1299-1310.
210. Adamczyk, B., T. Tharmalingam, and P.M. Rudd, *Glycans as cancer biomarkers.* Biochim Biophys Acta, 2012. **1820**(9): p. 1347-53.
211. Pierce, A., et al., *Levels of specific glycans significantly distinguish lymph node-positive from lymph node-negative breast cancer patients.* Glycobiology, 2010. **20**(10): p. 1283-1288.
212. Abbott, K.L., et al., *Targeted Glycoproteomic Identification of Biomarkers for Human Breast Carcinoma.* Journal of Proteome Research, 2008. **7**(4): p. 1470-1480.
213. North, S.J., et al., *Mass spectrometry in the analysis of N-linked and O-linked glycans.* Curr Opin Struct Biol, 2009. **19**(5): p. 498-506.
214. Wuhrer, M., et al., *Nano-scale liquid chromatography-mass spectrometry of 2-aminobenzamide-labeled oligosaccharides at low femtomole sensitivity.* International Journal of Mass Spectrometry, 2004. **232**(1): p. 51-57.
215. Ruhaak, L.R., et al., *Glycan labeling strategies and their use in identification and quantification.* Analytical and Bioanalytical Chemistry, 2010. **397**(8): p. 3457-3481.
216. Edge, C.J., et al., *Fast sequencing of oligosaccharides: the reagent-array analysis method.* Proceedings of the National Academy of Sciences, 1992. **89**(14): p. 6338-6342.
217. Campbell, M.P., et al., *GlycoBase and autoGU: tools for HPLC-based glycan analysis.* Bioinformatics, 2008. **24**(9): p. 1214-1216.
218. Desantos-Garcia, J.L., et al., *Enhanced sensitivity of LC-MS analysis of permethylated N-glycans through online purification.* Electrophoresis, 2011. **32**(24): p. 3516-25.
219. Jia, N., et al., *Glycomic characterisation of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies.* Journal of Biological Chemistry, 2014.
220. Kyselova, Z., et al., *Breast Cancer Diagnosis and Prognosis through Quantitative Measurements of Serum Glycan Profiles.* Clinical Chemistry, 2008. **54**(7): p. 1166-1175.
221. Alley, W.R., et al., *Chip-based Reversed-phase Liquid Chromatography−Mass Spectrometry of Permethylated N-Linked Glycans: A Potential Methodology for Cancer-biomarker Discovery.* Analytical Chemistry, 2010. **82**(12): p. 5095-5106.
222. Inamoto, Y., et al., *Liquid chromatography of guanidino compounds using a porous graphite carbon column and application to their analysis in serum.* J Chromatogr B Biomed Sci Appl, 1998. **707**(1-2): p. 111-20.
223. Melmer, M., et al., *Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis.* J Chromatogr A, 2011. **1218**(1): p. 118-23.

224. Ruhaak, L.R., A.M. Deelder, and M. Wuhrer, *Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry.* Anal Bioanal Chem, 2009. **394**(1): p. 163-74.
225. Everest-Dass, A., et al., *Structural Feature Ions for Distinguishing N- and O-Linked Glycan Isomers by LC-ESI-IT MS/MS*. Journal of The American Society for Mass Spectrometry, 2013. **24**(6): p. 895-906.
226. Ruhaak, L.R., A. Deelder, and M. Wuhrer, *Oligosaccharide analysis by graphitized carbon liquid chromatography–mass spectrometry.* Analytical and Bioanalytical Chemistry, 2009. **394**(1): p. 163-174.
227. Everest-Dass, A.V., et al., *Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by C. albicans.* Glycobiology, 2012.
228. Anugraham, M., et al., *Specific glycosylation of membrane proteins in epithelial ovarian cancer cell lines: glycan structures reflect gene expression and DNA methylation status.* Molecular & Cellular Proteomics, 2014.
229. Chik, J.H.L., et al., *Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies.* Journal of Proteomics, 2014. **108**(0): p. 146-162.
230. Wongtrakul-Kish, K., et al., *Characterization of N- and O-linked glycosylation changes in milk of the tammar wallaby (Macropus eugenii) over lactation.* Glycoconjugate Journal, 2013. **30**(5): p. 523-536.
231. Lee, A., et al., *Liver Membrane Proteome Glycosylation Changes in Mice Bearing an Extra-hepatic Tumor.* Molecular & Cellular Proteomics, 2011. **10**(9).
232. Lee, L.Y., et al., *Comprehensive N-Glycome Profiling of Cultured Human Epithelial Breast Cells Identifies Unique Secretome N-Glycosylation Signatures Enabling Tumorigenic Subtype Classification.* J Proteome Res, 2014.
233. Raman, R., et al., *Advancing glycomics: implementation strategies at the consortium for functional glycomics.* Glycobiology, 2006. **16**(5): p. 82r-90r.
234. Campbell, M.P., et al., *UniCarbKB: building a knowledge platform for glycoproteomics.* Nucleic Acids Research, 2014. **42**(D1): p. D215-D221.
235. Cooper, C.A., E. Gasteiger, and N.H. Packer, *GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data.* Proteomics, 2001. **1**(2): p. 340-9.
236. Cooper, C.A., et al., *GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources.* Nucleic Acids Research, 2001. **29**(1): p. 332-335.
237. Hashimoto, K., et al., *KEGG as a glycome informatics resource.* Glycobiology, 2006. **16**(5): p. 63r-70r.
238. Ceroni, A., et al., *GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans.* J Proteome Res, 2008. **7**(4): p. 1650-9.
239. Sakamoto, J., et al., *Expression of Lewisa, Lewisb, X, and Y Blood Group Antigens in Human Colonic Tumors and Normal Tissue and in Human Tumor-derived Cell Lines.* Cancer Research, 1986. **46**(3): p. 1553-1561.
240. Dennis, J., et al., *Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis.* Science, 1987. **236**(4801): p. 582-585.
241. Handerson, T., et al., *β1,6-Branched Oligosaccharides Are Increased in Lymph Node Metastases and Predict Poor Outcome in Breast Carcinoma.* Clinical Cancer Research, 2005. **11**(8): p. 2969-2973.
242. Fernandes, B., et al., *β1–6 Branched Oligosaccharides as a Marker of Tumor Progression in Human Breast and Colon Neoplasia.* Cancer Research, 1991. **51**(2): p. 718-723.
243. Drake, P.M., et al., *Lectin Chromatography/Mass Spectrometry Discovery Workflow Identifies Putative Biomarkers of Aggressive Breast Cancers.* Journal of Proteome Research, 2012. **11**(4): p. 2508-2520.
244. Brooks, S.A., D.M.S. Hall, and I. Buley, *GalNAc glycoprotein expression by breast cell lines, primary breast cancer and normal breast epithelial membrane.* Br J Cancer, 2001. **85**(7): p. 1014-1022.

245. Chen, S., et al., *Analysis of Cell Surface Carbohydrate Expression Patterns in Normal and Tumorigenic Human Breast Cell Lines Using Lectin Arrays.* Analytical Chemistry, 2007. **79**(15): p. 5698-5702.
246. Madjd, Z., et al., *High expression of Lewisy/b antigens is associated with decreased survival in lymph node negative breast carcinomas.* Breast Cancer Research, 2005. **7**(5): p. R780 - R787.
247. Potapenko, I.O., et al., *Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression.* Mol Oncol, 2010. **4**(2): p. 98-118.
248. Carcel-Trullols, J., et al., *Characterization of the glycosylation profile of the human breast cancer cell line, MDA-231, and a bone colonizing variant.* Int J Oncol, 2006. **28**(5): p. 1173-83.
249. Julien, S., et al., *Selectin Ligand Sialyl-Lewis x Antigen Drives Metastasis of Hormone-Dependent Breast Cancers.* Cancer Research, 2011. **71**(24): p. 7683-7693.
250. Inoue, S. and K. Kitajima, *KDN (Deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family.* Glycoconjugate Journal, 2006. **23**(5-6): p. 277-290.
251. Varki, A., *Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution.* Am J Phys Anthropol, 2001. **Suppl 33**: p. 54-69.
252. Samraj, A., et al., *Involvement of a Non-Human Sialic Acid in Human Cancer.* Frontiers in Oncology, 2014. **4**.
253. Cui, H., et al., *Differential expression of the alpha2,3-sialic acid residues in breast cancer is associated with metastatic potential.* Oncol Rep, 2011. **25**(5): p. 1365-71.
254. Lin, S., et al., *Cell Surface [alpha]2,6-Sialylation Affects Adhesion of Breast Carcinoma Cells.* Experimental Cell Research, 2002. **276**(1): p. 101-110.
255. Alley, W.R., Jr. and M.V. Novotny, *Glycomic analysis of sialic acid linkages in glycans derived from blood serum glycoproteins.* J Proteome Res, 2010. **9**(6): p. 3062-72.
256. Narita, T., et al., *Association of expression of blood group-related carbohydrate antigens with prognosis in breast cancer.* Cancer, 1993. **71**(10): p. 3044-53.
257. Renkonen, J., T. Paavonen, and R. Renkonen, *Endothelial and epithelial expression of sialyl Lewisx and sialyl Lewisa in lesions of breast carcinoma.* International Journal of Cancer, 1997. **74**(3): p. 296-300.
258. Nakagoe, T., et al., *Expression of ABH/Lewis-related antigens as prognostic factors in patients with breast cancer.* J Cancer Res Clin Oncol, 2002. **128**(5): p. 257-64.
259. Saldova, R., et al., *Levels of specific serum N-glycans identify breast cancer patients with higher circulating tumor cell counts.* Ann Oncol, 2011. **22**(5): p. 1113-9.
260. Ronn, L.C., B.P. Hartz, and E. Bock, *The neural cell adhesion molecule (NCAM) in development and plasticity of the nervous system.* Exp Gerontol, 1998. **33**(7-8): p. 853-64.
261. Martersteck, C.M., et al., *Unique alpha 2, 8-polysialylated glycoproteins in breast cancer and leukemia cells.* Glycobiology, 1996. **6**(3): p. 289-301.
262. Raval, G.N., et al., *Clinical usefulness of alterations in sialic acid, sialyl transferase and sialoproteins in breast cancer.* Indian J Clin Biochem, 2004. **19**(2): p. 60-71.
263. Recchi, M.A., et al., *Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer.* Cancer Res, 1998. **58**(18): p. 4066-70.
264. Figdor, C.G., Y. van Kooyk, and G.J. Adema, *C-type lectin receptors on dendritic cells and Langerhans cells.* Nat Rev Immunol, 2002. **2**(2): p. 77-84.
265. Kurebayashi, J., et al., *Combined measurement of serum sialyl Lewis X with serum CA15-3 in breast cancer patients.* Jpn J Clin Oncol, 2006. **36**(3): p. 150-3.
266. Pillai, S., et al., *Siglecs and Immune Regulation.* Annual Review of Immunology, 2012. **30**(1): p. 357-392.
267. Hudak, J.E., S.M. Canham, and C.R. Bertozzi, *Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion.* Nat Chem Biol, 2014. **10**(1): p. 69-75.
268. Sabit, I., et al., *Binding of a Sialic Acid-recognizing Lectin Siglec-9 Modulates Adhesion Dynamics of Cancer Cells via Calpain-mediated Protein Degradation.* Journal of Biological Chemistry, 2013. **288**(49): p. 35417-35427.

269. Miyoshi, E., K. Moriwaki, and T. Nakagawa, *Biological Function of Fucosylation in Cancer Biology.* Journal of Biochemistry, 2008. **143**(6): p. 725-729.
270. Ferrara, C., et al., *Unique carbohydrate–carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose.* Proceedings of the National Academy of Sciences, 2011.
271. Junttila, T.T., et al., *Superior In vivo Efficacy of Afucosylated Trastuzumab in the Treatment of HER2-Amplified Breast Cancer.* Cancer Research, 2010. **70**(11): p. 4481-4489.
272. Goodarzi, M.T. and G.A. Turner, *Decreased branching, increased fucosylation and changed sialylation of alpha-1-proteinase inhibitor in breast and ovarian cancer.* Clinica Chimica Acta, 1995. **236**(2): p. 161-171.
273. Yuan, K., et al., *Cell surface associated alpha-L-fucose moieties modulate human breast cancer neoplastic progression.* Pathol Oncol Res, 2008. **14**(2): p. 145-56.
274. Elola, M.T., et al., *Lewis x antigen mediates adhesion of human breast carcinoma cells to activated endothelium. Possible involvement of the endothelial scavenger receptor C-type lectin.* Breast Cancer Res Treat, 2007. **101**(2): p. 161-74.
275. Larrain, M., et al., *Breast cancer humoral immune response: involvement of Lewis y through the detection of circulating immune complexes and association with Mucin 1 (MUC1).* Journal of Experimental & Clinical Cancer Research, 2009. **28**(1): p. 121.
276. Ali, S., et al., *Leukocyte extravasation: an immunoregulatory role for alpha-L-fucosidase?* J Immunol, 2008. **181**(4): p. 2407-13.
277. Fukuda, M., *Possible roles of tumor-associated carbohydrate antigens.* Cancer Res, 1996. **56**(10): p. 2237-44.
278. Jeschke, U., et al., *Expression of sialyl lewis X, sialyl Lewis A, E-cadherin and cathepsin-D in human breast cancer: immunohistochemical analysis in mammary carcinoma in situ, invasive carcinomas and their lymph node metastasis.* Anticancer Res, 2005. **25**(3A): p. 1615-22.
279. Takada, A., et al., *Contribution of carbohydrate antigens sialyl Lewis A and sialyl Lewis X to adhesion of human cancer cells to vascular endothelium.* Cancer Res, 1993. **53**(2): p. 354-61.
280. Matsuura, N., et al., *Gene expression of fucosyl- and sialyl-transferases which synthesize sialyl Lewis(x), the carbohydrate ligands for E-selectin, in human breast cancer.* International Journal of Oncology, 1998. **12**(5): p. 1157-1164.
281. Zen, K., et al., *CD44v4 is a major E-selectin ligand that mediates breast cancer cell transendothelial migration.* PLoS One, 2008. **3**(3): p. e1826.
282. Powlesland, A.S., et al., *Targeted glycoproteomic identification of cancer cell glycosylation.* Glycobiology, 2009. **19**(8): p. 899-909.
283. Riethdorf, S., et al., *High incidence of EMMPRIN expression in human tumors.* International Journal of Cancer, 2006. **119**(8): p. 1800-1810.
284. Nielsen, J.S. and K.M. McNagny, *Novel functions of the CD34 family.* Journal of Cell Science, 2008. **121**(22): p. 3683-3692.
285. Dennis, J.W. and S. Laferte, *Oncodevelopmental expression of -GlcNAcβ1-6Mana1-6Manβ1-branched asparagine-linked oligosaccharides in murine tissues and human breast carcinomas.* Cancer Research, 1989. **49**(4): p. 945-950.
286. Ihara, S., et al., *Prometastatic effect of N-acetylglucosaminyltransferase V is due to modification and stabilization of active matriptase by adding beta 1-6 GlcNAc branching.* J Biol Chem, 2002. **277**(19): p. 16960-7.
287. Guo, H.B., et al., *N-acetylglucosaminyltransferase V expression levels regulate cadherin-associated homotypic cell-cell adhesion and intracellular signaling pathways.* J Biol Chem, 2003. **278**(52): p. 52412-24.
288. Guo, H.B., M. Randolph, and M. Pierce, *Inhibition of a specific N-glycosylation activity results in attenuation of breast carcinoma cell invasiveness-related phenotypes: inhibition of epidermal growth factor-induced dephosphorylation of focal adhesion kinase.* J Biol Chem, 2007. **282**(30): p. 22150-62.

289. Seberger, P.J. and W.G. Chaney, *Control of metastasis by Asn-linked, β1–6 branched oligosaccharides in mouse mammary cancer cells.* Glycobiology, 1999. **9**(3): p. 235-241.
290. Machado, E., et al., *N-Glycosylation of total cellular glycoproteins from the human ovarian carcinoma SKOV3 cell line and of recombinantly expressed human erythropoietin.* Glycobiology, 2011. **21**(3): p. 376-86.
291. Hua, S., et al., *Differentiation of Cancer Cell Origin and Molecular Subtype by Plasma Membrane N-Glycan Profiling.* Journal of Proteome Research, 2013. **13**(2): p. 961-968.
292. Mun, J.-Y., et al., *Efficient Adhesion-Based Plasma Membrane Isolation for Cell Surface N-Glycan Analysis.* Analytical Chemistry, 2013. **85**(15): p. 7462-7470.
293. de Leoz, M.L.A., et al., *High-Mannose Glycans are Elevated during Breast Cancer Progression.* Molecular & Cellular Proteomics, 2011. **10**(1).
294. Johns, T.G., et al., *The antitumor monoclonal antibody 806 recognizes a high-mannose form of the EGF receptor that reaches the cell surface when cells over-express the receptor.* The FASEB Journal, 2005.
295. Rouzier, R., et al., *Breast Cancer Molecular Subtypes Respond Differently to Preoperative Chemotherapy.* Clinical Cancer Research, 2005. **11**(16): p. 5678-5685.
296. Adam, P.J., et al., *Comprehensive Proteomic Analysis of Breast Cancer Cell Membranes Reveals Unique Proteins with Potential Roles in Clinical Cancer.* Journal of Biological Chemistry, 2003. **278**(8): p. 6482-6489.
297. Stastna, M. and J.E. Van Eyk, *Secreted proteins as a fundamental source for biomarker discovery.* Proteomics, 2012. **12**(4-5): p. 722-35.
298. Bakheet, T.M. and A.J. Doig, *Properties and identification of human protein drug targets.* Bioinformatics, 2009. **25**(4): p. 451-457.
299. Lundby, A. and J.V. Olsen, *GeLCMS for in-depth protein characterization and advanced analysis of proteomes.* Methods Mol Biol, 2011. **753**: p. 143-55.
300. Wong, J.H. and G. Cagney, *An Overview of Label-Free Quantitation Methods in Proteomics by Mass Spectrometry*, in *Proteome Bioinformatics*, S.J. Hubbard and A.R. Jones, Editors. 2010, Humana Press. p. 273-283.
301. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nat Protoc, 2007. **2**(10): p. 2366-82.
302. Samali, A., et al., *Methods for Monitoring Endoplasmic Reticulum Stress and the Unfolded Protein Response.* International Journal of Cell Biology, 2010. **2010**: p. 11.
303. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nat Meth, 2011. **8**(10): p. 785-786.
304. Karagiannis, G.S., M.P. Pavlou, and E.P. Diamandis, *Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology.* Molecular Oncology, 2010. **4**(6): p. 496-510.
305. van der Pol, E., et al., *Classification, Functions, and Clinical Relevance of Extracellular Vesicles.* Pharmacological Reviews, 2012. **64**(3): p. 676-705.
306. Bendtsen, J.D., et al., *Feature-based prediction of non-classical and leaderless protein secretion.* Protein Eng Des Sel, 2004. **17**(4): p. 349-56.
307. Kalra, H., et al., *Vesiclepedia: A Compendium for Extracellular Vesicles with Continuous Community Annotation.* PLoS Biol, 2012. **10**(12): p. e1001450.
308. Trinh, H.V., et al., *iTRAQ-Based and Label-Free Proteomics Approaches for Studies of Human Adenovirus Infections.* International Journal of Proteomics, 2013. **2013**: p. 16.
309. Neilson, K.A., M. Mariani, and P.A. Haynes, *Quantitative proteomic analysis of cold-responsive proteins in rice.* Proteomics, 2011: p. n/a-n/a.
310. Fanayan, S., et al., *Proteogenomic Analysis of Human Colon Carcinoma Cell Lines LIM1215, LIM1899, and LIM2405.* J Proteome Res, 2013.
311. Szklarczyk, D., et al., *The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.* Nucleic Acids Res, 2011. **39**(Database issue): p. D561-8.

312. Villarreal, L., et al., *Unconventional secretion is a major contributor of cancer cell line secretomes.* Molecular & Cellular Proteomics, 2012.
313. Bordier, C., *Phase separation of integral membrane proteins in Triton X-114 solution.* Journal of Biological Chemistry, 1981. **256**(4): p. 1604-1607.
314. Hansson, S., et al., *Membrane Protein Profiling of Human Islets of Langerhans Using Several Extraction Methods.* Clinical Proteomics, 2010. **6**(4): p. 195-207.
315. Liang, X., et al., *Identification and quantification of proteins differentially secreted by a pair of normal and malignant breast-cancer cell lines.* Proteomics, 2009. **9**(1): p. 182-93.
316. Logozzi, M., et al., *High Levels of Exosomes Expressing CD63 and Caveolin-1 in Plasma of Melanoma Patients.* PLOS ONE, 2012.
317. Taylor, D.D. and C. Gercel-Taylor, *MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer.* Gynecol Oncol, 2008. **110**(1): p. 13-21.
318. Ogata-Kawata, H., et al., *Circulating Exosomal microRNAs as Biomarkers of Colon Cancer.* PLoS ONE, 2014. **9**(4): p. e92921.
319. Masyuk, A.I., T.V. Masyuk, and N.F. LaRusso, *Exosomes in the pathogenesis, diagnostics and therapeutics of liver diseases.* Journal of Hepatology, 2013. **59**(3): p. 621-625.
320. Marx, C., et al., *Proteasome-Regulated ERBB2 and Estrogen Receptor Pathways in Breast Cancer.* Molecular Pharmacology, 2007. **71**(6): p. 1525-1534.
321. Chen, L. and K. Madura, *Increased Proteasome Activity, Ubiquitin-Conjugating Enzymes, and eEF1A Translation Factor Detected in Breast Cancer Tissue.* Cancer Research, 2005. **65**(13): p. 5599-5606.
322. Dees, E.C. and R.Z. Orlowski, *Targeting the ubiquitin–proteasome pathway in breast cancer therapy.* Future Oncology, 2006. **2**(1): p. 121-135.
323. Agyin, J.K., et al., *BU-32: a novel proteasome inhibitor for breast cancer.* Breast Cancer Res, 2009. **11**(5): p. R74.
324. Graff, J.R., et al., *Targeting the Eukaryotic Translation Initiation Factor 4E for Cancer Therapy.* Cancer Research, 2008. **68**(3): p. 631-634.
325. Orlowski, R.Z., G.W. Small, and Y.Y. Shi, *Evidence that inhibition of p44/42 mitogen-activated protein kinase signaling is a factor in proteasome inhibitor-mediated apoptosis.* J Biol Chem, 2002. **277**(31): p. 27864-71.
326. Cheng, Y., et al., *XPO1 (CRM1) inhibition represses STAT3 activation to drive a survivin-dependent oncogenic switch in triple-negative breast cancer.* Mol Cancer Ther, 2014. **13**(3): p. 675-86.
327. Kass, L., et al., *Mammary epithelial cell: influence of extracellular matrix composition and organization during development and tumorigenesis.* Int J Biochem Cell Biol, 2007. **39**(11): p. 1987-94.
328. Willipinski-Stapelfeldt, B., et al., *Changes in Cytoskeletal Protein Composition Indicative of an Epithelial-Mesenchymal Transition in Human Micrometastatic and Primary Breast Carcinoma Cells.* Clinical Cancer Research, 2005. **11**(22): p. 8006-8014.
329. Yamaguchi, H. and J. Condeelis, *Regulation of the actin cytoskeleton in cancer cell migration and invasion.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2007. **1773**(5): p. 642-652.
330. Muschler, J. and C.H. Streuli, *Cell-matrix interactions in mammary gland development and breast cancer.* Cold Spring Harb Perspect Biol, 2010. **2**(10): p. a003202.
331. Bergstraesser, L.M., et al., *Expression of hemidesmosomes and component proteins is lost by invasive breast cancer cells.* Am J Pathol, 1995. **147**(6): p. 1823-39.
332. Zajchowski, D.A., et al., *Identification of Gene Expression Profiles That Predict the Aggressive Behavior of Breast Cancer Cells.* Cancer Research, 2001. **61**(13): p. 5168-5178.
333. Zaidel-Bar, R., et al., *Functional atlas of the integrin adhesome.* Nat Cell Biol, 2007. **9**(8): p. 858-867.

334. Sutoh Yoneyama, M., et al., *Vimentin intermediate filament and plectin provide a scaffold for invadopodia, facilitating cancer cell invasion and extravasation for metastasis.* European Journal of Cell Biology, 2014. **93**(4): p. 157-169.
335. Iyengar, R., *Introduction: Overview of Pathways and Networks and GPCR Signaling.* Vol. 2005. 2005. tr4-tr4.
336. Davey, J., *G-protein-coupled receptors: new approaches to maximise the impact of GPCRS in drug discovery.* Expert Opin Ther Targets, 2004. **8**(2): p. 165-70.
337. Komatsu, H., et al., *A <italic>GNAS</italic> Mutation Found in Pancreatic Intraductal Papillary Mucinous Neoplasms Induces Drastic Alterations of Gene Expression Profiles with Upregulation of Mucin Genes.* PLoS ONE, 2014. **9**(2): p. e87875.
338. Liu, Z., et al., *Overexpression of GNAO1 correlates with poor prognosis in patients with gastric cancer and plays a role in gastric cancer cell proliferation and apoptosis.* Int J Mol Med, 2014. **33**(3): p. 589-96.
339. Dono, M., et al., *Mutation frequencies of GNAQ, GNA11, BAP1, SF3B1, EIF1AX and TERT in uveal melanoma: detection of an activating mutation in the TERT gene promoter in a single case of uveal melanoma.* Br J Cancer, 2014. **110**(4): p. 1058-1065.
340. Garcia-Murillas, I., et al., *An siRNA screen identifies the GNAS locus as a driver in 20q amplified breast cancer.* Oncogene, 2014. **33**(19): p. 2478-86.
341. Bull, C., et al., *Sialic Acids Sweeten a Tumor's Life.* Cancer Res, 2014.
342. Isaji, T., et al., *Introduction of Bisecting GlcNAc into Integrin α5β1 Reduces Ligand Binding and Down-regulates Cell Adhesion and Cell Migration.* Journal of Biological Chemistry, 2004. **279**(19): p. 19747-19754.
343. Song, Y., et al., *The Bisecting GlcNAc on N-Glycans Inhibits Growth Factor Signaling and Retards Mammary Tumor Progression.* Cancer Research, 2010. **70**(8): p. 3361-3371.
344. Huober, J., et al., *Prognosis of medullary breast cancer: analysis of 13 International Breast Cancer Study Group (IBCSG) trials.* Annals of Oncology, 2012.
345. Burchell, J., et al., *An alpha2,3 sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas.* Glycobiology, 1999. **9**(12): p. 1307-11.
346. Wu, C., et al., *N-Acetylgalactosaminyltransferase-14 as a potential biomarker for breast cancer by immunohistochemistry.* BMC Cancer, 2010. **10**: p. 123.
347. Saldova, R., et al., *Association of N-Glycosylation with Breast Carcinoma and Systemic Features Using High-Resolution Quantitative UPLC.* J Proteome Res, 2014. **13**(5): p. 2314-2327.
348. Saldova, R., et al., *Glycosylation Changes on Serum Glycoproteins in Ovarian Cancer May Contribute to Disease Pathogenesis.* Disease Markers, 2008. **25**(4-5).
349. *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
350. Diamandis, E., *The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem?* BMC Medicine, 2012. **10**(1): p. 87.
351. Taketa, K., et al., *A collaborative study for the evaluation of lectin-reactive alpha-fetoproteins in early detection of hepatocellular carcinoma.* Cancer Res, 1993. **53**(22): p. 5419-23.
352. White, K.Y., et al., *Glycomic characterization of prostate-specific antigen and prostatic acid phosphatase in prostate cancer and benign disease seminal plasma fluids.* J Proteome Res, 2009. **8**(2): p. 620-30.
353. Chandler, K. and R. Goldman, *Glycoprotein Disease Markers and Single Protein-omics.* Molecular & Cellular Proteomics, 2013. **12**(4): p. 836-845.
354. Elinav, E., et al., *Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms.* Nat Rev Cancer, 2013. **13**(11): p. 759-771.
355. DeNardo, D. and L. Coussens, *Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression.* Breast Cancer Research, 2007. **9**(4): p. 212.
356. Ralin, D., et al., *Kinetic Analysis of Glycoprotein–Lectin Interactions by Label-Free Internal Reflection Ellipsometry.* Clinical Proteomics, 2008. **4**(1-2): p. 37-46.

357. Kelder, T., et al., *Finding the Right Questions: Exploratory Pathway Analysis to Enhance Biological Discovery in Large Datasets*. PLoS Biol, 2010. **8**(8): p. e1000472.