MCP Papers in Press. Published on December 6, 2011 as Manuscript M111.010660

Extensive Determination of Glycan Heterogeneity Reveals an Unusual Abundance of High-Mannose Glycans in Enriched Plasma Membranes of Human Embryonic Stem Cells

Hyun Joo An[a], PhungGip[b], JaehanKim[c], ShuaiWu[d],Kun Wook Park[a], Cheryl T. McVaugh[b], David V. Schaffer[e], Carolyn R. Bertozzi[b,f], and Carlito B. Lebirlla[d,g,1].

a. *Graduate School of Analytical Science and Technology and Cancer Research Institute, Chungnam National University, Daejeon, South Korea.*
b. *Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720*
c. *Department of Food Nutrition,Chungnam National University, Daejeon, South Korea.*
d. *Department of Chemistry, University of California, Davis, CA 95616*
e. *Department of Chemical Engineering and Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720*
f. *Departments of Chemistry and Howard Hughes Medical Institute, University of California, and Materials Sciences Division and The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*
g. *Department of Biochemistry and Molecular Medicine, University of California, Davis, CA 95616*

Hyun Joo An and PhungGip have contributed equally to this work as co-first authors.
Carlito Lebrilla, Carolyn Bertozzi, and David Schaffer have contributed equally.

[1]To whom correspondence should be addressed:
Carlito B. Lebrilla, Email address: cblebrilla@ucdavis.edu; Tel: +1-530-752-0504; Fax: +1-530-752-8995

Running Title: Cell Surface Glycome using Mass Spectrometry

Abbreviations

SPE, Solid-Phase Extraction

ACN, acetonitrile

PGC, Porous Graphitized Carbon

MALDI, Matrix-Assisted Laser Desorption/Ionization

MS, Mass Spectrometry

TOF, Time of Flight

Hex, Hexose

HexNAc, N-acetylhexosamine

NeuAc, Sialic acid

Fuc, Fucose

Man, Mannose

Summary

Most cell membrane proteins are known or predicted to be glycosylated in eukaryotic organisms, where surface glycans are essential in many biological processes including cell development and differentiation. Nonetheless, the glycosylation on cell membranes remains not well characterizeddue to the lack of sensitive analytical methods. This study introduces a technique for the rapid profiling and quantitation of N- and O-glycans on cell membranes using membrane enrichment and nanoflow liquid chromatography/mass spectrometry of native structures. Using this new method, the glycome analysis of cell membranes isolated from human embryonic stem cells and somatic cell lines was performed. Human embryonic stem cells (hESCs) were found to have high levels of high- mannose glycans, which contrasts with IMR-90 fibroblasts and a human normal breast cell line, where complex glycans are by far the most abundant and high-mannose glycans are minor components. O-glycosylation is relatively minor components of cell surfaces. To verify the quantitation and localization of glycans on the hESCs membranes, flow cytometry and immunocytochemistry were performed. Proteomics analyses were also performed and confirmed enrichment of plasma membrane proteins with some contamination from endoplasmic reticulum and other membranes. These findings suggest that high-mannose glycans are the major component of cell surface glycosylation with even terminal glucoses. High mannose glycans are not commonly presented on the surfaces of mammalian cells or in serum,yet may play important roles in stem cell biology. The results also mean that distinguishingstem cells from other mammalian cells may be facilitated by the major difference in the glycosylation of the cell membrane. The deep structural analysis enabled by this new method will enable future mechanistic studies on the biological significance of high-

mannose glycans on stem cell membranes and provide a general tool to examine cell surface glycosylation.

Introduction

Glycosylation is the process by which oligosaccharides, termed glycans, are appended onto membrane and secreted proteins and lipids. It is the most common and complex form of post-translational modification, with approximately 50% of all eukaryotic proteins glycosylated(1, 2). A majority of glycans are found on the cell surface, where they are optimally poised to be the first cellular components encountered by approaching cells, pathogens, antibodies, or other molecules, as well as advertise information about the cell's internal state and homeostasis(3, 4). Therefore, glycans play an essential role in many biological processes, including cell development and differentiation, cell-cell or cell-matrix communication, and pathogen-host recognition(3, 5-7). In fact, differences in glycan profiles between healthy and diseased states are utilized for clinical diagnosis(7), providing targets for many novel classes of therapeutics including cancer chemotherapy, diabetes treatment, and antibiotic and anti-viral medicine(5, 8). Glycans are highly heterogeneous in nature, varying in the composition of individual monosaccharide building blocks, the positions with which these building blocks link to each other, and the stereochemical disposition of the linkages (α or β). This complexity has presented a significant challenge for obtaining structural information about glycans at the molecular level, particularly in contextsrelevant to native cellular physiology. For this reason, the technology for elucidating their structures has lagged behind other major classes of biomolecules - such as protein, DNA, and RNA - in the molecular revolution currently underway in biology and medicine(9). Although several approaches including lectin binding(10-12), cell surface shaving(13), cellsurface labeling(14, 15), and antibody-mediated membrane enrichment

have been developed to identify surface glycans or glycoproteins, comprehensive and conclusive structural elucidation and identification remain laborious and difficult (16).

Human embryonic stem cells (hESCs) are of particular biomedical interest as they hold enormous potential for regenerative medicine and drug discovery. As a model system, they can also contribute to the understanding of human development and potentially help to guide cancer research as hESCs and cancer cells share similar characteristics. (17-19). Therefore, structural elucidation of the components present on hESC membranes may provide a basis for understanding their role in hESC maintenance and differentiation. Recent studies suggest that glycans on the plasma membrane of hESCs change during differentiation, these changes can have profound effects on cellular function (20) and could be harnessed to meet the need to identify cell surface markers for isolating and purifying specific cell populations for therapeutic application. Indeed, one of the earliestpluripotent stem cell markers is stage-specific embryonic antigen (SSEA)-1, a glycan, otherwise known as Lexis X antigen, expressed on mouse embryonic stem cells(21, 22). Two other antigenic epitopes discovered were SSEA-3 and SSEA-4(23), which are both glycolipids that have become the most common cell surface markers used to characterize hESCs (24).

Cell surface glycosylation may play an important role in development and may provide important new sources of markers for differentiation. Studies regarding the glycosylation of stem cell surfacesare limited. Wearne et al. recently reported the use of fluorescence-labeled lectins to identify a number of specific structural motifs including mannose residues, α2,3- and α2-6-linked N-acetylneuraminic acid, α1,6-linked L-Fucosyl, and β-D-galactosyl groups(12). In addition, they also found a number of common antigens including T, Tn, and sialyl-Tn. A more

structurally intensive study of whole stem cell glycosylation was reported by Satomaa et al. (25). The cell surface studies were limited to lectins, which cannot be used quantitatively. Structural methods including nuclear magnetic resonance, mass spectrometry and glycosidase digestion,were used on whole cells where they showed that the N-glycan profile was rich high mannose glycans as well as complex type structures that are terminated by both α2,3 and α2,6-sialylated oligosaccharides, and fucosyl structures.

Here we characterize the N-glycan profile of human embryonic stem cell membrane glycans using a method that enables specific detection and quantitation, and acquisition of structural information. Interestingly, hESCs have high levels of high-mannose glycans on the cell surface, which is largely unprecedented in mammalian cells. Moreover, the hESCs were particularly rich in Man8 and Man9 structures, including Man 9 with terminal glycoside still intact. This unusual glycomic signature might have functional implications as well as practical utility in characterization and isolation of hESCs.

Experimental Procedures

Human Embryonic Stem Cell (hESC) Culture. The NIH approved hESC lines,H1 and HSF-6, were maintained under feeder-free conditions using a chemically defined media, X-VIVO 10 (Cambrex, Walkersville, MD) supplemented with human recombinant growth factors FGF and TGF-β1 (80 ng/mL and 0.5ng/mL, respectively; R&D Systems[10]. Cells are propagated on hESC-qualified Matrigel-coated plates (BD Biosciences). Media was exchanged daily after the first 48 hr in culture, and cells were passaged every 5 to 7 days using collagenase IV (200 units/ml; Invitrogen) and mechanically removed. For glycan analysis, cells were collected after

collagenase IV treatment, centrifuged, washed with PBS (Invitrogen) pelleted and frozen on dry ice. Approximately 50 million cells were counted and collected at different passage numbers to obtain biological triplicates. Karyotype analysis was routinely performed and indicated that all samples were diploid and had no chromosomal abnormalities. Cells were routinely stained with pluripotent markers Oct4 and SSEA-4.

Somatic Cell Line Culture.IMR-90 human fibroblast cells (UC Berkeley Tissue Culture Facility) were grown in DMEM high glucose media (Invitrogen) supplemented with 10% FBS (fetal bovine serum, HyClone). Cells were passaged every 3 days using trypsin 0.25% and EDTA solution (Invitrogen), For glycan analysis, IMR-90's were collected using 0.5 mM EDTA, centrifuged and washed with PBS, pelleted and frozen on dry ice.MCF10A human breast epithelial cells (ATCC, CRL-10317) were grown in DMEM with high glucose (Gibco, 31053) supplemented with 10% FBS (PAA Laboratories), penicillin/streptomycin (10 U/ml and 10 μg/ml; Gibco), 2.5 μg/ml fungizone (Gibco), 20 ng/ml EGF (epidermal growth factor; Biovision,), 0.5 μg/ml hydrocortisone (VWR), 100 ng/ml cholera toxin (VWR), and 10 μg/ml recombinant human insulin (Sigma). MCF10A's were passaged weekly by trypsinization. For glycan analysis, cells were detached using a cell scraper, washed with PBS, pelleted and frozen on dry ice. Approximately 50 million cells were counted and collected from both cell lines at different passages to obtain biological triplicates.

Cell Membrane Extraction.Membrane extraction was performed using ultracentrifugation. Pellets were thawed on ice with the addition of a homogenization buffer (HB) consisting of 0.25

M sucrose, 20 mM Hepes-KOH, pH 7.4 and protease inhibitor mixture (1:100; Calbiochem/EMD Chemicals). Cells were sonicated on ice, and cell lysates were centrifuged at 1,000g for 10 min to remove the nuclear fraction and debris. The supernatant was collected and additional HB was added for ultracentrifugation at 200,000g for 45 min at 4°C to remove the cytoplasmic fraction. The pellets were resuspended in 0.2 M Na_2CO_3 (pH 11) to break up the microsomes. The samples were spun twice more at 200,000g for 45 min to wash the samples of the cytoplasmic fraction. The supernatant was removed and the membrane fractions were frozen at -20 °C.

Western Blot Analysis.All fractions (nuclear, cytoplasmic, and membranes) were analyzed by SDS-PAGE followed by Western blot using known organelle-specific markers for the nucleus (nuclear pore complex proteins; Covance), endoplasmic reticulum (Bip/GRP78; BD Biosciences), cytosol (α-tubulin; Sigma), and the plasma membrane (CD49b; BD Biosciences). Primary antibodies were probed with a horseradish peroxidase conjugated anti-mouse secondary antibody (HRP-anti-mouse IgG). Before Western blot analysis, membrane pellets were resuspended in 4% SDS buffer and protein concentration was determined by the BCA assay (Pierce). The samples (4ug) were separated by SDS/PAGE (4-12%, Bio-Rad).

Glycan Release.For the analysis of N-glycans,100 μL of 100 mM ammonium bicarbonate (NH_4HCO_3), 5 mM dithiothreitol (DTT, from Promega) was added to the samples and heated to 100^oC for 2 min to denature the protein. After cooling at room temperature, 2 μL of Peptide N-glycosidase F (PNGase F, from New England Biolabs) was added to the mixture (pH 7.5)

andincubated at 37^{o}C for 12 hours in a water bath. 800 μL of chilled ethanol was added, and the mixture was frozen at -80°C for 1 hour and then centrifuged to separate glycans from deglycosylated proteins. The supernatant was completely dried down to remove the ethanol prior to solid phase extraction (SPE) using a graphitized carbon cartridge (GCC, from Alltech).

For O-glycan analysis, alkaline borohydride solution (500μL, mixture of 1.0 M sodium borohydride and 0.1 M sodium hydroxide) was added to the membrane fraction. The mixture was incubated at 42°C for 12 hours in a water bath. The addition of 1.0 M hydrochloric acid solution was slowly added in ice bath to stop the reaction and destroy excess sodium borohydride.

Glycan Enrichment.Released N- and O-glycans were purified and enriched by SPE-GCC. Prior to use, graphitized carbon cartridge (150 mg bed weight, 4 mL cartridge volume) was washed with nanopure water followed by 80% acetonitrile (AcN) in 0.05% (v/v) trifluoroacetic acid (TFA) (v/v) and again with nanopure water. Glycan solutions were applied to the GCC cartridge and subsequently washed with several cartridge volumes of nanopure water at a flow rate of 1 mL/min to remove salts. Glycans were eluted stepwise with 10% AcN in H_2O (v/v), 20% AcN in H_2O (v/v), and 40% AcN in 0.05% TFA in H_2O (v/v). Each fraction was collected and concentrated *in vacuo* prior to mass spectrometry analysis. Fractions were reconstituted in nanopure water prior to MS analysis.

Mass Spectrometric Analysis.Mass spectra were recorded on a Fourier transform-ion cyclotron resonance mass spectrometer (FT-ICR MS) with an external source HiResMALDI (IonSpec

Corporation) equipped with a 7.0 Tesla magnet. The HiResMALDI was equipped with a pulsed Nd:YAG laser (355 nm). 2,5-Dihydroxy-benzoic acid (DHB) was used as a matrix (5 mg/100 mL in 50% AcN:H_2O) for both positive and negative modes. A saturated solution of NaCl in 50% AcN in H_2O was used as a cation dopant to increase signal sensitivity. The glycan solution (0.7 μL) was applied to the MALDI probe followed by matrix solution (0.7 μL). The sample was dried under vacuum prior to mass spectrometric analysis.

Structural Determination Using Infrared Multiphoton Dissociation (IRMPD).A desired ion was readily selected in the analyzer with the use of an arbitrary-wave form generator and a frequency synthesizer. A continuous wave Parallax CO_2 laser with 20 W maximum power and 10.6 μm wavelength was installed at the rear of the magnet and was used to provide the photons for IRMPD. The laser beam diameter is 6 mm as specified by the manufacturer. The laser beam was expanded to 12 mm by means of a 2× beam expander (Synrad) to ensure complete irradiation of the ion cloud through the course of the experiment. The laser was aligned and directed to the center of the ICR cell through a BaF_2 window (Bicron Corporation). Photon irradiation time was optimized to produce the greatest number and abundance of fragment ions. The laser was operated at an output of approximately 13 W.

NanoLC Mass Spectrometry. GCC fractions were analyzed using a microfluidic HPLC-Chip-TOF MS (Agilent, CA). The microfluidic HPLC-CHIP consists of an enrichment column, an LC separation column packed with porous graphitized carbon, and a nanoelectrospray tip. Separation was performed by a binary gradient A: 3% acetonitrile in 0.1% formic acid solution

and B: 90% acetonitrile in 0.1% formic acid solution. The column was initially equilibrated and eluted with the flow rate at 0.3 μL/min for nanopump and 4 μL/min for capillary pump. The 65 minute gradient was programmed as follows: 2.5-20 minute, 0%-16% B; 20-30 minute, 16%-44% B; 30-35 minute, B increased to 100%, then continued 100% B to 45 minute, finally 0% B for 20 minute to equilibrate the chip column before next sample injection. Each possible composition of N-glycans were identified with the in-house program "GlycoX"(26)and "N-glycan library" (27)according to the mass tolerance with additional retention times and abundance information noted at the same time.

Proteomic Analyses by Liquid Chromatography/Tandem Mass Spectrometry (LC-MS/MS) Analyses. Membrane proteins were dried and solubilized with 60 μL of 8 M urea. Samples were then reduced with DTT and alkylated with IoAA. After dilution in 180 μL of water, an overnight digestion with trypsin was performed. Peptides were then concentrated and desalted using C18 peptide trap (MichromeBioResources, Inc. Auburn, CA, USA) before LC separation and online MS/MS. A nanoLC-2Dsystem (Eksigent, Dublin,CA) coupled with an LTQ ion trap mass spectrometer (Thermo Finnigan) was used with a homemade fritless reverse phase microcapillary column (75 μm×180 mm; packed with Magic C18AQ, 3 μm 100 Å : Michrom Bio Resources) and vented column configuration. Digested samples were transferred from the autosampler to the online trap column (0.15 mm×20mm; packed with Magic C18AQ, 3 μm100Å) and desalted. Peptides were eluted from the trap and separated on the capillary column using a reverse-phase gradient at a flow rate of 300 nL/min and directly electro sprayed into the mass spectrometer. A cycle of one MS survey scan followed by 10 MS/MS scans was repeatedly

acquired over the LC gradient. Dynamic exclusion for 1 min duration was utilized. Buffers were 0.1% formic acid in water (buffer A)and 0.1% formic acid in acetonitrile (buffer B). A 107min gradient (2-40% B for 95 min, followed by 40-80% B for 12 min) was used. Protein identification based on LC-MS/MS was performed using X! Tandem with a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 1.8 Da. IoAA(iodoacetamide) derivatization of cysteine was specified as a fixed modification. Peptide identifications were accepted if they could be established at greater than 95.0%probability as specified by the Peptide Prophet algorithm. Protein identifications were accepted if they could be established at greater than 99.0%probability and contained at least two identified peptides.

Immunofluorescence. hESCs were fixed with 2% paraformaldehyde and rinsed 3 times in PBS. Cells were blocked with staining buffer (2% fetal bovine serum in PBS) and then stained with pluripotent marker SSEA-4 (2.5μg/500 μL/well; Milllpore) for 30 minutes at RT. The wells were rinsed in PBS before adding FITC-conjugated lectins (20 μg/ml; EY Labs) and Alexa Fluor 594-conjugated goat-anti-mouse secondary antibody to SSEA-4 (1:400; Invitrogen). Wells were rinsed in PBS and stained with a solution of 1X Hoechst 23187 (Sigma) as a nuclear stain and analyzed using an Olympus IX71 fluorescent microscope. Control wells were stained with mouse IgG3 isotype (Invitrogen) and lectins incubated with their respective inhibitory sugar.

Flow Cytometry. hESCs were collected after incubation with collagenase IV (200 units/mL; Invitrogen) and mechanically removed. Colonies were dissociated into single cell suspensions in 0.5mM EDTA, were then filtered through a 40 micron cell strainer,and counted. Cells were

blocked with staining buffer (2% fetal bovine serum in PBS) and then stained with pluripotent marker SSEA-4 (2.5μg/500 μL/500,000 cells; Millipore) for 30 minutes on ice. Cells were washed, stained with APC-conjugated goat-anti- mouse secondary antibody to SSEA-4, and 5, 10, 20 or 40 μg/mL of the following FITC-conjugated lectins: *Canavaliaensiformis*(Con A) or*Galanthusnivalis* (GNA; EY Labs). To validate binding specificity, hESCs were also stained with lectins pre-incubated with sugar haptens: methy-α-mannoside, yeast mannan, respectively (Sigma). After 30 minutes on ice, cells were washed and resuspended in staining buffer with propidium iodide to distinguish dead cells from live cells. Flow cytometry (BD FACs Calibur from BD Biosciences) was performed and data was analyzed using FlowJo software (TreeStarInc). At least three independent assays were carried out for each lectin. The final quantitation represents live hESCs that were double-labeled with SSEA-4 and FITC-conjugated lectins. hESCswere also stained with mouse IgG3 isotype (Invitrogen), as a control for SSEA-4 labeling.

Results

The experimental strategy - including i) the purification of cell membrane fractions from whole cell lysates by ultracentrifugation, ii) release and enrichment of surface glycans by solid phase extraction (SPE) using a graphitized carbon, iii) glycan profiling by high performance mass spectrometry, and iv) isomer separation and quantitation by nanoflow liquid chromatography (nanoLC) - is outlined in **Fig.S1**(*SI Appendix*). These methods were developed and streamlined to profile cell membrane glycans effectively and selectively by mass spectrometry. As further described below, the quantitation and localization of glycans on the hESCs membrane were also validated by flow cytometry and immunocytochemistry.

Isolation of Membrane Fractions by Ultracentrifugation in hESCs

Selective isolation of membrane fractions from whole cells with a compatible buffer that allows MS detection of glycans was a crucial component of the methodology development. Although plasma membrane purification would have been a more rigorous approach, the sensitivity of the method at this point is still insufficient for the analysis of glycans from less than 50 million cells.

Ultracentrifugationwas employed as the technique to isolate membrane fractions from cells. However, while mass spectrometry can be a powerful method for analyzing biomolecules because of its intrinsic speed and sensitivity, coupling the two techniques hasproven challenging due in part to such incompatibilities in sample processing (16). For example, the typical buffer solution used in ultracentrifugation contains high concentration of detergents and solvents such as sodium dodecyl sulfate (SDS), Triton X-100, EDTA, sucrose, sodium, and protease inhibitors, and several of these components are known to be deleterious to the mass spectrometry analyses.

The buffer solutions usedin ultracentrifugation were thus optimized to render them more compatible with mass spectrometry.Homogenization buffer (HB) consisting of 0.25 M sucrose, 20 mM Hepes-KOH, 1 mM EDTA, 1% SDS, and a protease inhibitor mixture was initially used. However, in the initialMS analysis only polymeric material with a regular mass spacing andnon-carbon isotopic distribution were observed in **Fig.S2** (*SI Appendix*), indicative of chemical background material only and an absence or complete suppression of the signal from glycans.To resolve this issue, the buffer conditions were reformulated. One key change was the

removal of EDTA from the homogenization buffer - a change that did not affect the quality of the cellular fractionation step, but more importantly allowed observation of masses correlating to glycans in the mass spectra. Other changes included eliminating SDS and adding rigorous washes with nanopure water.The resulting purity of the membrane fractions was validated by SDS-PAGE gel electrophoresis followed by Western blotting using organelle-specific antibodies (*SI Appendix,***Fig. S3**).

Proteomic analyses were performed to determine the gross protein and membrane protein content in H1 stem cell. LC-MS/MS analyses identified a total of 335 proteins present in H1 membrane fraction. Two algorithms using HMMTOP2.0 and TMHMM2.0were used in order to predict integral membrane proteins and trans-membrane helices. Of the 335 proteins identified in the membrane fractions, 224 (59.1%) proteins contain at least one sequence predicted to be trans-membrane helices by the HMMTOP (Hidden Markov Mode for Topology prediction) method(28),while 161 (42.5%) proteins containat least one by the TMHMM2.0 (Tied Mixture Hidden Markov Model) method(29). We further performed Gene Ontology (GO) term analysis of the H1 stem cell membrane fraction using Database for Annotation, Visualization and Integrated Discovery (DAVID) V6.7(30) to determinethe membrane portion relative to other cell parts. In the process, seven proteins were excluded because their gene symbolswere not assigned. Of the remaining 321 proteins used in the GO analysis, 217 proteins (67.6%) are categorized as cell membrane (*SI Appendix,***Fig. S4**). We examined these proteins further and found plasma membrane proteins that have been confirmed by earlier publications (31-33).These proteins are listed in **Table S1** (SI Appendix) along with their potential sites of glycosylation based on known consensus sequences.The entire list of proteins identified in the

membrane fraction is provided as **Supplemental Data** (SI Appendix, excel format). Included in the list are proteins categorized as belonging to membranes other the cell membranes such as ribosome and hence endoplasmic reticulum suggesting that there are contributions from other membranes in the fraction. However, as membranes sometimes share the same proteins, and the complete identification of cell membrane proteins are still ongoing it may be difficult to provide conclusive determination as to what is and is not a cell membrane protein.

Mass Profiling of Glycans from hESC Cell Membranes

Global release methods were used to separate glycan from protein, including chemical and enzymatic to access O- and N-glycans, respectively. The exact glycan compositions of hexoses (Hex), N-acetylhexosamines (HexNAc), sialic acids (NeuAc), and fucoses (Fuc) were deduced based on their accurate masses. The most abundant glycans were further analyzed with tandem mass spectrometry using infrared multiphoton dissociation (IRMPD) to obtain further information about the structural connectivity of the individual monosaccharides within the native glycans.

Representative spectra of N-glycans found on hESC membranes of the two different hES cell lines HSF-6 and H1 are shown in **Fig. 1**. The putative structures of onlythe abundant species were assignedbased on known glycobiology (or N-glycan biosynthesis) and tandem mass spectrometry; however, zooming in on the low abundance signals resulted in the identification of even more glycans.Overall, glycan mass profiling from H1 and HSF-6 cellsis very similar. The *m/z* 1905.643 ($[M+Na]^+$ corresponding to $GlcNAc_2Man_9$ (Man9) is the base peak,and high

mannose glycans are in abundance in both hES cell lines. In this analysis, approximately 42 N-glycan compositions were identified on the hES cell membranes (*SI Appendix,***Table S2**).

To confirm glycan compositions,as well as obtain detailed structural information such as the putative glycan structures shown in **Fig. 1**, selected ions were subjected to tandem mass spectrometry using IRMPD. Tandem MS allows for the observance of ions arising from sequential loss of individual monosaccharides from the native glycan structure, thus supplementingthe composition data obtained above via MALDI-FTICR MS with information on the connections of the monosaccharides to each other. One representative IRMPD analysis is shown for the ion at *m/z* 2067.687 ($[M+Na]^+$), corresponding to 2HexNAc and 10Hex (9Man+1Hex), for which tandem MS afforded extensive fragments in a single MS/MS event(*SI Appendix,***Fig. S5**). The glycosidic bond cleavages corresponding to loss of the two core GlcNAc residues present in all N-glycans (*m/z* 1847 and 1643, respectively) were readily observed, along with fragments corresponding to subsequent monosaccharide losses (i.e., *m/z* 1847, 1685, 1523, 1361, 1198 and *m/z* 1643, 1481, 1319, 1157, 995, 833). The ion at *m/z* 833 corresponds to the trimannosyl core with two extra mannoses ($[Man5-H_2O+Na]^+$), confirming that this glycan is a high mannose type N-glycan. By contrast, the IRMPD spectrum of a complex type N-glycan (*m/z* 2012.719) is shown in *SI Appendix,***Fig S5**.

O-glycans were released from the membrane glycoproteins of hESCs by reductive β-elimination and enriched by SPE. The results indicate that O-glycans are presenton hESC membranes, butatsignificantlylower abundances than N-glycans. The O-glycans are primarily neutral oligosaccharides comprised of mucin core type structures containing N-acetyl hexosamine (HexNAc) and hexose (Hex) residues. The dominant mass in three biological

replicates, at relatively high abundances, corresponded to *m/z* 772 (2HexNAc:2Hex). All O-glycans found on the cell membrane commonly have a 2HexNAc:2Hex motif, indicating the presence of the *core 1* structure(34). However, as shown in *SI Appendix***FigS6**, the relatively high abundances ofresidual N-glycans even after treatment with PNGase F suppressed the signals from the O-glycans. With the somatic cell membranes, O-glycosylation was even less as to be undetectable with this method. As such, this report focused on N-glycan profiling of hESC membranes.

Mass Profiling of Glycans from Somatic Cell Membranes

To determine if the MS based method employed in this study can be applied to other cell lines, and if somatic cells present high-mannose glycans at the same abundance as hESCs, N-glycans were released from membrane fractions of IMR-90 fibroblast and MCF-10Ahuman breast cell line and profiled by mass spectrometry. These somatic cells lines are commonly used, in part because they can go through several passages before cellular senescence(35, 36). The methodprovides sufficient cellular material necessary for MS analysis of membrane N-glycans but appear to provide little or no O-glycans suggesting that O-glycans are significantly lower abundant than N-glycans, as previously mentioned. Glycans found in membranes of the two somatic cell lines were also summarized in *SI Appendix,***Table S2**.In general, the abundant species correspond to complex type structures for both cell lines, although high mannose glycans are more abundant in IMR-90 fibroblast than in breast cell lines. For example, the complex-type composition with *m/z* 1809.651 (5Hex:4HexNAc:1Fuc) is the base peak in IMR-90 fibroblast and the composition with*m/z* 2539.910 (7Hex:6HexNAc:1Fuc) is the base peak for

MCF-10A breast cells. Both compositions are minor components in the stem cells. Therefore, complex type glycans containingHex:HexNAc:Fuc:NeuAc are the major species found inhuman somatic cell lines, which contrast withhESCs.

Isomer Separation and Quantitation of Cell Surface Glycans

Glycan composition can be readily obtained based solely on MALDI-FTICR MS, and the abundances can also be determined with reasonable precisionusing peak intensitiesif the homogeneity of the sample surface can be guaranteed with the proper matrix (37, 38). However, this analysis does not provide information on the number of isomers associated with the compositions. Therefore, to analyze the isomers, glycans were further examined with nanoflow liquid chromatography (nanoLC). Amicrochip packed with graphitized carbon was used to chromatographically separate the glycans in this study(39). The microchip was interfaced with TOF mass analyzer that routinely provides a mass measurement accuracy of less than 5 ppm.This technique achieves both isomer separation and direct quantitationwith greater precision than MALDI MS.

Aliquots of each SPE fraction were combined, and the mixtureswereanalyzed by nanoLC MS in triplicates. Representative base peak chromatograms of N-glycans for four different cell membranes are shown in **Fig.2A-D**, with the putative structures for the most abundant glycans assigned to specific peaks, keeping in mind that each peak may still be composed of small mixtures.Both stem cells, H1 and HSF-6, show very similar base peak chromatograms and contrasts to the IMR-90 and MCF-10A cell lines. In both human embryonic stem cell membranes, high-mannose structures are the most abundant glycans and eluted earlier

(between 14 and 20 min). The IMR-90 and MCF-10A, contain more complex type glycans, which elute later(between 20 and 28 min). In general, the mass profilesperformed by MALDI-FTICR MS are consistent with the abundancesfrom nanoLC/MS (**Fig. 2A-B**and*SI Appendix,***Table S2**). In addition, differentiation between IMR-90 and MCF-10A is also possible, because the profile ofIMR-90 appears to be in intermediate between the stem cells and MCF10A.

We identified an average of 170 distinct features, which includes anomers and other isomers, arising from an average of 69 glycan compositions in H1 and HSF-6 hESC membranes (*SI Appendix,***Table S3**).Representative extracted ion chromatograms of high mannose glycans found on HSF-6 hESC membrane, $GlcNAc_2Man_8$(Man8) and $GlcNAc_2Man_7$ (Man7), are shown in **Figure 3**to illustrate the separation and the number of isomers that may comprise a specific composition. The anomeric contribution due to the separation on porous graphitized carbon (40)is small for these species so that the peaks correspond in these cases to linkage isomers. The extracted ion chromatogram of Man8 shows two structural isomers for *m/z* 861.302 ($[M+H]^{+2}$) with retention times of 15.66 and 16.66 min. Similarly, Man7 shows three structural isomers with retention times of 15.54, 16.44, and 16.56 min. H1 hESC membrane also shows the same number of isomers from Man7 and Man8 but with different peak abundances (*SI Appendix*,**Figure S7**and **Table S3**).

Chromatographic peak areaswere used for overall glycan quantitation. N-glycans are classified into two groups, namely high mannose and complex/hybrid type (non-high mannose glycan). The relative abundances of N-glycan types are depicted in **Fig. 4A**. Indeed, high-mannose glycans are the most abundant in both H1 and HSF-6 hESCs, accounting for 74% in H1 and 85% in HSF-6 of the total N-glycans identified, while complex/hybrid type N-glycans

comprise the remaining 26% in H1 and 15% in HSF-6, respectively. Man8and Man9 are the major glycans present in two hESC membranes, accounting for more than 50% of total high mannose glycans (**Fig. 4B**). However, human IMR-90 fibroblasts and MCF-10A breast cells had significantly lower levels of high-mannose glycans (45% and 28%, respectively) compared to hESCs. For the MCF-10A cells the complex-type glycans dominate accounting for more than 70% (**Fig. 4B**), while for IMR-90 it is 55%. Finally, the complex/hybrid type glycans are generally fucosylated in both hESC lines, with more than 70% of such glycans containing at least one fucose residue (**Fig. 4B**). For IMR-90 and MCF-10A the degree of fucosylation in the complex type glycans is76% and 68%, respectively (data not shown).

The reproducibility between biological replicates and analytical replicates was examined. Biological replicates consisted of hESCs collected on different days from different passage numbers, whereas analytical replicates were from the same biological sample analyzed multiple times. For the analytical replicates, normalized absolute peak intensities measured by three MS experiments were plotted against one other (*SI Appendix,***Fig. S8A**). High correlation coefficients (r: 0.96-0.97) were obtained from three analytical replicates. The variation among biological experiments was further examined. The correlation coefficients (r) between the normalized intensity of glycan peaks were in the range of 0.86 to 0.92, showing good correlation between biological replicates (*SI Appendix,***Fig.S8B**). The results clearly show that profiling cell membrane glycans by MS is highly reproducible.

Validation of Glycan Expression using Lectins

The majority of glycans found on hESCs surfacevia the MS approach were high-mannose

N-glycans. To determine whether these types of glycans, observed in the general cellular membrane fractions, are specifically found on hESC surfaces, hESCs were labeled with lectins -- plant and animal proteins with natural carbohydrate binding functionality. *Canavaliaensiformis*(Con A) was used for identifying N-glycans generally, and*Galanthusnivalis* (GNA) detected high-mannose N-glycans specificallywhile*Artocarpusintegrifolia*(JAC) was used for identifying *O*-glycans. Con A recognizes branched α-mannosidic structures,high-mannose, and hybrid and biantennary complex type N-glycans. GNA is a highly specific mannose lectin used to detectterminal α 1-3 linked high-mannose structures. JAC is a specific for α-D-galactose, and oligosaccharides terminating with α-D-galactose (i.e. T antigen, Galß1-3GalNAc), which is the *core* structure of mucin. These lectins were utilized on both fixed and live hESCs and analyzed via microscopy and flow cytometry.

To visualize the presence of glycans byfluorescence microscopy, hESCs were fixed and labeled with an antibody against SSEA-4 (a marker for pluripotent hESCs) followed by lectins conjugated to the fluorescein isothiocyanate (FITC) fluorophore. Binding of Con A (**Fig. 5A**) and GNA (**Fig. 5B**) indicated the presence of both N-glycans and terminal high mannose glycans, respectively. These results are consistent with the type of glycans analyzed inthis mass spectrometry method.

To quantify the cell surfaceglycans, live hESCs were stained with low orhigh concentrations of the lectins, and for the pluripotent marker SSEA-4, for flow cytometryanalysis. The majority of cells stained positive for both SSEA-4 and Con A(black, **Fig. 5C**), and the majority of cells were also stained by GNA(**Fig. 5C-D**). In addition, live hESCs expressed dose-dependent binding of both lectins Con A and GNA (*SI Appendix,***Figure S9**),indicating the specificity of lectin

binding and presence of N-glycans and terminal high mannose glycans on the cell surface. These results suggest that high-mannose glycans detected inthis mass spectrometry approach were derived from the cell surface.Similar to the absence of JAC binding (*SI Appendix,***Figure S9**), there was weak or no binding of JAC on the cell surface. These results confirmed our mass spectrometry approach to profile surface glycans, showing low O-glycan expression.

Discussion

In this study, we developed a method to measure cell surface glycosylation and provide extensive heterogeneity of the hESC glycome from membrane proteins. Enrichment of membrane proteins followed by high performance mass spectrometry and chromatography makes this approach both highly specific and sensitive.The resultingapproach provides comprehensive and highly quantitative structural information including isomer separation.

Stem cell surface glycosylation is dominated by high mannose glycans (**Fig. 4A**). These results are consistent with both the lectin cell surface studies by Wearne et al. (12)and the glycomic profile of whole hESC by Satomaa et al.(25)The major difference between this study and the previous two is that we enrich plasma membrane and determine the glycosylation specifically from the membrane. We confirm that the products we obtain are primarily from the membrane. We then confirm the results with lectins and flow cytometry analyses. Furthermore, we quantitate the relative contribution of each glycan types as well as specific isomers on the cell surface.These experiments differ from those by Wearne et al. (12) and Satomaa et al.(25) in that those were performed on whole cell lysates. The cell surface glycosylation were inferred indirectly using lectins, which are not quantitative when comparing

different glycan types. Therefore, while high mannose may show on surfaces with lectins, so would complex and hybrid types. There is no method using lectins to determine whether high-mannose is more abundant compared to the other glycan types. Glycans from whole cells were indeed examined by MS and by NMR, but there were no efforts to determine the relative concentrations of each glycan types on the cell surface.

The amount of high mannose is unprecedented consisting of approximately 75% of the total N-glycan species, which is the most abundant type of glycosylation. Moreover, hESCs are rich in Man8 and Man9 glycans (**Fig. 4B**) and contain a large amount of Man9+Glc, suggesting that hESC membranes have an unprecedented amount of terminal glucose. These results contrast with fully differentiated cell lines that are richer in complex-type glycansfurthersuggesting new types of glycan markers for determining hESC cells.

Thehigh abundance of high-mannose glycans on hESCs further contrasts with the levels, for example, found in blood and human somatic cell lines, where complex types are by far the most abundant and high-mannose glycans are very minor components (**Figure 4**)(40). More generally, high-mannose glycans are not commonly observed on the surfaces of mammalian cells (41), with the exception of macrophages (42). High-mannose glycans are involved in initial steps in N-glycan processing and control of protein folding in the endoplasmic reticulum. The initiating event is the processing of N-glycan precursor, Man9, and three glucoses by glucosidases I and II to yield to Man9+Glucose, which is subsequently trimmed back in the process of generating complex-type glycans (43). If the membrane fractions were primarily plasma membrane, the abundance of Man 9 suggests that the trimming process is incomplete in hESCs, perhaps yielding “immature” glycoproteins. Indeed, the ion at *m/z* 2067.68,

corresponding to Man9 with one extra hexose, potentially a glucose, was readily identified by tandem MS(*SI Appendix,* **Fig S5**).Sincethe membrane extraction included all types of membranes from the cell, glycans derived from the endoplasmic reticulum (ER) were included in the analyses. However, human somatic cell lines such as IMR-90 fibroblasts and MCF-10A breast cells had significantly lower levels of high mannose glycans compared to hESCs, and in particular MCF-10A cells had predominantly complex-type glycans. Interestingly, IMR-90fibroblasts which are a cell type derived from human fetal lung(35), also have high-mannose glycans on the cell surface, assessed by flow cytometry (data not shown). It is unclear whether this represents an "embryonic" characteristic of IMR-90 fibroblasts that is shared with hESCs, or possibly that high-mannose glycans are serving a similar function in both type of cells. In contrast to IMR-90's,MCF-10A's are a cell type derived from adult breast epithelium(36).

Contrary to mammalian cellsyeast and viruses have high-mannose glycans on the cell surface. Cells in humanimmune system such as macrophages express mannose receptors that can bind to terminal/ high-mannose glycans on the surfaces of non-self entities (e.g. bacteria, yeast and viruses) to detect and initiate pathogen phagocytosis. Thus, investigations of the role of high mannose glycanshave been mostly directed totheir involvement in initiating the innate immune system or in protein folding in the ER. A few studies, however, have suggestedthat high mannose glycans mediate cell-cell fusion (42, 44), mediating sperm-egg fusion,myoblast fusion and osteoclast formation. Thus high-mannose glycans on hESCs may play a role in cellular binding and recognition. More recently, our group and othershave shown that high-mannose glycans are prevalent on the cell surface of various tumor cells compared to normal cells (45, 46). Understanding the role of high-mannose glycans on hESCs could be applicable to cancer

biology since similar characteristicshave been observed between tumor and stem cells, such as the ability to self-renew,the expression of cell surface markersand the activation of signaling pathways(47).

Further analyses will be performed to identify theproteins that high-mannose glycans are modifyingandthe sites of attachment, to discern whether high-mannose glycans are functionally modifying cell surface proteins and the functional significance on both hESCs compared tosomatic cell lines. Also, future studies include investigating changes in high-mannose glycans on the cell surface during differentiation as well as confirming that the glycans come only from surface proteins and not from contamination by the endoplasmic reticulum, although the same contamination should also be present in the other cell lines.

If indeed the glycan structures come primarily from the surface proteins, these findings suggest that the abundance and presence of high-mannose glycansspecifically on the cell surface may have biological significance. However, it has immediate consequences. It suggests that stem cells have less sialylation than differentiated cells, because sialylation accompanies complex and hybrid but are not found in high mannose glycans. Furthermore, the presence of Man9+glucose in high abundances also suggests the unprecedented presence of terminal glucose residues on the surfaces. Thus, the deep structural analysis enabled by this methodology motivates future mechanistic studies on the biological significance cell-surface glycosylation.

ACKNOWLEDGMENTS

We are grateful for funds provided by the National Institute of Health (RO1GM049077 for C.B.L and GM66047 for C.R.B) and California Institute for Regenerative Medicine (CIRM, RS1-00365) for C.R.B.; P.G. was supported by CIRM and National Science Foundation and CIRM pre-doctoral fellowships. The research was also supported by the Converging Research Center Program through the Ministry of Education, Science and Technology (2011K000968 for H.J.A). We would like to thank Dr. Jong Shin Yoo for his assistance in proteomics data analysis.

AUTHOR CONTRIBUTIONS

H.J.A., P.G., and C.B.L. designed the overall research;P.G., D.V.S., and C.R.B.designed the hESCs cells culture, membrane isolation, lectin assays, and immunocytochemistry of hESCs; H.J.A. did all of the MS analyses; S.W. performed the nanoLC-Chip/TOF analyses; J.H.K. did data analysis for peak normalization and reproducibility; G.W.P. did proteomics data analysis and H.J.A., P.G., C.T.M., and C.B.L. wrote the paper.

REFERENCES

1. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Bba-Gen Subjects* 1473, 4-8
2. Lebrilla, C. B., and Mahal, L. K. (2009) Post-translation modifications. *Curr Opin Chem Biol* 13, 373-374
3. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* 126, 855-867
4. Paulson, J. C., Blixt, O., and Collins, B. E. (2006) Sweet spots in functional glycomics. *Nature Chemical Biology* 2, 238-248
5. Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery* 4, 477-488
6. Cooke, C. L., An, H. J., Kim, J., Canfield, D. R., Torres, J., Lebrilla, C. B., and Solnick, J. V. (2009) Modification of Gastric Mucin Oligosaccharide Expression in Rhesus Macaques After Infection With Helicobacter pylori. *Gastroenterology* 137, 1061-1071
7. An, H. J., Kronewitter, S. R., de Leoz, M. L., and Lebrilla, C. B. (2009) Glycomics and disease markers. *Curr Opin Chem Biol*
8. Fuster, M. M., and Esko, J. D. (2005) The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nature Reviews Cancer* 5, 526-542
9. Turnbull, J. E., and Field, R. A. (2007) Emerging glycomics technologies. *Nature Chemical Biology* 3, 74-77
10. La Belle, J. T., Gerlach, J. Q., Svarovsky, S., and Joshi, L. (2007) Label-free impedimetric detection of glycan-lectin interactions. *Anal Chem* 79, 6959-6964
11. Ito, H., Kuno, A., Sawaki, H., Sogabe, M., Ozaki, H., Tanaka, Y., Mizokami, M., Shoda, J., Angata, T., Sato, T., Hirabayashi, J., Ikehara, Y., and Narimatsu, H. (2009) Strategy for glycoproteomics: identification of glyco-alteration using multiple glycan profiling tools. *J Proteome Res* 8, 1358-1367
12. Wearne, K. A., Winter, H. C., O'Shea, K., and Goldstein, I. J. (2006) Use of lectins for probing differentiated human embryonic stem cells for carbohydrates. *Glycobiology* 16, 981-990
13. Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates, J. R., 3rd (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 21, 532-538
14. Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., and Watts, J. D. (2009) Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology* 27, 378-386
15. Luchansky, S. J., Hang, H. C., Saxon, E., Grunwell, J. R., Danielle, C. Y., Dube, D. H., and Bertozzi, C. R. (2003) Constructing azide-labeled cell surfaces using polysaccharide biosynthetic pathways. *Recognition of Carbohydrates in Biological Systems Pt A: General Procedures* 362, 249-272
16. Cordwell, S. J., and Thingholm, T. E. (2010) Technologies for plasma membrane proteomics. *Proteomics* 10, 611-627
17. Lovell-Badge, R. (2001) The future for stem cell research. *Nature* 414, 88-91
18. Nagano, K., Yoshida, Y., and Isobe, T. (2008) Cell surface biomarkers of embryonic stem cells. *Proteomics* 8, 4025-4035
19. McNeish, J. (2004) Embryonic stem cells in drug discovery. *Nature Reviews Drug Discovery* 3, 70-80
20. Lanctot, P. M., Gage, F. H., and Varki, A. P. (2007) The glycans of stem cells. *Curr Opin Chem Biol* 11, 373-380

21. Solter, D., and Knowles, B. B. (1978) Monoclonal Antibody Defining a Stage-Specific Mouse Embryonic Antigen (Ssea-1). *Proceedings of the National Academy of Sciences of the United States of America* 75, 5565-5569
22. Gooi, H. C., Feizi, T., Kapadia, A., Knowles, B. B., Solter, D., and Evans, M. J. (1981) Stage-Specific Embryonic Antigen Involves Alpha-1-]3 Fucosylated Type-2 Blood-Group Chains. *Nature* 292, 156-158
23. Kannagi, R., Cochran, N. A., Ishigami, F., Hakomori, S., Andrews, P. W., Knowles, B. B., and Solter, D. (1983) Stage-Specific Embryonic Antigens (Ssea-3 and Ssea-4) Are Epitopes of a Unique Globo-Series Ganglioside Isolated from Human Teratocarcinoma Cells. *Embo J* 2, 2355-2361
24. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145-1147
25. Satomaa, T., Heiskanen, A., Mikkola, M., Olsson, C., Blomqvist, M., Tiittanen, M., Jaatinen, T., Aitio, O., Olonen, A., Helin, J., Hiltunen, J., Natunen, J., Tuuri, T., Otonkoski, T., Saarinen, J., and Laine, J. (2009) The N-glycome of human embryonic stem cells. *BMC Cell Biol* 10, 42
26. An, H. J., Tillinghast, J. S., Woodruff, D. L., Rocke, D. M., and Lebrilla, C. B. (2006) A new computer program (GlycoX) to determine simultaneously the glycosylation sites and oligosaccharide heterogeneity of glycoproteins. *J Proteome Res* 5, 2800-2808
27. Kronewitter, S. R., An, H. J., de Leoz, M. L., Lebrilla, C. B., Miyamoto, S., and Leiserowitz, G. S. (2009) The development of retrosynthetic glycan libraries to profile and classify the human serum N-linked glycome. *Proteomics* 9, 2986-2994
28. Tusnady, G. E., and Simon, I. (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* 283, 489-506
29. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6, 175-182
30. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44-57
31. Pfeiffer, R., Rossier, G., Spindler, B., Meier, C., Kuhn, L., and Verrey, F. (1999) Amino acid transport of y+L-type by heterodimers of 4F2hc/CD98 and members of the glycoprotein-associated amino acid transporter family. *EMBO J* 18, 49-57
32. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21, 660-666
33. Kayano, T., Fukumoto, H., Eddy, R. L., Fan, Y. S., Byers, M. G., Shows, T. B., and Bell, G. I. (1988) Evidence for a family of human glucose transporter-like proteins. Sequence and gene localization of a protein expressed in fetal skeletal muscle and other tissues. *J Biol Chem* 263, 15245-15248
34. Brockhausen, I. (1997) Biosynthesis and functions of O-glycans and regulation of mucin antigen expression in cancer. *Biochem Soc Trans* 25, 871-874
35. Nichols, W. W., Murphy, D. G., Cristofalo, V. J., Toji, L. H., Greene, A. E., and Dwight, S. A. (1977) Characterization of a new human diploid cell strain, IMR-90. *Science* 196, 60-63
36. Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Jr., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F., and Brooks, S. C. (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res* 50, 6075-6086
37. Harvey, D. J. (1993) Quantitative Aspects of the Matrix-Assisted Laser-Desorption Mass-Spectrometry of Complex Oligosaccharides. *Rapid Communications in Mass Spectrometry* 7, 614-619

38. Grey, C., Edebrink, P., Krook, M., and Jacobsson, S. P. (2009) Development of a high performance anion exchange chromatography analysis for mapping of oligosaccharides. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 877, 1827-1832
39. Ninonuevo, M., An, H., Yin, H., Killeen, K., Grimm, R., Ward, R., German, B., and Lebrilla, C. (2005) Nanoliquid chromatography-mass spectrometry of oligosaccharides employing graphitized carbon chromatography on microchip with a high-accuracy mass analyzer. *Electrophoresis* 26, 3641-3649
40. Chu, C. S., Ninonuevo, M. R., Clowers, B. H., Perkins, P. D., An, H. J., Yin, H. F., Killeen, K., Miyamoto, S., Grimm, R., and Lebrilla, C. B. (2009) Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. *Proteomics* 9, 1939-1951
41. Tao, S. C., Li, Y., Zhou, J. B., Qian, J., Schnaar, R. L., Zhang, Y., Goldstein, I. J., Zhu, H., and Schneck, J. P. (2008) Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* 18, 761-769
42. Morishima, S., Morita, I., Tokushima, T., Kawashima, H., Miyasaka, M., Omura, K., and Murota, S. (2003) Expression and role of mannose receptor/terminal high-mannose type oligosaccharide on osteoclast precursors during osteoclast formation. *Journal of Endocrinology* 176, 285-292
43. Molinari, M. (2007) N-glycan structure dictates extension of protein folding or onset of disposal. *Nature Chemical Biology* 3, 313-320
44. Kurachi, T., Morita, I., Oki, T., Ueki, T., Sakaguchi, K., Enomoto, S., and Murota, S. (1994) Expression on Outer Membranes of Mannose Residues, Which Are Involved in Osteoclast Formation Via Cellular Fusion Events. *Journal of Biological Chemistry* 269, 17572-17576
45. Johns, T. G., Mellman, I., Cartwright, G. A., Ritter, G., Old, L. J., Burgess, A. W., and Scott, A. M. (2005) 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 Journal* 19, 780-+
46. de Leoz, M. L., Young, L. J., An, H. J., Kronewitter, S. R., Kim, J., Miyamoto, S., Borowsky, A. D., Chew, H. K., and Lebrilla, C. B. (2011) High-mannose glycans are elevated during breast cancer progression. *Mol Cell Proteomics* 10, M110 002717
47. Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001) Stem cells, cancer, and cancer stem cells. *Nature* 414, 105-111

Figure Legends.

Figure 1. Representative MALDI-FTICR mass spectra of N-glycans found in hESC membranes in the positive detection ion mode. Glycans eluted from SPEfractions (10%, 20%, and 40% AcN) were combined prior to MS analysis. Glycan mass profile of (A) HSF-6 hESC and (B) H1 hESC membrane.Structures are putative and are based on accurate masses and tandem mass spectrometry.

Figure 2. Quantitation and isomer separation of N-glycans on cell membranes.The combined SPE fraction (10%, 20%, and 40%) was analyzed by nanoLC-Chip/TOF. Representative base peak chromatogram of N-glycans found on four different cell membranes: (A) HSF-6 hESC, (B) H1 hESC, (C) IMR-90 fibroblast, and (D) normal breast cell line, MCF10A. Proposed structures correspond to the most abundant glycan at that retention time are provided. An asterisk (*) represents non-glycan peaks.

Figure 3. Extracted ion chromatogram of high mannose type glycans found in HSF-6 hESC. Left panel shows $GlcNAc_2Man_7$ (Man7) isomers and right panel shows $GlcNAc_2Man_8$ (Man8) isomers.

Figure 4. Glycan quantitation using peak areas of chromatograms from nanoLC-Chip/TOF MS. (A) Overall abundances based on N-glycan types. High mannose glycans are the most abundant in both H1 hESC and HSF-6 hESC membranes while complex/hybrid glycans are in abundance in IMR-90 fibroblast and normal breast cell line, MCF 10A. (B) Glycan abundance profile of high mannose and complex/hybrid type glycans in H1 hESC and HSF-6 hESC membranes, respectively. Man8 and Man9 are the most abundant glycans in both hESC. The complex/hybrid type glycans were further sorted into four groups based on glycan composition and the number of fucose (n=1-3). Sialic acid represents sialylated (but non-fucoyslated) glycans and Hex:HexNAc represents non-sialylated and non-fucosylated glycans. The complex/hybrid type glycans are generally fucosylated in both hESC membranes.

Figure 5. Validation of glycan expression using lectins. (A)-(B) are fixed hESC colonies that were stained with Hoescht dye (nuclear stain), antibody against SSEA-4 with secondary antibody (depicted by AlexaFluor 594) and Con A-FITC and GNA-FITC (20 μg/mL). All colonies were positively stained with Hoescht and expressed SSEA-4. (A) Cells bound to Con A compared to colonies that were labeled with Con A+I (inhibitory control). (B) GNA-FITC also bound to cells, compared to the inhibitory control (GNA+I). All images were taken with 10X magnification and the scale bar: 20 μm. (C)-(D), Flow cytometry was used to quantify amount of lectin binding on live hESCs. (C)This dot plot shows the populations of cells (black) that were double-labeled with both lectin (Con A-FITC) and SSEA-4 positive (APC) compared to the unstained population (grey) and inhibitory control (blue). (D) Thesehistograms represent the distribution of live hESCs depicted in (C) that were stained with Con A-FITC and GNA- FITC (40 μg/mL) compared to unstained (grey fill) and inhibitory controls (blue).

Figure 1

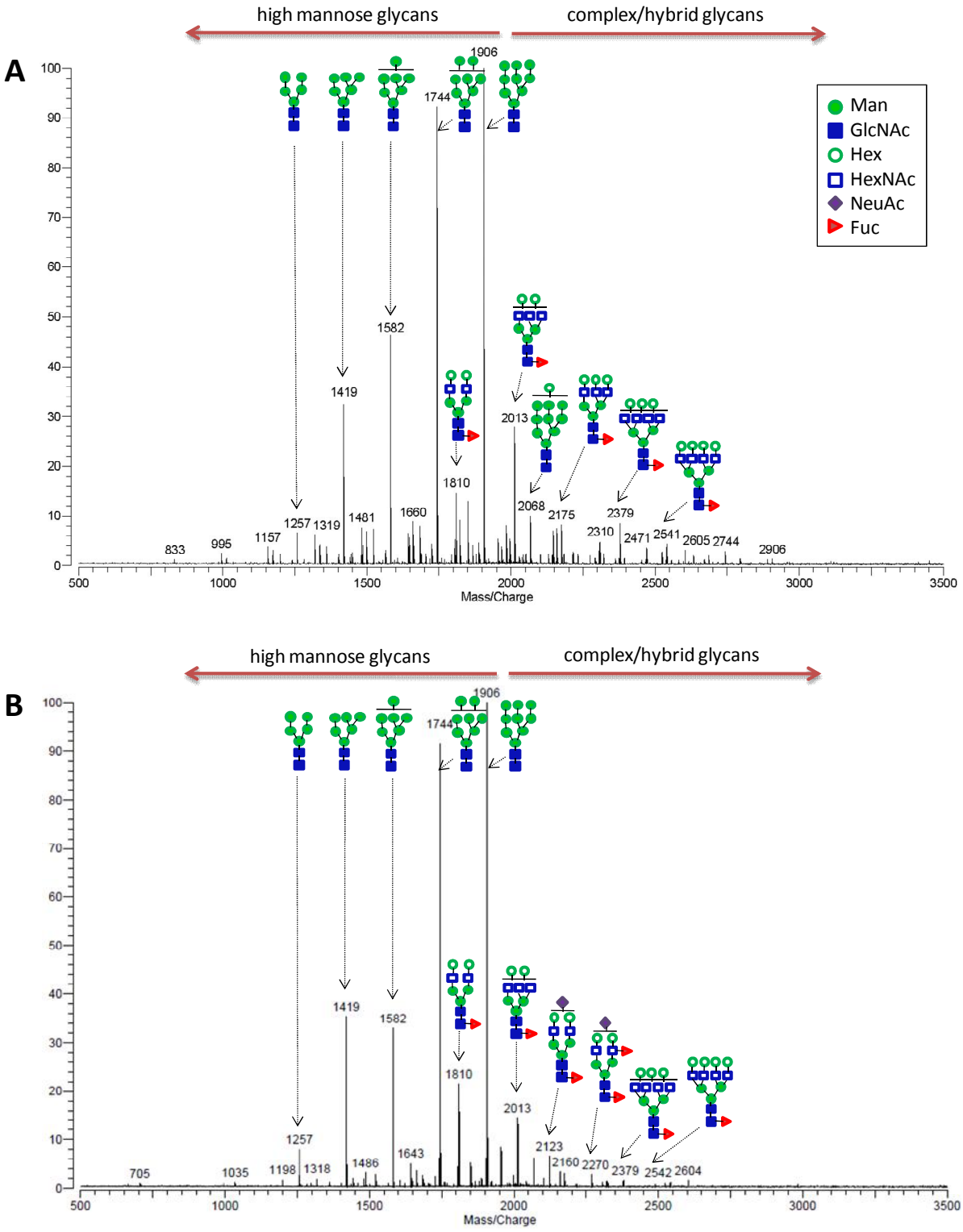

A
high mannose glycans
complex/hybrid glycans
Man
GlcNAc
Hex
HexNAc
NeuAc
Fuc
Mass/Charge
B
high mannose glycans
complex/hybrid glycans
Mass/Charge

Figure 2

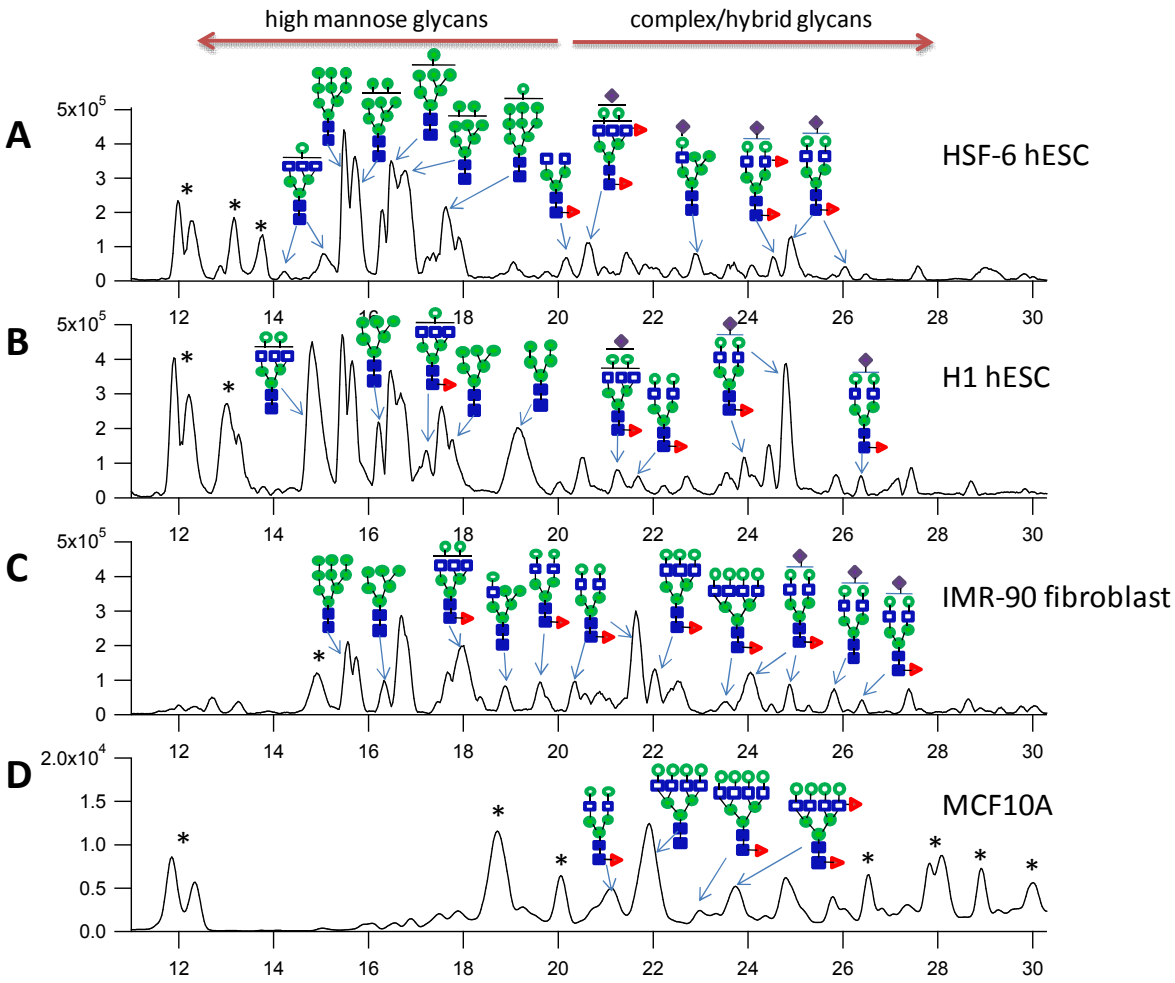

high mannose glycans
complex/hybrid glycans
A
HSF-6 hESC
B
H1 hESC
C
IMR-90 fibroblast
D
MCF10A
5x10^5
2.0x10^4

Figure 3.

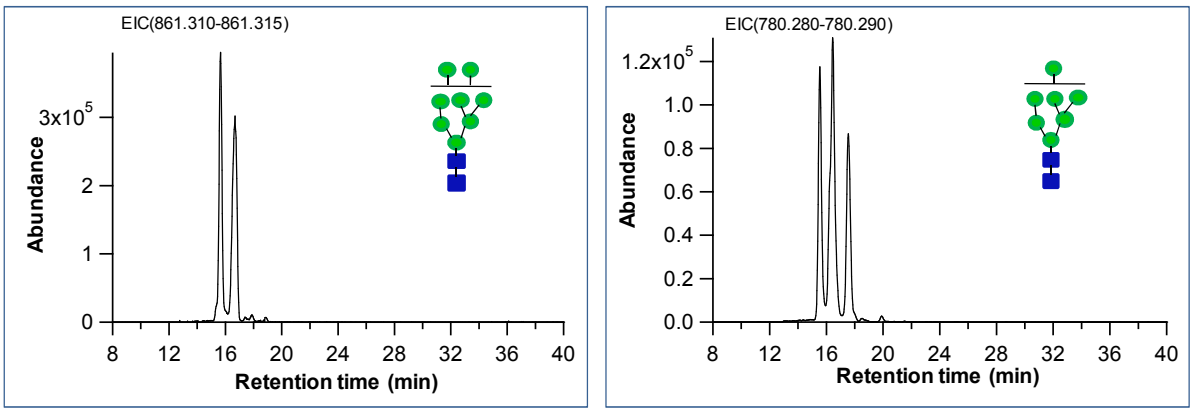

EIC(861.310-861.315)
Abundance
3x10^5
2
1
0
8
12
16
20
24
28
32
36
40
Retention time (min)
EIC(780.280-780.290)
Abundance
1.2x10^5
1.0
0.8
0.6
0.4
0.2
0.0
8
12
16
20
24
28
32
36
40
Retention time (min)

Figure 4

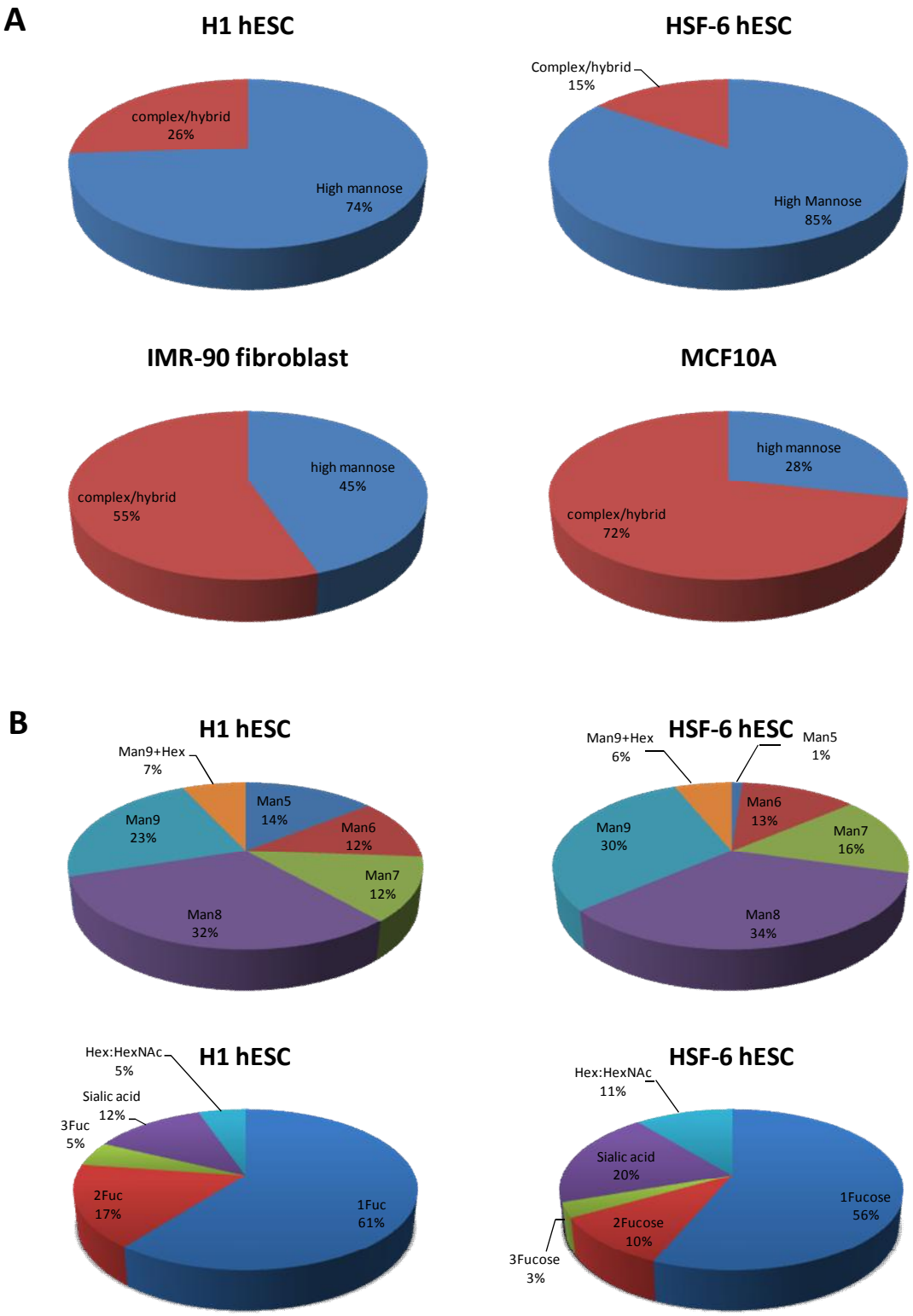
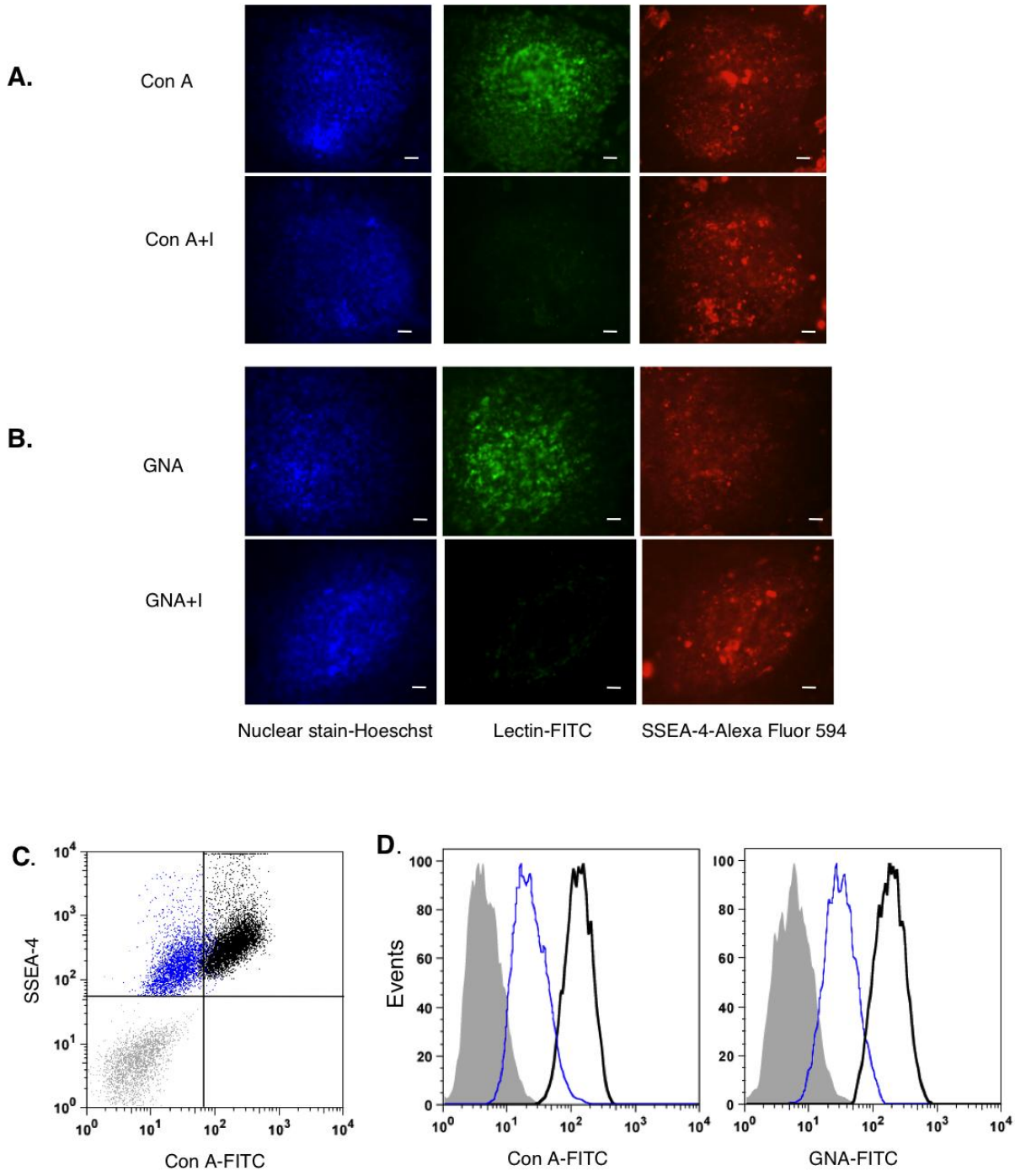

A
H1 hESC
complex/hybrid 26%
High mannose 74%
HSF-6 hESC
Complex/hybrid 15%
High Mannose 85%
IMR-90 fibroblast
high mannose 45%
complex/hybrid 55%
MCF10A
high mannose 28%
complex/hybrid 72%
B
H1 hESC
Man9+Hex 7%
Man5 14%
Man6 12%
Man7 12%
Man8 32%
Man9 23%
HSF-6 hESC
Man9+Hex 6%
Man5 1%
Man6 13%
Man7 16%
Man8 34%
Man9 30%
H1 hESC
Hex:HexNAc 5%
Sialic acid 12%
3Fuc 5%
2Fuc 17%
1Fuc 61%
HSF-6 hESC
Hex:HexNAc 11%
Sialic acid 20%
1Fucose 56%
2Fucose 10%
3Fucose 3%

Figure 5


A.
Con A
Con A+I
B.
GNA
GNA+I
Nuclear stain-Hoeschst
Lectin-FITC
SSEA-4-Alexa Fluor 594
C.
SSEA-4
Con A-FITC
D.
Events
Con A-FITC
GNA-FITC