Title

O-linked α2,3 sialylation defines stem cell populations in breast cancer.

Journal

Science Advances, 8(1)

Authors

Walker, Melanie
Goel, Hira
Mukhopadhyay, Dimpi
et al.

Publication Date

2022-01-07

DOI

10.1126/sciadv.abj9513

CANCER

O-linked α2,3 sialylation defines stem cell populations in breast cancer

Melanie R. Walker[1], Hira Lal Goel[1], Dimpi Mukhopadhyay[1], Peter Chhoy[1], Emmet R. Karner[1], Jennifer L. Clark[2], Haibo Liu[1], Rui Li[1], Julie Lihua Zhu[1], Shuhui Chen[3], Lara K. Mahal[3,4], Barbara A. Bensing[5], Arthur M. Mercurio[1]*

We pursued the hypothesis that specific glycans can be used to distinguish breast cancer stem cells (CSCs) and influence their function. Comparison of CSCs and non-CSCs from multiple breast cancer models revealed that CSCs are distinguished by expression of α2,3 sialylated core2 O-linked glycans. We identified a lectin, SLBR-N, which binds to O-linked α2,3 sialic acids, that was able to enrich for CSCs in vitro and in vivo. This O-glycan is expressed on CD44 and promotes its interaction with hyaluronic acid, facilitating CD44 signaling and CSC properties. In contrast, FUT3, which contributes to sialyl Lewis X (sLeX) production, is preferentially expressed in the non-CSC population, and it antagonizes CSC function. Collectively, our data indicate that SLBR-N can be more efficient at enriching for CSCs than CD44 itself because its use avoids the issues of CD44 splicing and glycan status. These data also reveal how differential glycosylation influences CSC fate.

INTRODUCTION

Solid tumors such as breast and other carcinomas are characterized by extensive heterogeneity in the cell populations that comprise them (*1*, *2*). This heterogeneity exists not only between tumor and stromal cells but also within tumor cells themselves, a finding that has been made apparent by recent single-cell RNA sequencing (RNA-seq) studies on many different carcinomas (*3*–*5*). The concept of intratumor heterogeneity is highly relevant because distinct tumor cell populations differ in many aspects including their differentiation status, response to therapy, and ability to initiate a recurrent tumor (*6*). In this context, considerable evidence supports the hypothesis that distinct populations of tumor cells with the ability to resist therapy and initiate new tumors are characterized by properties associated with stem cells (*6*, *7*). For this reason, they are often referred to as tumor-initiating cells or cancer stem cells (CSCs). Although the term "CSC" has been used to describe a range of cell populations that share similar properties such as the ability to form spheroids in three-dimensional culture and to resist chemo- and radiation therapy, it is evident that cells of this nature are enriched in more aggressive tumors and contribute to their aggressive behavior (*7*). On the basis of this information, there is intense interest in identifying specific markers that distinguish CSC populations from their less aggressive counterparts and to exploit such markers for therapeutic targeting.

We hypothesized that the cell surface glycome provides a rich resource for the identification of specific glycans that distinguish CSCs from other tumor cells. Cell surface glycans are highly complex, with various cores and termini regulated by multiple enzymes. Moreover, different sets of structures can be displayed on protein N-linked and O-linked glycans, as well as glycolipids. There is a literature that spans decades on glycosylation differences between normal and transformed cells as highlighted by the more recent finding that the carbohydrate antigen CA19-9 [sialyl Lewis A (sLeA)] is highly expressed in aggressive pancreatic tumors with minimal expression in normal pancreatic tissue (*8*, *9*). Moreover, glycans can be effective therapeutic targets. For example, targeting a glycosylated form of the programmed death-ligand 1 (PD-L1) with drug-conjugated antibodies in breast cancer promoted immune reactivation and a "bystander-killing effect" with minimal toxicity to normal tissues (*10*). Although previous studies have investigated glycosyltransferases expressed in CSCs (*11*), glycans are not used routinely as markers to enrich for CSCs in heterogeneous tumor populations, and relatively little is known about how specific glycans promote CSC properties. Thus, it is essential to identify the unique and targetable glycans and glycosylated proteins expressed in CSC populations and to assess how protein glycoforms regulate CSC functions. To achieve this goal, we used unbiased approaches to identify specific glycans that distinguish breast cancer cells with CSC properties from other tumor cells, and we investigated how these glycans affect CSC function.

RESULTS

α2,3 sialylation characterizes CSC-enriched populations of breast tumor cells

Initially, we sought to identify specific glycans that distinguish populations of breast cancer cells enriched for CSC properties from the bulk population. For this purpose, we used human mammary luminal epithelial (HMLE) cells generated by the Weinberg laboratory that had been transformed with H-Ras (HMLER) (*12*). A previously published study from our laboratory had shown that HMLER cells could be sorted on the basis of the expression of CD104 and CD24 into a distinct population that is mesenchymal and enriched for CSC properties ($CD104^{low}/CD24^{low}$) and a more epithelial population that has diminished CSC properties ($CD104^{high}/CD24^{high}$) (*13*). In this context, we define CSC properties as the ability to form mammospheres and resist chemotherapy as demonstrated in the previous publication. These two distinct populations referred to as CSC-enriched (CSC) and non–CSC-enriched (non-CSC) were compared

[1]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA. [2]Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA. [3]Biomedical Research Institute, Department of Chemistry, New York University, New York, NY, USA. [4]Department of Chemistry, University of Alberta, Edmonton, AB, Canada. [5]Department of Medicine, The San Francisco Veterans Affairs Medical Center, and the University of California, San Francisco, San Francisco, CA, USA.
*Corresponding author. Email: arthur.mercurio@umassmed.edu

for potential differences in glycosylation using an unbiased approach that involved lectin microarray analysis and RNA-seq and subsequent validation of these data in vitro and in vivo including glycoprotein identification and functional analysis (Fig. 1A).

We analyzed the glycomes of the CSC and non-CSC HMLER populations using our dual-color lectin microarray technology (*14*). In brief, equal amounts of orthogonally labeled fluorescent samples and a mixed reference containing all samples were analyzed on the array, which contains >100 lectin probes. For more information, see tables S1 and S2. Although few significant differences between the two populations were observed, a subtle but clear difference was seen in O-linked α2,3 sialic acids [lectins: SK678, SLBR-H (n.s., not significant), and SLBR-N (n.s.)], which were enriched in CSC cells, and N-linked α2,6 sialic acids, which were enriched in non-CSC cells (Fig. 1B and fig. S1B; for complete heatmap, see fig. S1A). Subsequently, RNA-seq was performed on the same populations to assess potential differences in glycosyltransferase expression (Fig. 1C). We observed that the CSC-enriched population displayed increased expression of the core2 β-1,6-*N*-acetylglucosaminyltransferase-1 (GCNT1) and α2,3 sialyltransferases [ST3GAL2, ST3GAL5, and ST3GAL6 (n.s.)] and decreased expression of the α2,6 sialyltransferases (ST6GAL1) compared with the non–CSC-enriched population. The differential expression of these glycosyltransferases correlated with the lectin microarray data, explaining the alterations in glycan structures observed between the CSC-enriched and non–CSC-enriched populations. We also observed that the α1,3 fucosyltransferase FUT3, which

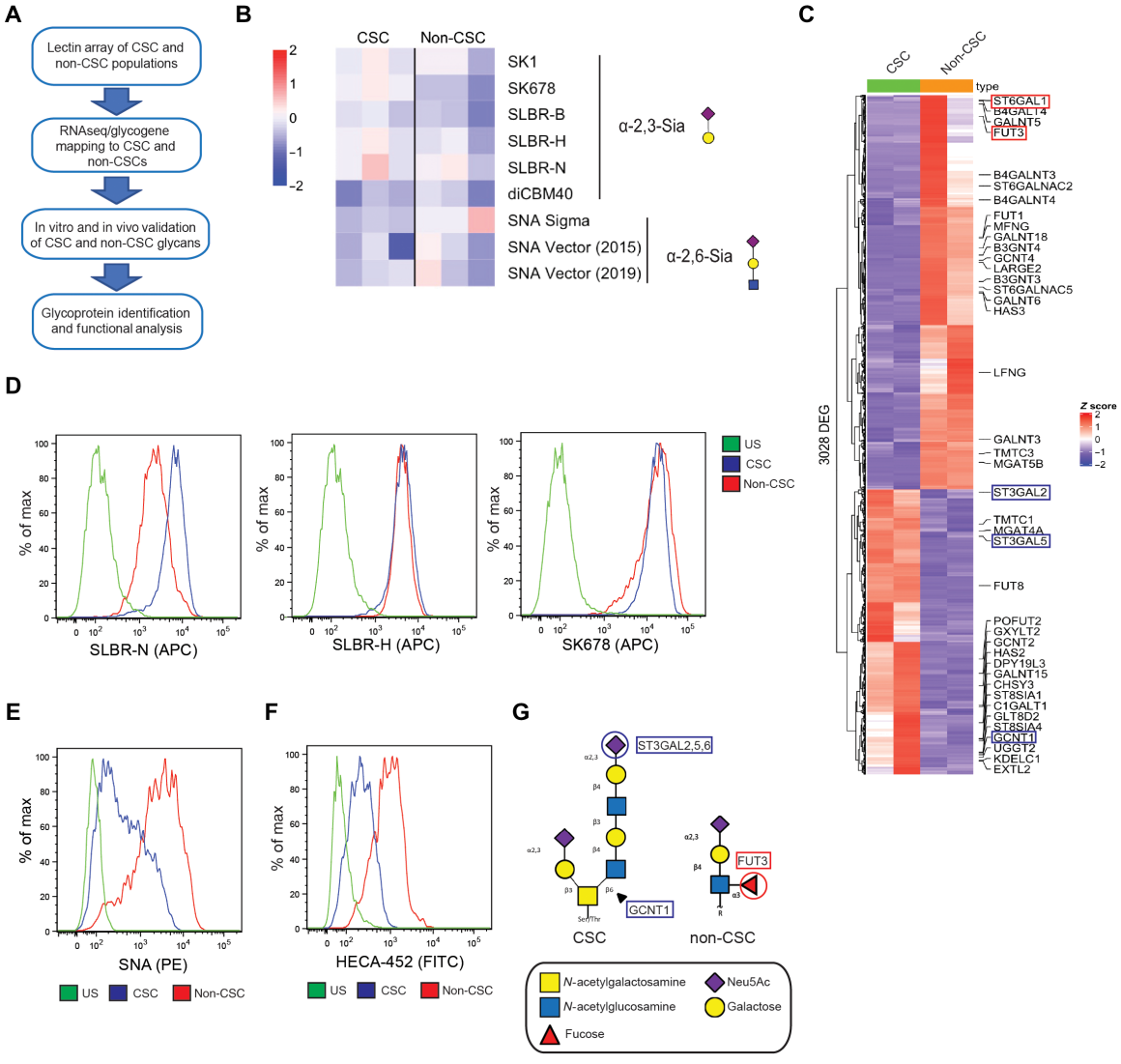


Fig. 1. CSCs and non-CSCs are distinguished by their glycomes. (**A**) Schematic of our unbiased approach to identify unique glycans and glycosyltransferases expressed in the CSC population. (**B**) CSC and non-CSC cells were subject to dual-color lectin microarray. Shown are the binding affinities of the CSC and non-CSC extracts to lectins identifying α2,3 sialoglycans and α2,6 sialoglycans. Shown are the results of three replicates. (**C**) Heatmap of differentially expressed glycosyltransferases expressed in the CSC and non-CSC populations; FDR < 0.05 and |log2FC| > 1. Glycosyltransferases corresponding to core2 α2,3 sialoglycans are highlighted in blue, and α2,6 sialoglycans are highlighted in red. CSC and non-CSC cells were subjected to flow cytometric analysis for glycan expression. Populations were stained for (**D**) SLBR-N, SLBR-H, SK678, (**E**) SNA, and (**F**) HECA-452. US, unstained control. Representative results of three independent experiments. (**G**) Structures of the two target glycans expressed in the CSC and non-CSC populations. CSCs express the core2 α2,3 sialoglycan initiated by GCNT1 and sialylated by the ST3GAL enzymes. Non-CSCs express N-linked sLeX driven by FUT3.

is linked to the sialyl Lewis X (sLeX) N-glycan, was preferentially expressed in the non–CSC-enriched population.

The lectin microarray and RNA-seq data provided key information that enabled us to investigate the putative glycan differences in more detail. Specifically, we made use of the lectins identified by the lectin microarray to characterize sialylation differences between the CSC- and non–CSC-enriched HMLER populations by flow cytometry and observed the most significant difference with SLBR-N, which is derived from the bacterium *Streptococcus gordonii* and binds preferentially to core2 O-linked α2,3 sialoglycans on glycoproteins (*15*). These glycans are initiated by the glycosyltransferase GCNT1 and are further modified by other glycosyltranferases including terminal α2,3 sialylation by ST3GAL2, ST3GAL5, and ST3GAL6, which we confirmed to be enriched in the CSC population by quantitative polymerase chain reaction (qPCR; fig. S1C). Notably, SLBR-N exhibited significantly higher binding to the CSC-enriched population compared with the non–CSC-enriched population (Fig. 1D). In contrast, no such difference was observed with SLBR-H, which prefers sialylated core1 and core2 O-glycans. Also, the SK678 lectin, which recognizes other core2 α2,3 sialoglycan structures, did not select for the CSC-enriched population, although it showed significant binding in the lectin microarray (*15*). The SNA (Sambucus Nigra Lectin) lectin, which detects α2,6 sialic acids produced by ST6GAL1, exhibited preferential binding to the non–CSC-enriched population (Fig. 1E and fig. S1C), consistent with our glycomic and transcriptomic datasets. Our RNA-seq data on FUT3 prompted us to compare the expression of sLeX between the two populations. Although a specific lectin that detects α1,3 fucose is not available, we used the HECA-452 antibody for flow cytometry because it is specific for N-linked sLeX, which is fucosylated by FUT3. HECA-452 bound much more significantly to the non–CSC-enriched populations, and FUT3 expression was enhanced in the non-CSC population as assessed by qPCR (Fig. 1F and fig. S1C). Last, we thought it was important to compare the expression of sLeA, also known as CA19-9, between the two populations because this glycan has been implicated in highly aggressive pancreatic cancer (*9*). The breast cancer–derived CSC- and non–CSC-enriched populations did not significantly differ in their expression of CA19-9, suggesting the glycome patterns of the breast and pancreatic CSCs are different (fig. S1D). Together, our unbiased approaches reveal that the CSC- and non–CSC-enriched populations exhibit significant differences in glycosylation. Specifically, the non-CSC population is distinguished by higher levels of α2,6 sialosides and the expression of sLeX (Fig. 1G). In contrast, the CSC-enriched population is characterized by increased expression of a core2 O-linked α2,3 sialoglycan detected by the SLBR-N lectin.

The SLBR-N lectin enriches for cells with stem cell properties in vitro and in vivo

To assess whether the SLBR-N lectin can be used to enrich for CSCs in heterogeneous cell populations, we sorted HMLER cells and another breast cancer cell line, HCC1806, using this lectin and obtained distinct SLBR-N^{high} and SLBR-N^{low} populations (fig. S2, A and C). We observed that the SLBR-N^{high} cells had enhanced self-renewal capacity as assessed by serial mammosphere formation compared with the SLBR-N^{low} cells in both cell lines (Fig. 2, A and D). Subsequently, these populations were analyzed by flow cytometry for surface expression of CD44 and CD24, whose differential expression is commonly used to identify CSC-like breast tumor populations (CD44^{high}/CD24^{low}) (*16*). HMLER SLBR-N^{high} and HCC1806 SLBR-N^{high} cells stained for a CD44^{high}/CD24^{low} population compared with the SLBR-N^{low} cells (Fig. 2, B and E). The SLBR-N^{high} cells exhibited a morphology and gene expression pattern characteristic of mesenchymal cells, compared with the SLBR-N^{low} cells that had an epithelial phenotype (Fig. 2, C and F, and fig. S2E). We also confirmed by qPCR that the genes GCNT1, ST3GAL2, ST3GAL5, and ST3GAL6 were enriched in the SLBR-N^{high} cells (fig. S2, B and D). This finding that SLBR-N can enrich for CSCs in cancer cell lines was substantiated by flow cytometry analysis of a patient-derived xenograft (PDX). The SLBR-N^{high} cells, which comprised 8.58% of the total population, exclusively stained for cells in the CD44^{high}/CD24^{low} population (Fig. 2G). In contrast, SLBR-H and SK678, which bind to other core1 and core2 α2,3 sialoglycan structures, did not select the CD44^{high}/CD24^{low} cell population as efficiently as SLBR-N (fig. S2F).

The results described above prompted us to evaluate a causal role for the SLBR-N glycan ligand in promoting CSC properties. We used a variant of MDA-MB-231 cells, termed TE3 cells, that were selected for their CSC-like properties (*17*). We observed by SLBR-N staining that the TE3 variant was enriched in core2 O-linked α2,3 sialoglycans compared with the MDA-MB-231 cells (Fig. 2H). We confirmed that SLBR-N identifies an O-linked α2,3 sialoglycan by treating TE3 cells with neuraminidase S, an α2,3-specific sialidase, and PNGaseF, which removes N-linked glycans (Fig. 2, I and J). Diminishing the expression of a single ST3Gal enzyme did not significantly alter the expression of the SLBR-N ligand (fig. S2, G and H), however, most likely because of the compensatory nature of the α2,3 sialyltransferases (*18*). To inhibit the synthesis of the SLBR-N ligand, we knocked down expression of GCNT1, the enzyme responsible for the formation of core2 glycans, and observed significantly reduced SLBR-N binding to the cell surface (Fig. 2, K and L) and reduced self-renewal capacity compared with control cells (Fig. 2M). Knockdown of GCNT1 did not diminish the ability of SLBR-H to bind to the CSCs, because it also can bind to core1 structures, but it did diminish SK678 binding (fig. S2I).

On the basis of the above findings, we sought to obtain additional data to link SLBR-N^{high} cells to a CSC phenotype including tumorigenicity and chemoresistance. Using organoids from a patient with triple-negative breast cancer (TNBC), we sorted SLBR-N^{high} and SLBR-N^{low} cell populations and assessed glycogene expression and mammosphere formation (fig. S3A). We found that the TNBC SLBR-N^{high} cells had enhanced GCNT1/ST3GAL expression and higher mammosphere-forming ability compared with the SLBR-N^{low} cells (fig. S3, B and C). To determine whether the SLBR-N lectin selects for cells that exhibit resistance to chemotherapy, we treated the SLBR-N^{high} and SLBR-N^{low} cells with 5-fluorouracil (5-FU). We observed that SLBR-N^{high} cells exhibited significantly more resistance to increasing doses of 5-FU compared with the SLBR-N^{low} cells (Fig. 3A and fig. S3D). Subsequently, we generated patient-derived organoids from a patient with TNBC that were resistant to 5-FU and cisplatin and observed that these chemoresistant organoids had significantly more SLBR-N staining compared with the chemosensitive organoids from the same patient (Fig. 3B and fig. S3, E to G), indicating that chemotherapy selects for the survival of SLBR-N^{high} cells. We also analyzed patient-derived organoids from an estrogen receptor–positive (ER$^+$) tumor and observed that fewer than 1% of the total population of cells were positive for SLBR-N (fig. S3H). Given that ER$^+$ breast cancer harbors a lower frequency of CSCs cells than does TNBC, this result suggests that CSCs from these different

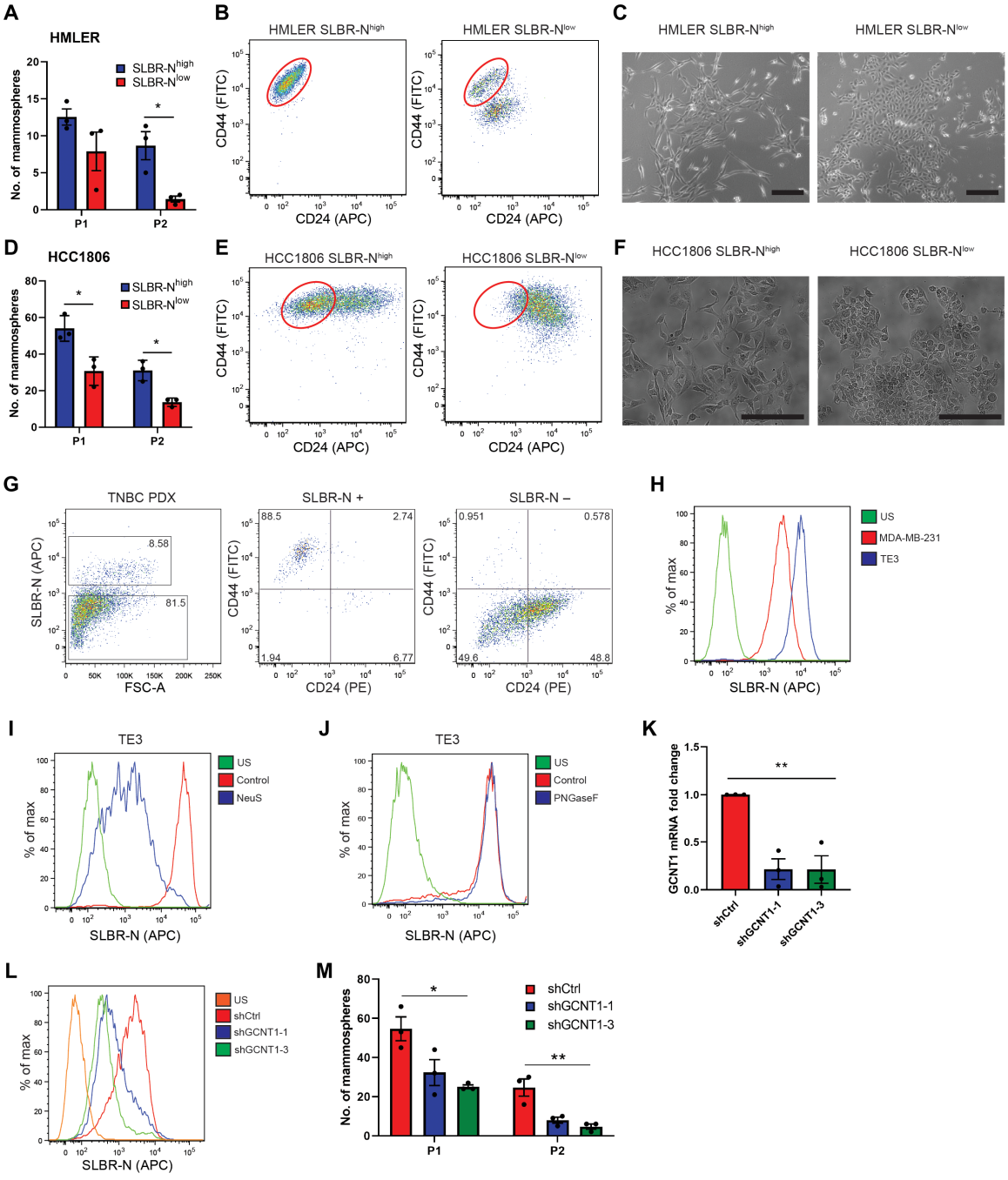
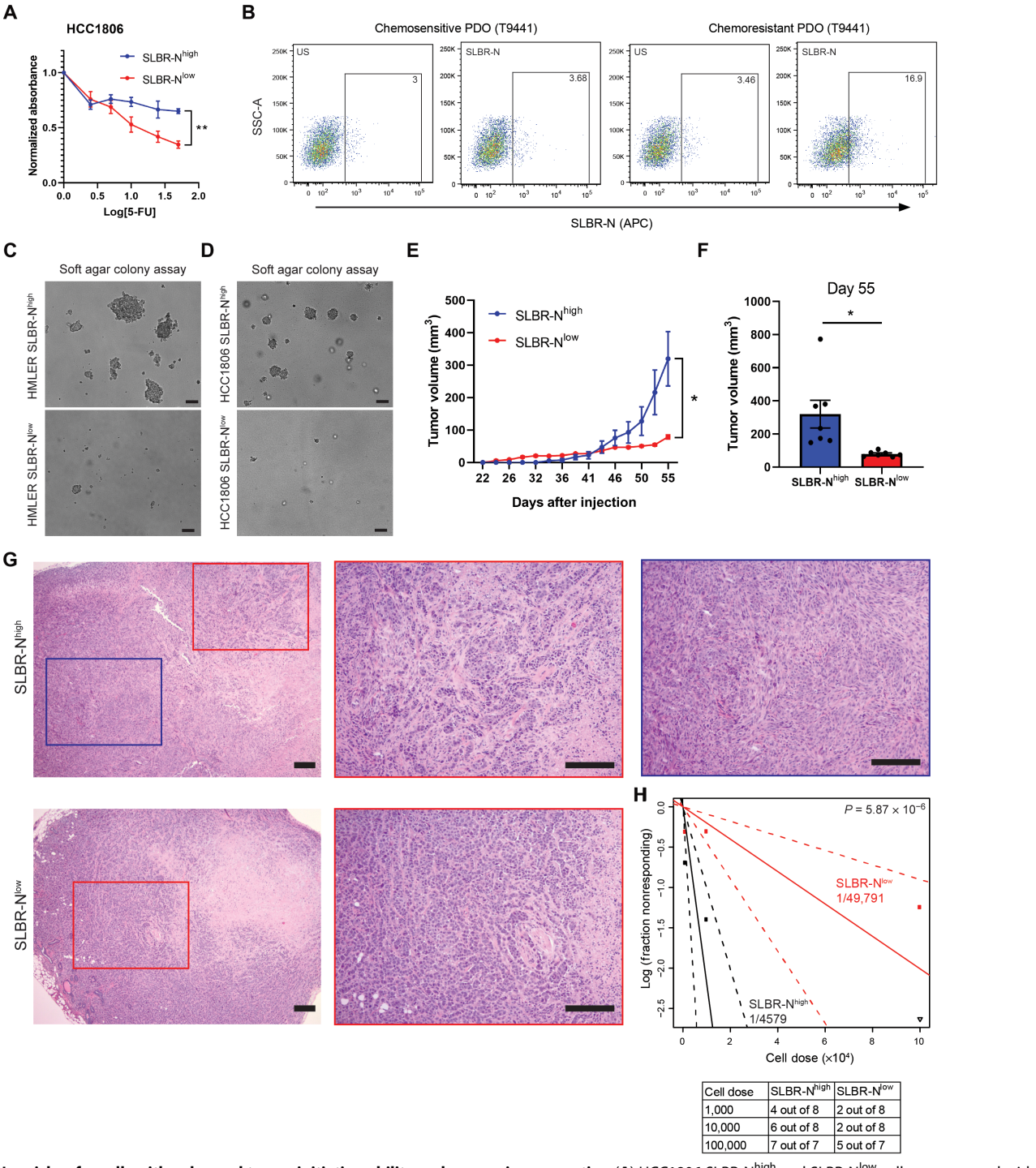


Fig. 2. SLBR-N identifies CSCs in vitro and in vivo. HMLER SLBR-N^{high} and SLBR-N^{low} cell populations were assessed for (**A**) self-renewal by serial passage mammosphere formation (P1, passage 1; P2, passage 2), (**B**) CD44/CD24 surface expression by flow cytometry, and (**C**) morphological changes by phase-contrast imaging (scale bars, 200 μm). Representative results of three independent experiments. HCC1806 SLBR-N^{high} and SLBR-N^{low} cell populations were assessed for (**D**) self-renewal by serial passage mammosphere formation, (**E**) CD44/CD24 surface expression by flow cytometry, and (**F**) morphological changes by phase-contrast imaging (scale bars, 200 μm). Representative results of three independent experiments. (**G**) PDX TNBC cells were analyzed for SLBR-N binding and CD44/CD24 expression by flow cytometry. Shown is one biological replicate. (**H**) MDA-MB-231 cells and TE3 cells were assessed for the expression of the core2 α2,3 sialoglycan via SLBR-N. Representative results of three independent experiments. (**I**) TE3 cells were treated with neuraminidase S and analyzed for SLBR-N binding by flow cytometry. Representative results of three independent experiments. (**J**) TE3 cells were treated with PNGaseF and analyzed for SLBR-N binding by flow cytometry. Representative results of three independent experiments. (**K**) TE3 cells were stably transfected with control shRNA (shCtrl) or two shRNA clones for GCNT1. GCNT1 expression was quantified by qPCR. Representative results of three independent experiments. (**L**) TE3 shCtrl and shGCNT1 cells were assessed for SLBR-N binding. Representative results of three independent experiments. (**M**) TE3 shCtrl and shGCNT1 cells were analyzed by serial passage mammosphere formation. Representative results of three independent experiments. *$P < 0.05$, **$P < 0.01$.



Cell dose	SLBR-N^{high}	SLBR-N^{low}
1,000	4 out of 8	2 out of 8
10,000	6 out of 8	2 out of 8
100,000	7 out of 7	5 out of 7

Fig. 3. SLBR-N enriches for cells with enhanced tumor-initiating ability and aggressive properties. (**A**) HCC1806 SLBR-N^{high} and SLBR-N^{low} cells were treated with 5-FU for 96 hours, and percentage of surviving cells was quantified. Representative results of three independent experiments. (**B**) Chemosensitive and chemoresistant patient-derived organoids (PDO) from TNBC patient tumor (T9441) were dissociated and analyzed for SLBR-N. Shown is one biological replicate. (**C**) HMLER SLBR-N^{high} and SLBR-N^{low} cells were plated in soft agar to measure colony formation. Representative results of three independent experiments. Scale bars, 200 μm. (**D**) HCC1806 SLBR-N^{high} and SLBR-N^{low} cells were plated in soft agar to measure colony formation. Representative results of three independent experiments. Scale bars, 200 μm. (**E**) HCC1806 SLBR-N^{high} and SLBR-N^{low} cells (2.5×10^5 cells) were transplanted into mammary fat pads of NSG mice, and tumor volume (mm^3) was measured over 55 days. Shown are the mean tumor volume ± SEM of seven mice per group. (**F**) Volume of tumors (mm^3) from HCC1806 SLBR-N^{high} and SLBR-N^{low} cells measured on day 55 of the study. Shown are the mean tumor volume ± SEM of seven mice per group. (**G**) Hematoxylin and eosin sections of tumors from HCC1806 SLBR-N^{high} and SLBR-N^{low} cells. Red box, epithelioid; blue box, spindle cell. Scale bars, 200 μm. (**H**) HCC1806 SLBR-N^{high} and SLBR-N^{low} cells were transplanted into mammary fat pads of NSG mice using 10-fold serial dilution. Data are presented as a log-log plot, and frequency of stem cells is calculated by extreme limiting dilution analysis. Red, SLBR-N^{low} (1/49,791); black, SLBR-N^{high} (1/4579). *$P < 0.05$.

breast cancer subtypes can be defined by their expression of the α2,3 sialoglycan.

Soft agar colony formation assays revealed that the SLBR-N^{high} cells had enhanced tumorigenicity compared with the SLBR-N^{low} cells (Fig. 3, C and D). Subsequently, we injected HCC1806 cells sorted for SLBR-N expression into immunodeficient mice. SLBR-N^{high} cells formed larger tumors compared with the SLBR-N^{low} cells (Fig. 3, E and F). Histological examination of these tumors revealed that the SLBR-N^{low} cells produced high-grade tumors with typical features of epithelioid, invasive ductal carcinoma. However, the SLBR-N^{high} cells produced high-grade, heterogeneous tumors with the features of metaplastic carcinoma, specifically spindle cell carcinoma (Fig. 3G). These metaplastic, spindle cell carcinomas are highly aggressive, associated with an epithelial-to-mesenchymal transition (EMT) phenotype, and chemoresistant (*19*), properties associated with CSCs. The SLBR-N^{high} tumors also displayed regions with epithelioid morphology, suggesting that the SLBR-N^{high} cells can differentiate into different tumor cell populations.

To assess whether the SLBR-N^{high} cells harbor a higher percentage of CSCs compared with the SLBR-N^{low} cells, we performed a limiting dilution experiment in vivo. Using the extreme limiting dilution analysis, the frequency of CSCs was calculated to be approximately 1 of 4579 for HCC1806 SLBR-N^{high} and 1 of 49,791 for HCC1806 SLBR-N^{low} (Fig. 3H). These data demonstrate that SLBR-N selects for a population of cells that is enriched with stem cell properties.

CD44s expresses O-linked α2,3 sialoglycan

Given that SLBR-N enriches for breast CSCs, we were interested in identifying CSC glycoproteins that express the core2 α2,3 sialoglycan. Lectin blotting revealed distinct glycoproteins that were positive for SLBR-N in the CSC population that were far less apparent in the non-CSC population (Fig. 4A). We then performed SLBR-N affinity capture on CSC lysates and subjected the captured glycoproteins to mass spectrometry. One of the most abundant SLBR-N captured proteins was CD44 (table S3), a known breast CSC marker (*16*). We validated that SLBR-N captured CD44 by immunoblotting and confirmed that CD44 is highly expressed in the CSC population (Fig. 4, B and C). We found that CSCs express primarily CD44s and not other CD44 variant forms, in agreement with previous studies (*20*), and that only CD44s binds to the SLBR-N lectin (Fig. 4D). We also confirmed that SLBR-N recognizes O-linked glycans on CD44 by PNGaseF treatment, which removes N-linked glycans from proteins, which is evident in the apparent molecular mass change and retention of signal in Fig. 4E. These results were substantiated by the finding that SLBR-N^{high} HCC1806 cells exclusively express CD44s, which is expressed at very low levels in the SLBR-N^{low} population (Fig. 4F). In a recent collaborative study, we demonstrated that the CD44s variant promotes the CSC properties in breast cancer by activation of the platelet-derived growth factor receptor β (PDGFRβ)/signal transducer and activator of transcription 3 (STAT3) signaling pathway (*20*). Consistent with this observation, we observed that the SLBR-N^{high} population expresses PDGFRβ and can activate STAT3 signaling in response to PDGF stimulation, which is not active in the SLBR-N^{low} population (Fig. 4G).

α2,3 sialylated core2 O-glycans on CD44s contributes to hyaluronan binding and signaling

CD44 is the primary receptor for hyaluronan (HA), a glycosaminoglycan that is a major component of the extracellular matrix (*21*). This interaction of CD44 with HA is facilitated by many factors including N-glycosylation of CD44 (*22*, *23*), but less is known about the contribution of specific O-glycans to this interaction. To determine whether the core2 α2,3 sialoglycan contributes to the interaction of CD44 with HA, we first verified that CD44s expresses this glycan in TE3 cells (Fig. 5A). To assess the contribution of CD44s to HA binding, we depleted CD44 expression using short hairpin RNAs (shRNAs; Fig. 5, B and C). To visualize and quantify HA binding, we used fluorescein isothiocyanate (FITC)–conjugated HA to monitor binding by flow cytometry and confirmed that loss of CD44 diminished HA binding (Fig. 5D and fig. S4A). Subsequently, we used the shGCNT1 cells to investigate whether the core2 glycan mediates the binding of CD44 to HA. Although surface expression of CD44 was not altered with knockdown of GCNT1 (Fig. 5E), loss of core2 O-glycan expression decreased HA binding compared with the control cells and resulted in a level of HA binding comparable to that observed in shCD44 cells (Fig. 5F and fig. S4B). To determine whether the α2,3 sialoglycan expressed on core2 O-glycans contributes to the CD44s-HA interaction, we pretreated TE3 cells with SLBR-N and then incubated the cells with HA-FITC. TE3 cells pretreated with SLBR-N exhibited lower HA binding than cells incubated with HA alone or with the SNA lectin, suggesting that the α2,3 sialoglycan promotes the CD44s interaction with HA (Fig. 5G and fig. S4, C and D). An important issue that arose from this result is whether the CD44-HA interaction mediated by the core2 α2,3 sialoglycan contributes to CSC properties. To address this issue, we treated CSCs with the hyaluronic acid synthase inhibitor 4-methylumbelliferone (4-MU), which significantly reduced production of HA after 24 hours (Fig. 5H) and reduced mammosphere formation compared with vehicle alone (Fig. 5I).

The results described above prompted us to investigate how the interaction of HA with CD44 via the core2 α2,3 sialoglycan promotes the CSC phenotype. We verified that loss of CD44 by shRNA decreases phosphorylation and activation of the PDGFRβ/STAT3 signaling pathway in TE3 cells, similar to previous reports suggesting that CD44s expression promotes stemness through this pathway (fig. S4E) (*20*). To determine whether the glycans expressed on CD44 are contributing to the activation of the PDGFRβ/STAT3 pathway, we stimulated shGCNT1 cells with PDGF and observed that diminishing GCNT1 expression reduced the phosphorylation of PDGFRβ and STAT3 compared with control cells (Fig. 5J). To determine whether the HA-CD44 interaction is necessary to activate the PDGFRβ/STAT3 pathway, we treated TE3 cells with 4-MU and stimulated them with PDGF. Control cells showed significantly more phosphorylation and activation of the PDGFRβ/STAT3 pathway compared with the cells treated with 4-MU (Fig. 5K). Given our finding that the core2 α2,3 sialoglycan mediates CD44s binding to HA, we infer from these data that this glycan promotes the stem cell phenotype through activation of the PDGFRβ/STAT3 pathway.

sLeX inhibits stem cell properties and CD44s signaling

As mentioned above, our RNA-seq data revealed that the expression of the fucosyltransferase FUT3 is significantly higher in the non-CSC population compared with the CSC population (Fig. 1C). This finding intrigued us because FUT3 is an α1,3 fucosyltransferase that is associated with the expression of N-linked sLeX, which has been implicated in aggressive breast cancer (*24*, *25*). However, we found that FUT3 expression is significantly higher in epithelial basal A breast cancer cell lines compared with mesenchymal basal B cell lines

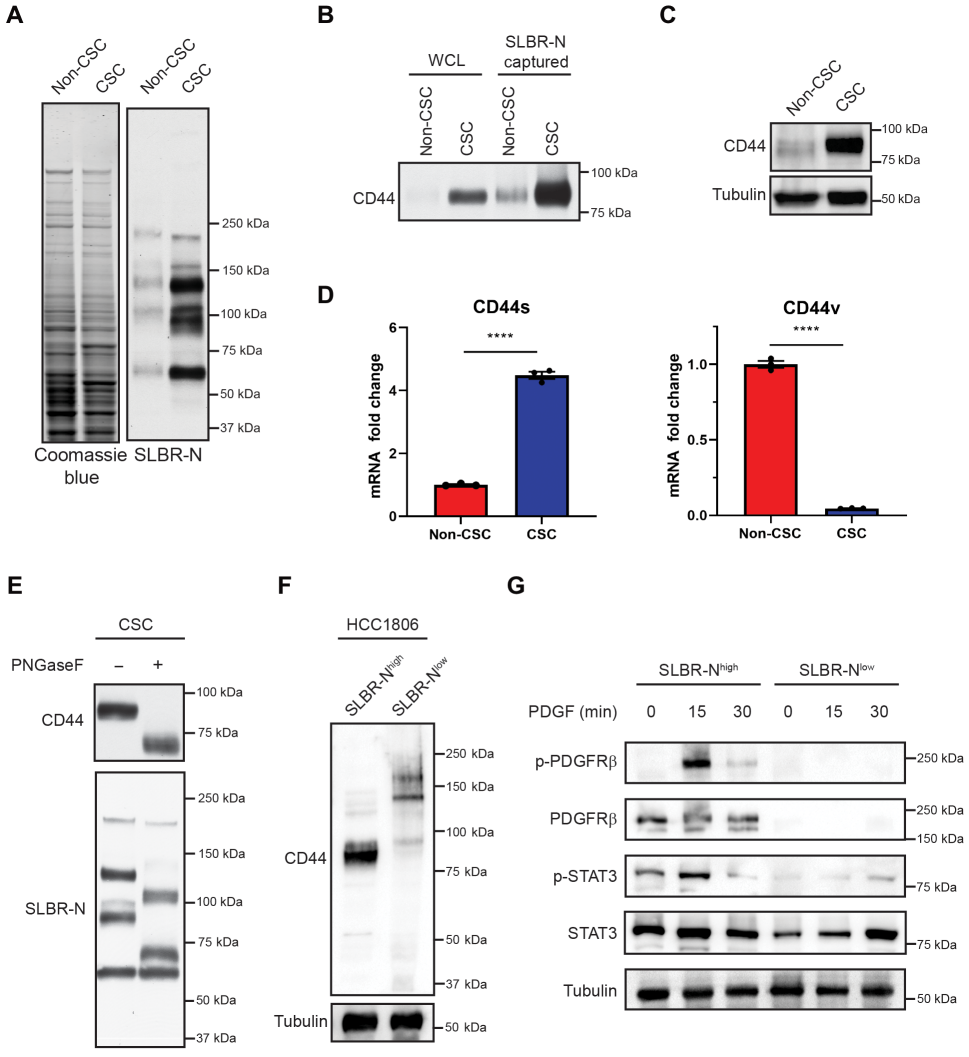


Fig. 4. CD44s expresses core2 O-linked α2,3 sialoglycan. (**A**) HMLER CSC and non-CSC whole-cell lysates were subjected to lectin blotting with SLBR-N to assess the expression of the core2 α2,3 sialoglycan. Loading was assessed by Coomassie stain. Representative results of two independent experiments. (**B**) HMLER CSC and non-CSC whole-cell lysates were subjected to affinity capture by SLBR-N. Captured proteins were assessed by immunoblotting for CD44. Representative results of two independent experiments. (**C**) HMLER CSC and non-CSC whole-cell lysates were assessed by immunoblotting for CD44 and tubulin. Representative results of three independent experiments. (**D**) mRNA was isolated from HMLER CSC and non-CSC cells, and CD44s and CD44v splice form expression was quantified by qPCR. Representative results of three independent replicates. (**E**) HMLER CSC whole-cell lysate was treated with or without 1 U of PNGaseF, and the migration pattern of SLBR-N–bound glycoproteins and CD44 was assessed by lectin and immunoblotting. Representative results of three independent experiments. (**F**) HCC1806 SLBR-N^{high} and SLBR-N^{low} whole-cell lysates were assessed by immunoblotting for CD44 and tubulin. Representative results of three independent experiments. (**G**) HCC1806 SLBR-N^{high} and SLBR-N^{low} were treated with PDGF (10 ng/ml) for 0, 15, or 30 min. Isolated protein was assessed by immunoblotting for phospho-PDGFRβ (Tyr^{1009}), total PDGFRβ, phospho-STAT3 (Tyr^{705}), STAT3, and tubulin. Representative results of three independent experiments. ****$P < 0.0001$.

by analyzing the Cancer Cell Line Encyclopedia (CCLE) database (Fig. 6A). This finding is consistent with our results that the SLBR-N^{low} population of HCC1806 cells, which are epithelial, had significantly higher expression of sLeX compared with the mesenchymal SLBR-N^{high} population as assessed by flow cytometry and immunoblotting using the HECA-452 antibody (Fig. 6B and fig. S5A). Likewise, the HMLER CSC population exhibited decreased HECA-452^{+} glycoproteins compared with the non-CSC population as assessed by immunoblotting (fig. S5B).

The above observations prompted us to assess the function of FUT3 and sLeX in the context of stem cell properties. For this purpose, we used TE3 cells because they have low expression of FUT3 and are enriched in CSC properties. Increasing FUT3 expression in these cells generated sLeX, but it did not affect surface expression of the SLBR-N glycan (Fig. 6, C to E). Also, PNGaseF treatment of the FUT3-expressing cells eliminated HECA-452 binding, demonstrating that sLeX is N-linked (Fig. 6F). FUT3 expression reduced mammosphere and soft agar colony formation significantly (Fig. 6G and fig. S5, D and E). Overexpression of FUT3 in the stem cell–enriched SLBR-N^{high} sorted cells significantly reduced mammosphere-forming ability (Fig. 6H and fig. S5C), indicating that sLeX antagonizes stem cell properties. We also observed that expression of FUT3 in TE3 cells induced expression of sLeX-modified CD44s (Fig. 6J) but that it did not alter CD44 surface expression (Fig. 6I). We then hypothesized

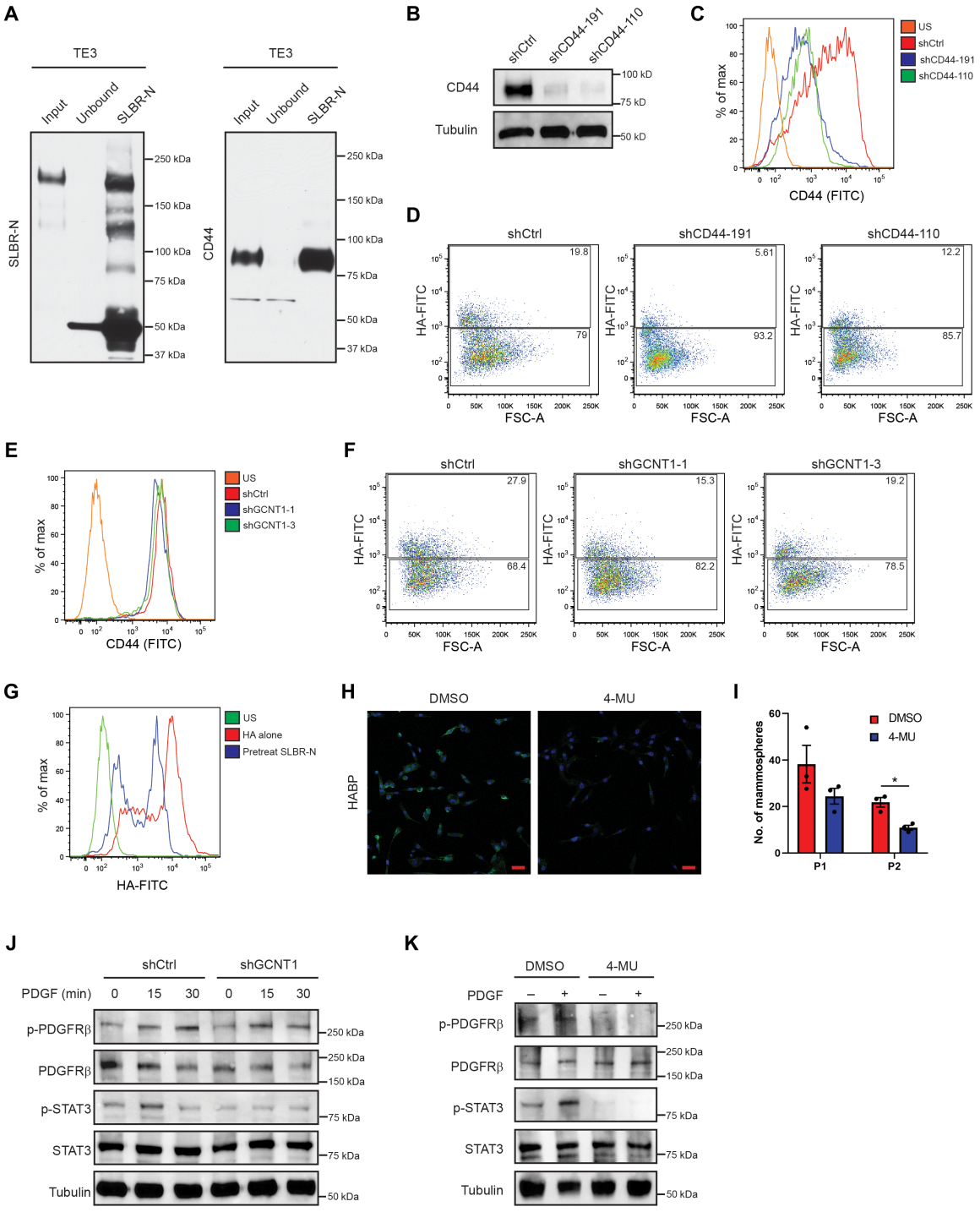


Fig. 5. O-linked α2,3 sialylation of CD44a contributes to HA binding and signaling. (**A**) TE3 lysate was subject to lectin affinity capture with SLBR-N and assessed for SLBR-N binding and CD44 expression. Representative results of two independent experiments. (**B**) TE3 cells were stably transfected with control shRNA (shCtrl) or two shRNA clones for CD44, and CD44 expression was assessed by immunoblotting. Representative results of three independent experiments. TE3 shCtrl and shCD44 cells were assessed for (**C**) CD44 expression and (**D**) HA binding affinity by flow cytometry. Representative results of three independent experiments. TE3 shCtrl and shGCNT1 cells were assessed for (**E**) CD44 expression and (**F**) HA binding affinity by flow cytometry. Representative results of three independent experiments. (**G**) TE3 cells were pretreated with SLBR-N and assessed for HA binding affinity by flow cytometry. Representative results of three independent experiments. (**H**) TE3 cells were treated with 4-MU, and HA production was measured by immunofluorescence microscopy. Representative results of three independent experiments. Scale bars, 50 μm. (**I**) TE3 cells were pretreated with 4-MU and analyzed by serial passage mammosphere formation. Representative results of three independent experiments. (**J**) TE3-shCtrl and TE3-shGCNT1 cells were treated with PDGF (10 ng/ml) for 0, 15, or 30 min, and expression of indicated proteins was assessed by immunoblotting. Representative results of three independent experiments. (**K**) TE3 cells pretreated with dimethyl sulfoxide (DMSO) or 4-MU and stimulated with PDGF (10 ng/ml) for 30 min; expression of indicated proteins was assessed by immunoblotting. Representative results of three independent experiments. *$P < 0.05$.

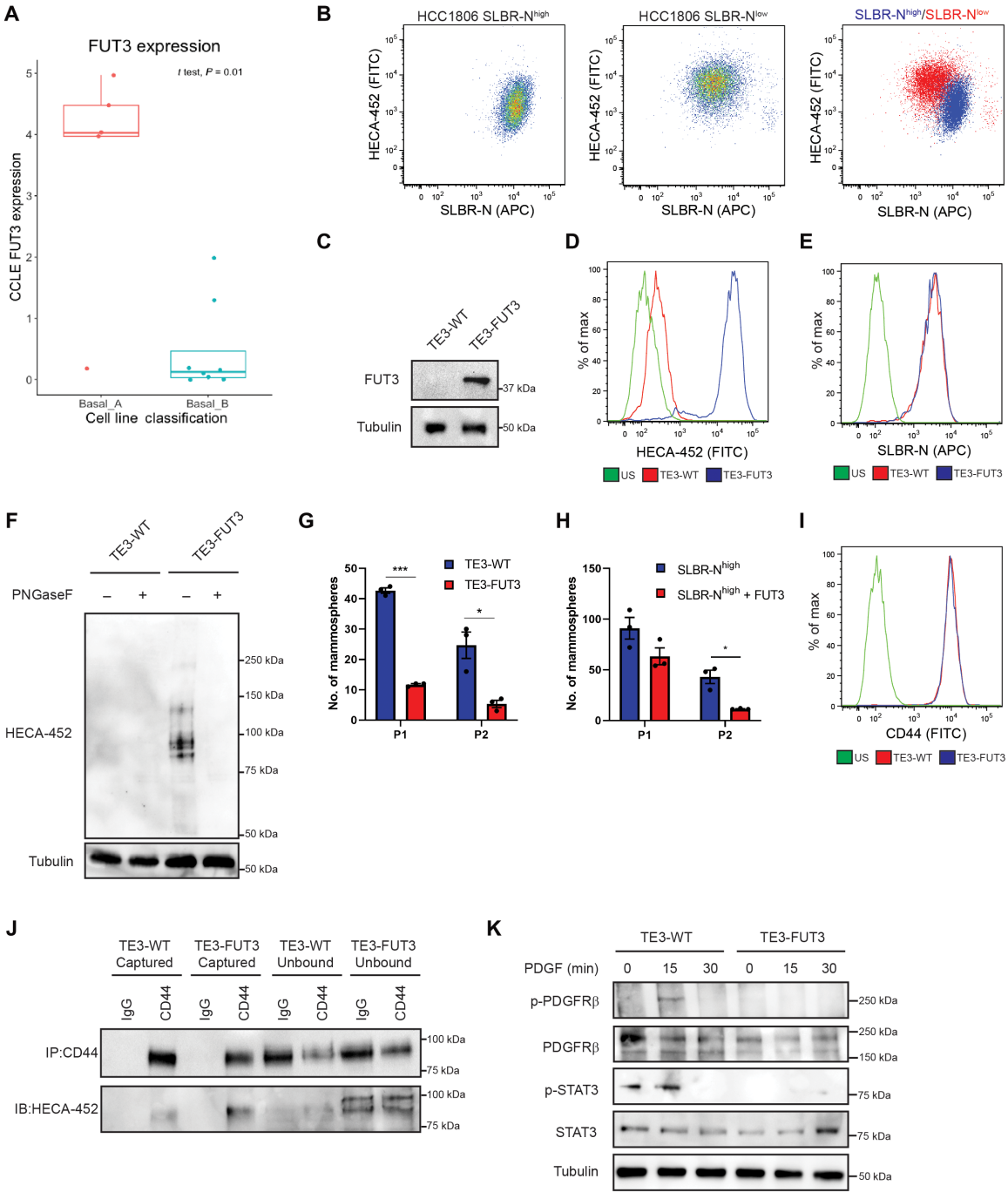


Fig. 6. Fucosylation of CD44s by FUT3 inhibits stem cell signaling. (**A**) The CCLE was analyzed for the expression of FUT3. Shown are the means ± SD of FUT3 expression in basal A and B breast cancer lines. (**B**) HCC1806 SLBR-N^{high} and SLBR-N^{low} cells were analyzed by flow cytometry for SLBR-N and HECA-452. Representative results of three independent experiments. (**C**) TE3 cells were stably transfected with a FUT3 overexpression construct and assessed by immunoblotting (IB) for FUT3 and tubulin. Representative results of three independent experiments. TE3-WTand TE3-FUT3 cells were analyzed by flow cytometry for (**D**) HECA-452 binding and (**E**) SLBR-N binding. Representative results of three independent experiments. (**F**) TE3-WT and TE3-FUT3 cells were treated with and without PNGaseF, and protein was analyzed by immunoblotting for HECA-452. Representative results of three independent experiments. (**G**) TE3-WT and TE3-FUT3 cells were analyzed by serial passage mammosphere formation. Representative results of three independent replicates. (**H**) HCC1806 SLBR-N^{high} and HCC1806 SLBR-N^{high} + FUT3-HA cells were analyzed by serial passage mammosphere formation. Representative results of three independent replicates. (**I**) TE3-WT and TE3-FUT3 cells were analyzed by flow cytometry for CD44 expression. Representative results of three independent experiments. (I) TE3-WT and TE3-FUT3 lysate was subject to immunoprecipitation (IP) for CD44 and immunoblotting for CD44 and HECA-452. Representative results of three independent experiments. (**J**) TE3-WT and TE3-FUT3 cells were treated with PDGF (10 ng/ml) for 0, 15, or 30 min, and expression of indicated proteins was assessed by immunoblotting. Representative results of three independent experiments. *P < 0.05 and ***P < 0.001.

that FUT3-mediated fucosylation of CD44s in breast cells affects the ability of CD44s to signal activation of the PDGFRβ/STAT3 pathway. Phosphorylation of PDGFRβ and STAT3 in response to PDGF was markedly reduced in FUT3-expressing cells compared with control cells. (Fig. 6K).

DISCUSSION

We demonstrate that the surface expression of a core2 (β-1,6-*N*-acetylglucosamine), O-linked α2,3 sialoglycan, which is recognized by the lectin SLBR-N, characterizes populations of breast CSCs. This lectin can be used to enrich for cells with stem cell properties in heterogeneous populations of tumor cells including PDX models of breast cancer. These conclusions are buttressed by RNA-seq data indicating that the expression of key enzymes involved in the biosynthesis of this glycan, including GCNT1 and the ST3GAL glycosyltransferases, is significantly elevated in CSC populations compared with non-CSC populations. We also demonstrate that this O-linked α2,3 sialoglycan is expressed on CD44s, a splice variant of CD44 critical for CSC function (*20*), and that it is necessary for mediating the interaction of CD44s with HA and signaling events that sustain CSC function.

Our results advance our understanding of the role of O-linked glycosylation in cancer, especially CSCs. Much of the work on this area has focused on mucins, heavily O-glycosylated proteins that express few N-glycans, that have been implicated in the behavior of breast and other cancers, including the function of CSCs (*26*, *27*). The O-linked core2 α2,3 sialoglycan that we have identified is not linked to a mucin, but its ability to modify the function of essential proteins such as CD44 underlies its contribution to CSC function. Although a previous study had shown that GCNT1, the enzyme that synthesizes the branched core2 glycan structure, contributes to the function of tumor-initiating cells in breast cancer (*28*), we argue that it is the terminal α2,3 sialylation that is the critical effector of CSC function. Sialylation is known to affect the function of tumor cells, including CSCs, and it is considered a potential therapeutic target (*29*). Of note, α2,6 sialylation mediated by ST6GAL1, primarily on N-linked glycans, has been implicated in the function of pancreatic and ovarian CSCs (*29*). Our results differ markedly from these studies because our data revealed that α2,6 sialylation ST6GAL1 expression is associated with a non-CSC population and that α2,3 sialylation of O-linked glycans is the critical effector of CSC function. The discrepancy between our data and other studies may relate to differences in the cancers studied or other factors that remain to be determined. Nonetheless, the dichotomy of α2,3 sialylation of O-glycans and α2,6 sialylation of N-glycans that we observed between CSCs and non-CSCs in breast cancer is notable, and it highlights the specificity of sialylation in affecting CSC function.

A major finding in this study is that CD44s expresses the O-linked, core2 α2,3 sialoglycan in breast CSCs and that this glycan is necessary for the ability of CD44s to interact with HA. Although CD44 expresses N- and O-linked glycans, most work has focused on the contribution of N-linked glycans to HA binding (*22*, *23*, *30*, *31*). A key study in this regard by Skelton *et al.* (*23*) demonstrated the role of glycosylation in regulating CD44-HA binding and the contribution of complex N-linked glycans to this binding. The study also indicated that O-linked glycans may be important based on the mutational analysis of Ser-Gly motifs in the membrane proximal domain. Subsequent studies have refined the nature of N-linked glycosylation on CD44, including CD44s, but much less is known about O-glycosylation (*32*, *33*). Recently, Mereiter and colleagues (*34*) found that truncation of CD44 O-glycans promotes interaction with HA in gastric cancer, suggesting that a variety of glycans can mediate this interaction; however, more work must be done to better understand the impact of O-glycosylation in this process. On the basis of the compensatory nature of α2,3 sialyltransferases and the technical constraints in diminishing this glycan, we targeted the α2,3 sialoglycan by knocking down the core2 synthase, GCNT1. Future studies delving into the specific structure of this glycan on CD44s would benefit from glycoproteomic analyses and O-glycosylation site mutants to better understand the mechanisms by which this α2,3 sialoglycan promotes the interaction with HA. The strength of our finding is that we identified a specific O-linked glycan that modifies CD44s and demonstrated that it contributes to the ability of CD44 to bind HA and facilitate breast CSC properties.

We conclude from our data that the ability of core2 α2,3 sialoglycan to promote the interaction of CD44s with HA underlies the ability of CD44s to promote CSC functions. Although the ability of CD44 to bind HA has been implicated in tumorigenicity and possibly CSC function (*35*), the contribution of CD44s glycosylation to these processes had not been established. An important contribution of our findings is that we link CD44s glycosylation and HA binding to signaling events that have been shown to be critical for the function of breast CSCs. In a recent collaborative study, we demonstrated that the breast CSC state is regulated by alternative splicing of CD44 and that CD44s determines CSC function (*20*). This study also revealed that CD44s interacts with PDGFRβ and activates a PDGFRβ/STAT3 signaling pathway that contributes to CSC function. In the current study, we advance our understanding of this mechanism by demonstrating that it is dependent on a core2 α2,3 sialylated glycoform of CD44s.

An unexpected finding in this study is that FUT3, a key enzyme in the genesis of sLeX, antagonizes CSC function. Both FUT3 and sLeX have been implicated in breast cancer progression and correlate with lower overall survival of patients (*24*, *25*). sLeX is the main ligand for E-selectin, which is expressed on endothelial cells, and the binding of sLeX on tumors to E-selectin on endothelial cells contributes to extravasation from the vasculature during metastasis (*24*). Our data reveal that sLeX antagonizes CSC function, including the function of CD44, and they highlight distinct differences in glycosylation between cells with stem cell properties and other populations of tumor cells. It is worth noting in this context that CD44s in hematopoietic stem and progenitor cells expresses N-linked sLeX, a glycoform termed "HCELL" (hematopoietic cell E-/L-selectin ligand) and that this glycan actually promotes stemness in these cells (*36*). The contrast with our data with respect to sLeX function on CD44s is notable, and it suggests key differences between hematopoietic and breast CSCs with respect to how glycosylation affects CD44 function.

Our work also highlights the confounding nature of CD44 as a breast CSC marker. CD44 was identified as a marker for the breast CSC population and was subsequently used to select for CSC populations in other cancer types (*16*, *37*–*40*). Our study supports the hypothesis that CD44 surface expression cannot be used exclusively to identify CSC populations in breast and other cancers (*20*, *41*–*43*). Not only is CD44 splice variant expression a determinant of CSC fate and function, but glycan expression on CD44 itself can alter its function (*20*, *44*). We argue that the SLBR-N lectin acts as a more

efficient CSC marker compared with conventional $CD44^+/CD24^-$ staining pattern because it enriches for cells with CSC properties while removing the confounding issues of CD44 splice forms and glycan status.

In summary, this study has identified a distinct glycan that distinguishes breast CSCs from non-CSCs. The ability of a specific lectin to recognize this glycan enables the enrichment of CSCs from heterogeneous tumor populations, a discovery that has potential implications for further mechanistic studies on CSCs, as well as for tumor diagnosis, prognosis, and therapy. The data provided also highlight how differences in sialylation and fucosylation have a profound impact on the regulation of CSC function.

MATERIALS AND METHODS

Reagents and cell culture

SLBR-N, SLBR-H, and SK678 were characterized and purified as previously described (*15*). Neuraminidase S and PNGaseF were purchased from New England Biolabs (NEB), and recombinant human PDGF was purchased from R&D Systems. Flow cytometry antibodies and lectins were acquired as follows: integrin β_4 (439-9b; Abcam, ab110167), CD24–allophycocyanin (APC; ML5, BioLegend), CD24–phycoerythrin (PE; ML5, BioLegend), CD44-FITC (IM7, BioLegend), SNA (B-1305-2, Vector Laboratories), PE streptavidin (405203, BioLegend), CLA (HECA-452, BioLegend), anti-rat immunoglobulin M (IgM) FITC (MRM-47, BioLegend), anti–glutathione *S*-transferase (GST) APC (Columbia Biosciences), HA-FITC (385906, EMD Millipore), and anti–CA19-9 (121SLE, Novus Biologicals). Immunoblotting antibodies were acquired as follows: tubulin [Cell Signaling Technology (CST), 3873], CD44 (Abcam, 157107), FUT3 (Abcam, 110082), anti-rat IgM horseradish peroxidase (Southern-Biotech), phospho-PDGFRβ Tyr^{1009} (CST, 3124), PDGFRβ (CST, 3169), phospho-STAT3 Tyr^{705} (CST, 9145), and STAT3 (CST, 9139).

HMLER cells were provided by R. Weinberg (Massachusetts Institute of Technology). Cells were cultured in MEBM mammary epithelial cell growth basal medium (Lonza). MDA-MB-231-TE3 was provided by S. Tavazoie (Rockefeller University). MDA-MB-231 and HCC1806 were obtained from the American Type Culture Collection.

Constructs and transfection

The following lentiviral shRNAs were used: shGCNT1 (TRCN0000035233 and TRCN0000035231), shCD44 (TRCN0000308110 and TRCN0000296191), and shST3GAL6 (TRCN0000288762 and TRCN0000288833) purchased from Sigma-Aldrich. FUT3 lentiviral overexpression plasmid was designed and purchased from VectorBuilder. Lipofectamine 3000 (Thermo Fisher Scientific) was used for plasmid transfection.

RNA-seq data analysis

For all analyses, default parameters were used unless otherwise specified. Read quality was checked by using FASTQC (version 0.11.5) (*45*). Then, reads were aligned to the human reference genome GRCh38 downloaded from Ensembl using Tophat2 (version 2.1.1) (*46*) with minimal and maximal intron length set as 20 base pairs (bp) and 100 kb, respectively. The resulting BAM files were sorted by name and processed by using MMR (*47*) to assign multimapper reads to their most likely genomic locations. The BAM file outputs by MMR were input to featureCounts (version 1.5.3) (*48*) to generate a gene-level count summary table with the following explicit option settings –d 35 –t exon –M and the GTF file for the reference genome GRCh38.89 as the genome annotation file. Genes of extremely low expression were removed from the count table if their expression level is not higher than 1 cpm in at least two samples, which is the number of samples in each group. Differential expression analysis was performed using DESeq2 (version 1.16.1) (*49*) by comparing gene expression between the two groups (CSC and non-CSC cells). Genes with a q value less than 0.05 were considered as significantly differentially expressed between the two groups.

The heatmap was plotted with R package ComplexHeatmap (*50*). Only significant differentially expressed genes (DEG) were visualized. The criteria for DEG were false discovery rate (FDR) < 0.05 and $|\log2FC| > 1$. All genes listed in the Glyco-enzyme Expression Repository (Complex Carbohydrate Research Center, University of Georgia) were marked on the heatmap, unless it is not in the DEG list.

Lectin microarray

Sample processing for lectin Array

Cell membrane extraction from CSCs and non-CSCs was prepared for lectin microarrays and labeled with AlexaFluor-555 as previously described (*51*). A pooled reference was labeled with AlexaFluor-647. Details of sample preparation are given in table S1.

Lectin microarray printing, hybridization, and analysis

Lectin microarrays were printed as previously described (*52*). The printlist, including lectin sources, is given in table S2. Equal amounts of sample and reference (5 μg) were hybridized on the array, and data analysis was performed as previously described (*53*). Additional experimental information can be found in table S1.

Mammosphere assays

Cells were plated in UltraLow attachment 24-well plates in Dulbecco's modified Eagle's medium/F12 medium supplemented with B27, epidermal growth factor, and fibroblast growth factor and passaged as previously described (*54*). Mammospheres were counted using the Celigo Imaging Cytometer (Nexcelom).

Cell-based assays

For flow cytometry assays, cells were incubated with antibodies as per the manufacturer's instructions. Cells treated with PNGaseF or NeuS (NEB) were incubated with 1 U of enzyme for 1 hour at 37°C before staining with antibodies. For cell viability assays, cells were seeded at a density of 5000 cells per well. Twenty-four hours later, cells were treated with a range of concentrations (2.5, 5, 10, 25, and 50 μM) of 5-FU and saline solution (control). Ninety-six hours after treatment, cells were assessed for viability using crystal violet. The absorbance measured at 600 nm was normalized to the saline solution control.

Patient-derived organoids

Tumor tissues of freshly resected biopsies from patients with TNBC were obtained from UMass Cancer Center Tumor Bank. These tumors were digested using a tissue dissociation kit (Miltenyi Biotech) and gentleMACS Dissociator. The digested tumors were washed thrice using phosphate-buffered saline, and partially digested tumor pieces were embedded into reduced growth factor basement membrane extract (Cultrex). The embedded organoids were cultured in organoid media as described previously (*55*). The drug-resistant

organoids were generated by culturing them in the presence of 5-FU and cisplatin for 2 weeks. Single-cell suspension was prepared from organoids using Accumax solutions and stained using SLBR-N as described above for flow cytometry. Cell viability was calculated using CellTiter-Glo (Promega).

In vivo studies

For xenograft experiments, cells were injected into the mammary fat pad of 4- to 6-week-old NOD.Cg-*Prkdc*^{scid} *IL2rg*^{tm1Wjl} (abbreviated as NSG) mice in 50 μl of Cultrex reduced growth factor basement membrane matrix (Trevigen). Mice were monitored for 7 to 8 weeks, and tumors were collected for histological analysis at the end of the experiment. Limiting dilution analysis was performed using the extreme limiting dilution analysis (*56*). All mouse procedures were done under the guidance of the University of Massachusetts Medical School Institutional Animal Care and Use Committee in accordance with the institutional and regulatory guidelines.

Lectin blotting, immunoblotting, and immunoprecipitation

Lectin blotting and immunoblotting were performed as previously described (*15*, *54*). Briefly, cells were harvested and lysed in radioimmunoprecipitation assay (RIPA) buffer (Sigma-Aldrich) with protease and phosphatase inhibitors (Roche). For PNGaseF experiments, equal concentrations of whole-cell lysate were treated with 1 U of PNGaseF (NEB) for 1 hour at 37°C. For PDGF experiments, PDGF (10 ng/ml; R&D Systems) was incubated with cells for the indicated time intervals. Antibodies used for immunoblotting were diluted to a concentration of 1:1000.

For lectin pulldown and immunoprecipitation, after preclearing the lysate with Glutathione Sepharose (GE Healthcare) or protein A agarose (Sigma-Aldrich) beads for 1 hour at 4°C, 0.5 to 1 mg of lysate was incubated with 1 μg of indicated lectin or antibodies overnight at 4°C. After adding Glutathione Sepharose or protein A agarose beads, lysates were incubated for 1 hour at 4°C.

Real-time qPCR

RNA extraction was accomplished using an RNA isolation kit (BS88133, Bio Basic Inc.), and complementary DNAs (cDNAs) were produced using an Azura cDNA synthesis kit (AZ-1996, Azura Genomics). AzuraView GreenFast qPCR Blue Mix LR was used as the qPCR Master Mix (AZ-1996, Azura Genomics). qPCR primers: 18S (forward, 5′-GTCGCTCGCTCCTCTCCTACT-3′; reverse, 5′-TCTGATAAATGCACGCATCCC-3′), GCNT1 (forward, 5′-AACCCCTTAGTAAAGAAGAGGCG-3′; reverse, 5′-AGCAGCCTGTCAAGCATTTCA-3′), ST3GAL2 (forward, 5′-CGTCTGGACCCGAGAGAAC-3′; reverse, 5′-GCCAGGCACTATCTGGAACA-3′), ST3GAL5 (forward, 5′-AGGAATGTCGTCCCAAGTTTG-3′; reverse, 5′-GGAGTAAGTCCACGCTATACCT-3′), ST3GAL6 (forward, 5′-ACTGCATTGCATATTATGGGGAA-3′; reverse, 5′-TGGCTTTGATAAACAAGGCTGG-3′), ST6GAL1 (forward, 5′-CTGAATGGGAGGGTTATCTGCC-3′; reverse, 5′-ACCTCAGGACTGCGTCATGATC-3′), FUT3 (forward, 5′-GCCGACCGCAAGGTGTAC-3′; reverse, 5′-TGACTTAGGGTTGGACATGATATCC-3′), CDH1 (forward, 5′-TGAAGGTGACAGAGCCTCTGGAT-3′; reverse, 5′-TGGGTGAATTCGGGCTTGTT-3′), CDH2 (forward, 5′-TCAGGCGTCTGTAGAGGCTT-3′; reverse, 5′-ATGCACATCCTTCGATAAGACTG-3′), CD44s (forward, 5′-TACTGATGATGACGTGAGCA-3′; reverse, 5′-GAATGTGTCTTGGTCTCTGGT-3′), and CD44v5/6 (forward, 5′-GTAGACAGAAATGGCACCAC-3′; reverse, 5′-CAGCTGTCCCTGTTGTCGAA-3′).

CCLE dataset analysis

Gene expression data were collected from publicly available gene expression datasets from the CCLE (*57*). Expression values represent log2-transformed TPM gene expression counts from RNA-seq data. Adherent TNBC cell lines were grouped by their molecular signatures into basal A or basal B subtypes as defined by available metadata for the CCLE dataset. Basal A subtypes represent cells with an epithelial gene expression signature positive for markers such as Krt5 and Krt14, while basal B subtypes represent cells with a mesenchymal gene expression signature positive for markers such as vimentin (*58*). Graphs were created using the ggboxplot function from the ggpubr package in R, and *P* value was calculated using a one-tailed *t* test.

Statistical analysis

Statistical comparison between only two groups was done with the unpaired Student's *t* test. Multiple-group comparisons were completed using one-way analysis of variance (ANOVA). Statistical tests were carried out using GraphPad Prism version 8.0, and a *P* value of less than 0.05 was considered significant. The bars in graphs represent means ± SEM. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abj9513

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. K. Polyak, Heterogeneity in breast cancer. *J. Clin. Invest.* **121**, 3786–3788 (2011).
2. G. Turashvili, E. Brogi, Tumor heterogeneity in breast cancer. *Front. Med.* **4**, 227 (2017).
3. M. Karaayvaz, S. Cristea, S. M. Gillespie, A. P. Patel, R. Mylvaganam, C. C. Luo, M. C. Specht, B. E. Bernstein, F. Michor, L. W. Ellisen, Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat. Commun.* **9**, 3588 (2018).
4. D. A. Lawson, N. R. Bhakta, K. Kessenbrock, K. D. Prummel, Y. Yu, K. Takai, A. Zhou, H. Eyob, S. Balakrishnan, C. Y. Wang, P. Yaswen, A. Goga, Z. Werb, Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
5. S. Zhou, Y.-E. Huang, H. Liu, X. Zhou, M. Yuan, F. Hou, L. Wang, W. Jiang, Single-cell RNA-seq dissects the intratumoral heterogeneity of triple-negative breast cancer based on gene regulatory networks. *Mol. Ther. Nucleic Acids* **23**, 682–690 (2021).
6. I. Januškevičienė, V. Petrikaitė, Heterogeneity of breast cancer: The importance of interaction between different tumor cell populations. *Life Sci.* **239**, 117009 (2019).
7. J. Zhao, Cancer stem cells and chemoresistance: The smartest survives the raid. *Pharmacol. Ther.* **160**, 145–158 (2016).
8. S. R. Stowell, T. Ju, R. D. Cummings, Protein glycosylation in cancer. *Annu. Rev. Pathol.* **10**, 473–510 (2015).
9. D. D. Engle, H. Tiriac, K. D. Rivera, A. Pommier, S. Whalen, T. E. Oni, B. Alagesan, E. J. Lee, M. A. Yao, M. S. Lucito, B. Spielman, B. Da Silva, C. Schoepfer, K. Wright, B. Creighton, L. Afinowicz, K. H. Yu, R. Grützmann, D. Aust, P. A. Gimotty, K. S. Pollard, R. H. Hruban, M. G. Goggins, C. Pilarsky, Y. Park, D. J. Pappin, M. A. Hollingsworth, D. A. Tuveson, The glycan CA19-9 promotes pancreatitis and pancreatic cancer in mice. *Science* **364**, 1156–1162 (2019).
10. C.-W. Li, S.-O. Lim, E. M. Chung, Y.-S. Kim, A. H. Park, J. Yao, J.-H. Cha, W. Xia, L.-C. Chan, T. Kim, S.-S. Chang, H.-H. Lee, C.-K. Chou, Y. L. Liu, H.-C. Yeh, E. P. Perillo, A. K. Dunn, C.-W. Kuo, K.-H. Khoo, J. L. Hsu, Y. Wu, J.-M. Hsu, H. Yamaguchi, T.-H. Huang, A. A. Sahin, G. N. Hortobagyi, S. S. Yoo, M.-C. Hung, Eradication of triple-negative breast cancer cells by targeting glycosylated PD-L1. *Cancer Cell* **33**, 187–201.e10 (2018).
11. N. M. Akella, G. Le Minh, L. Ciraku, A. Mukherjee, Z. A. Bacigalupa, D. Mukhopadhyay, V. L. Sodi, M. J. Reginato, O-GlcNAc transferase regulates cancer stem-like potential of breast cancer cells. *Mol. Cancer Res.* **18**, 585–598 (2020).
12. B. Elenbaas, L. Spirio, F. Koerner, M. D. Fleming, D. B. Zimonjic, J. L. Donaher, N. C. Popescu, W. C. Hahn, R. A. Weinberg, Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* **15**, 50–65 (2001).
13. Y. Geng, J. J. Amante, H. L. Goel, X. Zhang, M. R. Walker, D. C. Luther, A. M. Mercurio, V. M. Rotello, Differentiation of cancer stem cells through nanoparticle surface engineering. *ACS Nano* **14**, 15276–15285 (2020).

14. K. T. Pilobello, D. E. Slawek, L. K. Mahal, A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11534–11539 (2007).
15. B. A. Bensing, Q. Li, D. Park, C. B. Lebrilla, P. M. Sullam, Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: Implications for infective endocarditis. *Glycobiology* **28**, 601–611 (2018).
16. M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, M. F. Clarke, Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3983–9838 (2003).
17. J. B. Ross, D. Huh, L. B. Noble, S. F. Tavazoie, Identification of molecular determinants of primary and metastatic tumour re-initiation in breast cancer. *Nat. Cell Biol.* **17**, 651–664 (2015).
18. W. H. Yang, C. Nussbaum, P. K. Grewal, J. D. Marth, M. Sperandio, Coordinated roles of ST3Gal-VI and ST3Gal-IV sialyltransferases in the synthesis of selectin ligands. *Blood* **120**, 1015–1026 (2012).
19. M. R. Carter, J. L. Hornick, S. Lester, C. D. Fletcher, Spindle cell (sarcomatoid) carcinoma of the breast: A clinicopathologic and immunohistochemical analysis of 29 cases. *Am. J. Surg. Pathol.* **30**, 300–309 (2006).
20. H. Zhang, R. L. Brown, Y. Wei, P. Zhao, S. Liu, X. Liu, Y. Deng, X. Hu, J. Zhang, X. D. Gao, Y. Kang, A. M. Mercurio, H. L. Goel, C. Cheng, CD44 splice isoform switching determines breast cancer stem cell state. *Genes Dev.* **33**, 166–179 (2019).
21. A. Aruffo, I. Stamenkovic, M. Melnick, C. B. Underhill, B. Seed, CD44 is the principal cell surface receptor for hyaluronate. *Cell* **61**, 1303–1313 (1990).
22. A. Bartolazzi, A. Nocks, A. Aruffo, F. Spring, I. Stamenkovic, Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *J. Cell Biol.* **132**, 1199–1208 (1996).
23. T. P. Skelton, C. Zeng, A. Nocks, I. Stamenkovic, Glycosylation provides both stimulatory and inhibitory effects on cell surface and soluble CD44 binding to hyaluronan. *J. Cell Biol.* **140**, 431–446 (1998).
24. J. C. F. do Nascimento, E. I. C. Beltrão, C. R. C. Rocha, High *FUT3* expression is a marker of lower overall survival of breast cancer patients. *Glycoconj. J.* **37**, 263–275 (2020).
25. S. Julien, A. Ivetic, A. Grigoriadis, D. QiZe, B. Burford, D. Sproviero, G. Picco, C. Gillett, S. L. Papp, L. Schaffer, A. Tutt, J. Taylor-Papadimitriou, S. E. Pinder, J. M. Burchell, Selectin ligand sialyl-Lewis x antigen drives metastasis of hormone-dependent breast cancers. *Cancer Res.* **71**, 7683–7693 (2011).
26. K. Engelmann, H. Shen, O. J. Finn, MCF7 side population cells with characteristics of cancer stem/progenitor cells express the tumor antigen MUC1. *Cancer Res.* **68**, 2419–2426 (2008).
27. J. M. Curry, K. J. Thompson, S. G. Rao, D. M. Besmer, A. M. Murphy, V. Z. Grdzelishvili, W. A. Ahrens, I. H. McKillop, D. Sindram, D. A. Iannitti, J. B. Martinie, P. Mukherjee, The use of a novel MUC1 antibody to identify cancer stem cells and circulating MUC1 in mice and patients with pancreatic cancer. *J. Surg. Oncol.* **107**, 713–722 (2013).
28. J. Kim, R. Villadsen, T. Sørlie, L. Fogh, S. Z. Grønlund, A. J. Fridriksdottir, I. Kuhn, F. Rank, V. T. Wielenga, H. Solvang, P. A. W. Edwards, A.-L. Børresen-Dale, L. Rønnov-Jessen, M. J. Bissell, O. W. Petersen, Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6124–6129 (2012).
29. M. J. Schultz, A. T. Holdbrooks, A. Chakraborty, W. E. Grizzle, C. N. Landen, D. J. Buchsbaum, M. G. Conner, R. C. Arend, K. J. Yoon, C. A. Klug, D. C. Bullard, R. A. Kesterson, P. G. Oliver, A. K. O'Connor, B. K. Yoder, S. L. Bellis, The tumor-associated glycosyltransferase ST6Gal-I regulates stem cell transcription factors and confers a cancer stem cell phenotype. *Cancer Res.* **76**, 3978–3988 (2016).
30. S. Katoh, Z. Zheng, K. Oritani, T. Shimozato, P. W. Kincade, Glycosylation of CD44 negatively regulates its recognition of hyaluronan. *J. Exp. Med.* **182**, 419–429 (1995).
31. A. Dasgupta, K. Takahashi, M. Cutler, K. K. Tanabe, O-linked glycosylation modifies cd44 adhesion to hyaluronate in colon carcinoma cells. *Biochem. Biophys. Res. Commun.* **227**, 110–117 (1996).
32. P. S. Raman, C. S. Alves, D. Wirtz, K. Konstantopoulos, Distinct kinetic and molecular requirements govern CD44 binding to hyaluronan versus fibrin(ogen). *Biophys. J.* **103**, 415–423 (2012).
33. C. E. Faller, O. Guvench, Terminal sialic acids on CD44 N-glycans can block hyaluronan binding by forming competing intramolecular contacts with arginine sidechains. *Proteins* **82**, 3079–3089 (2014).
34. S. Mereiter, Á. M. Martins, C. Gomes, M. Balmaña, J. A. Macedo, K. Polom, F. Roviello, A. Magalhães, C. A. Reis, O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer. *FEBS Lett.* **593**, 1675–1689 (2019).
35. H. Jariyal, C. Gupta, A. Srivastava, Hyaluronic acid induction on breast cancer stem cells unfolds subtype specific variations in stemness and epithelial-to-mesenchymal transition. *Int. J. Biol. Macromol.* **160**, 1078–1089 (2020).
36. R. Sackstein, The biology of CD44 and HCELL in hematopoiesis: The 'step 2-bypass pathway' and other emerging perspectives. *Curr. Opin. Hematol.* **18**, 239–248 (2011).
37. L. Jin, K. J. Hope, Q. Zhai, F. Smadja-Joffe, J. E. Dick, Targeting of CD44 eradicates human acute myeloid leukemic stem cells. *Nat. Med.* **12**, 1167–1174 (2006).
38. P. Dalerba, S. J. Dylla, I. K. Park, R. Liu, X. Wang, R. W. Cho, T. Hoey, A. Gurney, E. H. Huang, D. M. Simeone, A. A. Shelton, G. Parmiani, C. Castelli, M. F. Clarke, Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10158–10163 (2007).
39. C. Li, D. G. Heidt, P. Dalerba, C. F. Burant, L. Zhang, V. Adsay, M. Wicha, M. F. Clarke, D. M. Simeone, Identification of pancreatic cancer stem cells. *Cancer Res.* **67**, 1030–1037 (2007).
40. M. E. Prince, R. Sivanandan, A. Kaczorowski, G. T. Wolf, M. J. Kaplan, P. Dalerba, I. L. Weissman, M. F. Clarke, L. E. Ailles, Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 973–978 (2007).
41. J. A. Magee, E. Piskounova, S. J. Morrison, Cancer stem cells: Impact, heterogeneity, and uncertainty. *Cancer Cell* **21**, 283–296 (2012).
42. J. E. Visvader, G. J. Lindeman, Cancer stem cells: Current status and evolving complexities. *Cell Stem Cell* **10**, 717–728 (2012).
43. B. Tang, A. Raviv, D. Esposito, K. C. Flanders, C. Daniel, B. T. Nghiem, S. Garfield, L. Lim, P. Mannan, A. I. Robles, W. I. Smith Jr., J. Zimmerberg, R. Ravin, L. M. Wakefield, A flexible reporter system for direct observation and isolation of cancer stem cells. *Stem Cell Reports* **4**, 155–169 (2015).
44. J. S. Merzaban, M. M. Burdick, S. Z. Gadhoum, N. M. Dagia, J. T. Chu, R. C. Fuhlbrigge, R. Sackstein, Analysis of glycoprotein E-selectin ligands on human and mouse marrow cells enriched for hematopoietic stem/progenitor cells. *Blood* **118**, 1774–1783 (2011).
45. S. Andrews, FASTQC;https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
46. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
47. A. Kahles, J. Behr, G. Rätsch, MMR: A tool for read multi-mapper resolution. *Bioinformatics* **32**, 770–772 (2016).
48. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
49. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
50. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
51. P. Agrawal, B. Fontanals-Cirera, E. Sokolova, S. Jacob, C. A. Vaiana, D. Argibay, V. Davalos, M. McDermott, S. Nayak, F. Darvishian, M. Castillo, B. Ueberheide, I. Osman, D. Fenyö, L. K. Mahal, E. Hernando, A systems biology approach identifies FUT8 as a driver of melanoma metastasis. *Cancer Cell* **31**, 804–819.e7 (2017).
52. K. T. Pilobello, P. Agrawal, R. Rouse, L. K. Mahal, Advances in lectin microarray technology: Optimized protocols for piezoelectric print conditions. *Curr. Protoc. Chem. Biol.* **5**, 1–23 (2013).
53. S. Koppolu, L. Wang, A. Mathur, J. A. Nigam, C. S. Dezzutti, C. Isaacs, L. Meyn, K. E. Bunge, B. J. Moncla, S. L. Hillier, L. C. Rohan, L. K. Mahal, Vaginal product formulation alters the innate antiviral activity and glycome of cervicovaginal fluids with implications for viral susceptibility. *ACS Infect. Dis.* **4**, 1613–1622 (2018).
54. A. L. Elaimy, S. Guru, C. Chang, J. Ou, J. J. Amante, L. J. Zhu, H. L. Goel, A. M. Mercurio, VEGF–neuropilin-2 signaling promotes stem-like traits in breast cancer cells by TAZ-mediated repression of the Rac GAP β2-chimaerin. *Sci. Signal.* **11**, eaao6897 (2018).
55. N. Sachs, J. de Ligt, O. Kopper, E. Gogola, G. Bounova, F. Weeber, A. V. Balgobind, K. Wind, A. Gracanin, H. Begthel, J. Korving, R. van Boxtel, A. A. Duarte, D. Lelieveld, A. van Hoeck, R. F. Ernst, F. Blokzijl, I. J. Nijman, M. Hoogstraat, M. van de Ven, D. A. Egan, V. Zinzalla, J. Moll, S. F. Boj, E. E. Voest, L. Wessels, P. J. van Diest, S. Rottenberg, R. G. J. Vries, E. Cuppen, H. Clevers, A living biobank of breast cancer organoids captures disease heterogeneity. *Cell* **172**, 373–386.e10 (2018).
56. Y. Hu, G. K. Smyth, ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).
57. Broad DepMap, DepMap 20Q4 Public. figshare. Dataset (2020);https://doi.org/10.6084/m9.figshare.13237076.v4.
58. R. M. Neve, K. Chin, J. Fridlyand, J. Yeh, F. L. Baehner, T. Fevr, L. Clark, N. Bayani, J. P. Coppe, F. Tong, T. Speed, P. T. Spellman, S. DeVries, A. Lapuk, N. J. Wang, W. L. Kuo, J. L. Stilwell, D. Pinkel, D. G. Albertson, F. M. Waldman, F. McCormick, R. B. Dickson, M. D. Johnson, M. Lippman, S. Ethier, A. Gazdar, J. W. Gray, A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527 (2006).

Acknowledgments

Funding: This work was supported by NIH grants 1F31CA239509 (M.R.W.), R50 CA221780 (H.L.G.), CA CA168464 and CA218085 (A.M.M.), and NIDCR R03DE029516 (B.A.B.); DOD grant CA171043 (L.K.M.); and the Canada Excellence Research Chairs Program (L.K.M.).

Author contributions: M.R.W., B.A.B., and A.M.M. designed the experiments and analyzed the data. M.R.W., H.L.G., D.M., P.C., and B.A.B. performed the experiments. E.R.K., H.L., R.L., and J.L.Z. performed the bioinformatics analysis. S.C. and L.K.M. performed and analyzed the lectin microarray. J.L.C. assisted with the pathological analysis of tumor samples. M.R.W. and A.M.M. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The lectin microarray data are available through Synapse.org (doi:10.7303/syn26451619). The RNA-seq data are available through the GEO database (GSE115058). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 15 June 2021
Accepted 11 November 2021
Published 7 January 2022
10.1126/sciadv.abj9513