Science Advances
AAAS

Supplementary Materials for

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

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

*Corresponding author. Email: arthur.mercurio@umassmed.edu

Published 7 January 2022, *Sci. Adv.* **8**, eabj9513 (2022)
DOI: 10.1126/sciadv.abj9513

This PDF file includes:

Figure S1

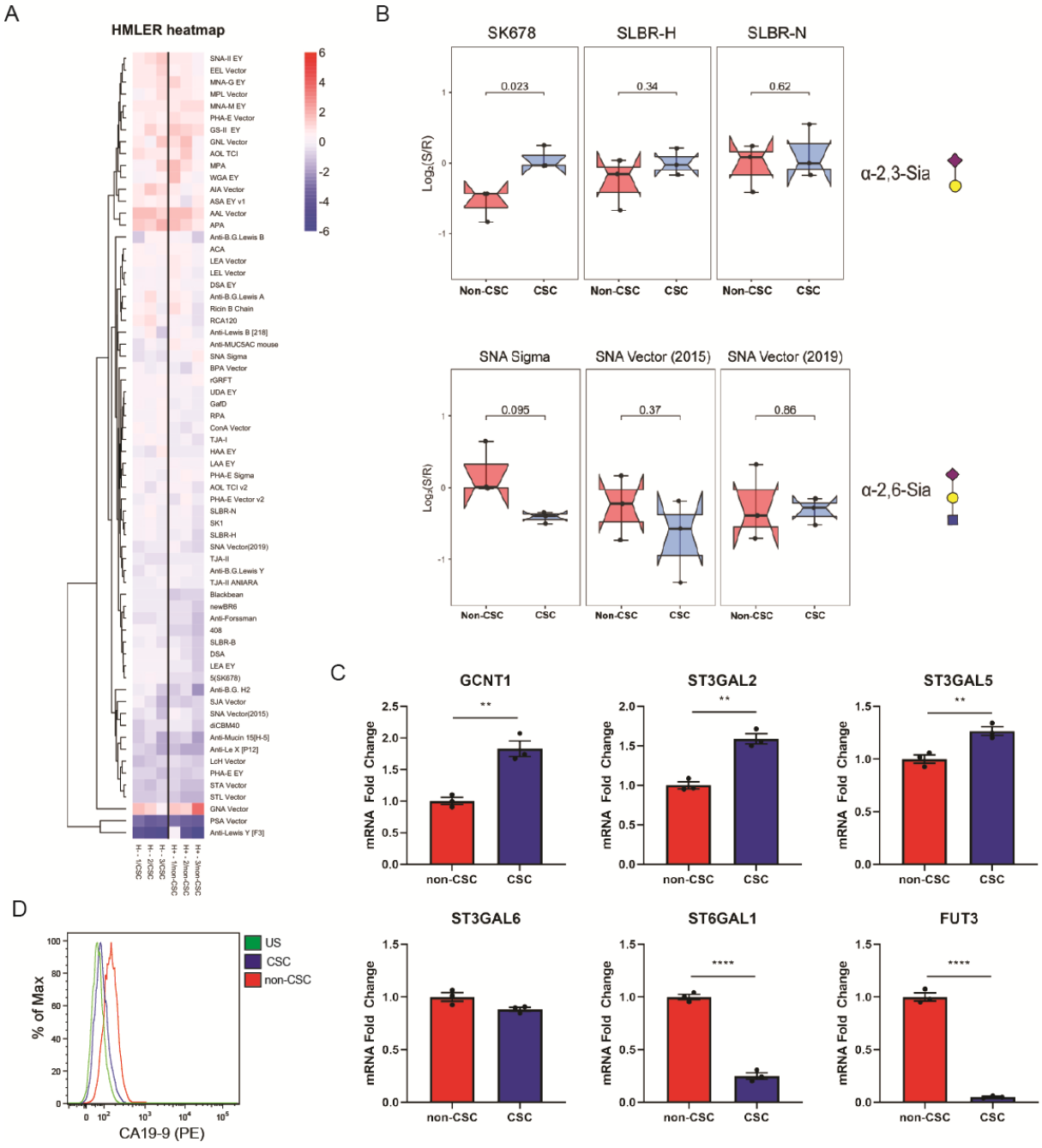

Figure S1: A) Heatmap of lectin clustering of CSC and non-CSC populations by dual color lectin microarray technology (n=3). Red, log2(S/R) > log2(Smedian/Rmedian); blue, log2(Smedian/Rmedian) > log2(S/R). Experimental information and lectin printlist are detailed in Tables S1 and S2. **B)** Boxplots depicting 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 with standard deviation. **C)** CSC and non-CSC RNA was analyzed by qPCR for

glycosyltransferases of interest (GCNT1, ST3GAL2, ST3GAL5, ST3GAL6, ST6GAL1, and FUT3). Shown are the means ± SEM of a representative experiment of three independent replicates. **D)** HMLER CSC and non-CSC cells were subject to flow cytometric analysis and sLea surface expression was assessed by staining with a CA19-9 antibody. Shown is one replicate of two independent experiments.

Figure S2

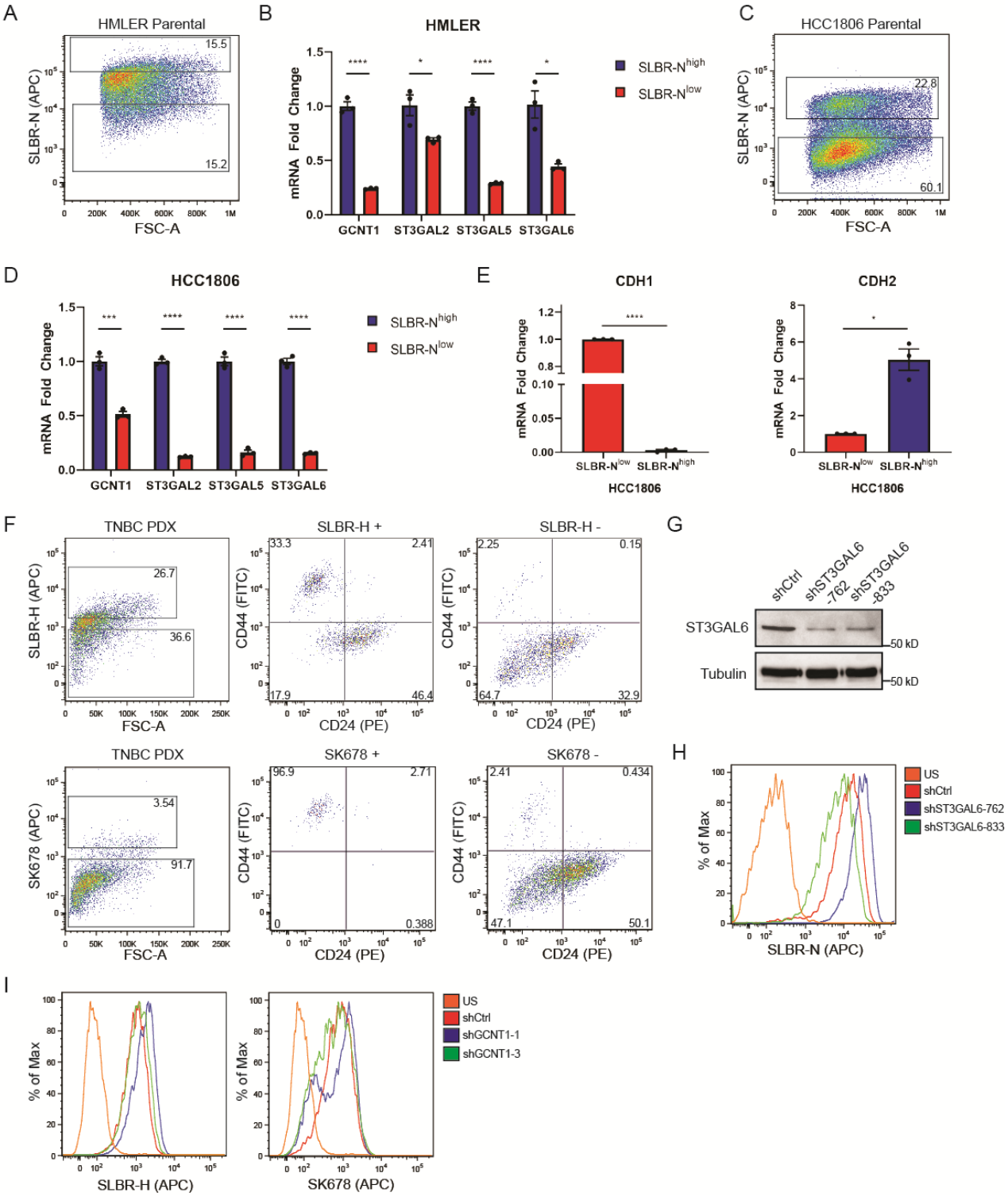
A
HMLER Parental
15.5
15.2
SLBR-N (APC)
FSC-A
B
HMLER
mRNA Fold Change
GCNT1
ST3GAL2
ST3GAL5
ST3GAL6
SLBR-N^high
SLBR-N^low
C
HCC1806 Parental
22.8
60.1
SLBR-N (APC)
FSC-A
D
HCC1806
mRNA Fold Change
GCNT1
ST3GAL2
ST3GAL5
ST3GAL6
SLBR-N^high
SLBR-N^low
E
CDH1
CDH2
mRNA Fold Change
SLBR-N^low
SLBR-N^high
HCC1806
F
TNBC PDX
SLBR-H +
SLBR-H -
SLBR-H (APC)
CD44 (FITC)
CD24 (PE)
FSC-A
26.7
36.6
33.3
2.41
17.9
46.4
2.25
0.15
64.7
32.9
TNBC PDX
SK678 +
SK678 -
SK678 (APC)
3.54
91.7
96.9
2.71
0
0.388
2.41
0.434
47.1
50.1
G
shCtrl
shST3GAL6 -762
shST3GAL6 -833
ST3GAL6
Tubulin
50 kD
50 kD
H
% of Max
SLBR-N (APC)
US
shCtrl
shST3GAL6-762
shST3GAL6-833
I
% of Max
SLBR-H (APC)
SK678 (APC)
US
shCtrl
shGCNT1-1
shGCNT1-3

Figure S2: A) HMLER cells were analyzed and sorted into SLBR-N^{high} and SLBR-N^{low} cell populations. Shown is one representative plot of FACS analysis and sorting. **B)** HMLER SLBR-N^{high} and SLBR-N^{low} RNA was analyzed by qPCR for GCNT1, ST3GAL2, ST3GAL5, and ST3GAL6. Shown are the means ± SEM of a representative experiment of three independent replicates. **C)** HCC1806 cells were analyzed and sorted into SLBR-N^{high} and SLBR-N^{low} cell populations. Shown is one representative plot of FACS analysis and sorting. **D)** HCC1806 SLBR-N^{high} and SLBR-N^{low} RNA was analyzed by qPCR for GCNT1, ST3GAL2, ST3GAL5, and ST3GAL6. Shown are the means ± SEM of a representative experiment of three independent replicates. **E)** HCC1806 SLBR-N^{high} and SLBR-N^{low} RNA was analyzed by qPCR for CDH1 and CDH2. Shown are the means ± SEM of three independent replicates. **F)** PDX cells isolated from a TNBC tumor were analyzed for SLBR-H and SK678 binding by flow cytometry. The lectin+ and lectin- populations were then assessed for expression of CD44 and CD24. Shown is one biological replicate. **G)** TE3 cells were stably transfected with control shRNA (shCtrl) or two shRNA clones for ST3GAL6 (shST3GAL6-762 and shST3GAL6-833). Protein lysate was isolated and ST3GAL6 expression was quantified by western blotting. Shown is one replicate of three independent experiments. **H)** TE3 shCtrl and shST3GAL6 cells were assessed for SLBR-N binding by flow cytometric analysis. Shown is one replicate of three independent experiments. **I)** TE3 shCtrl and shGCNT1 cells were assessed for SLBR-H and SK678 binding by flow cytometric analysis. Shown is one replicate of three independent experiments.

Figure S3

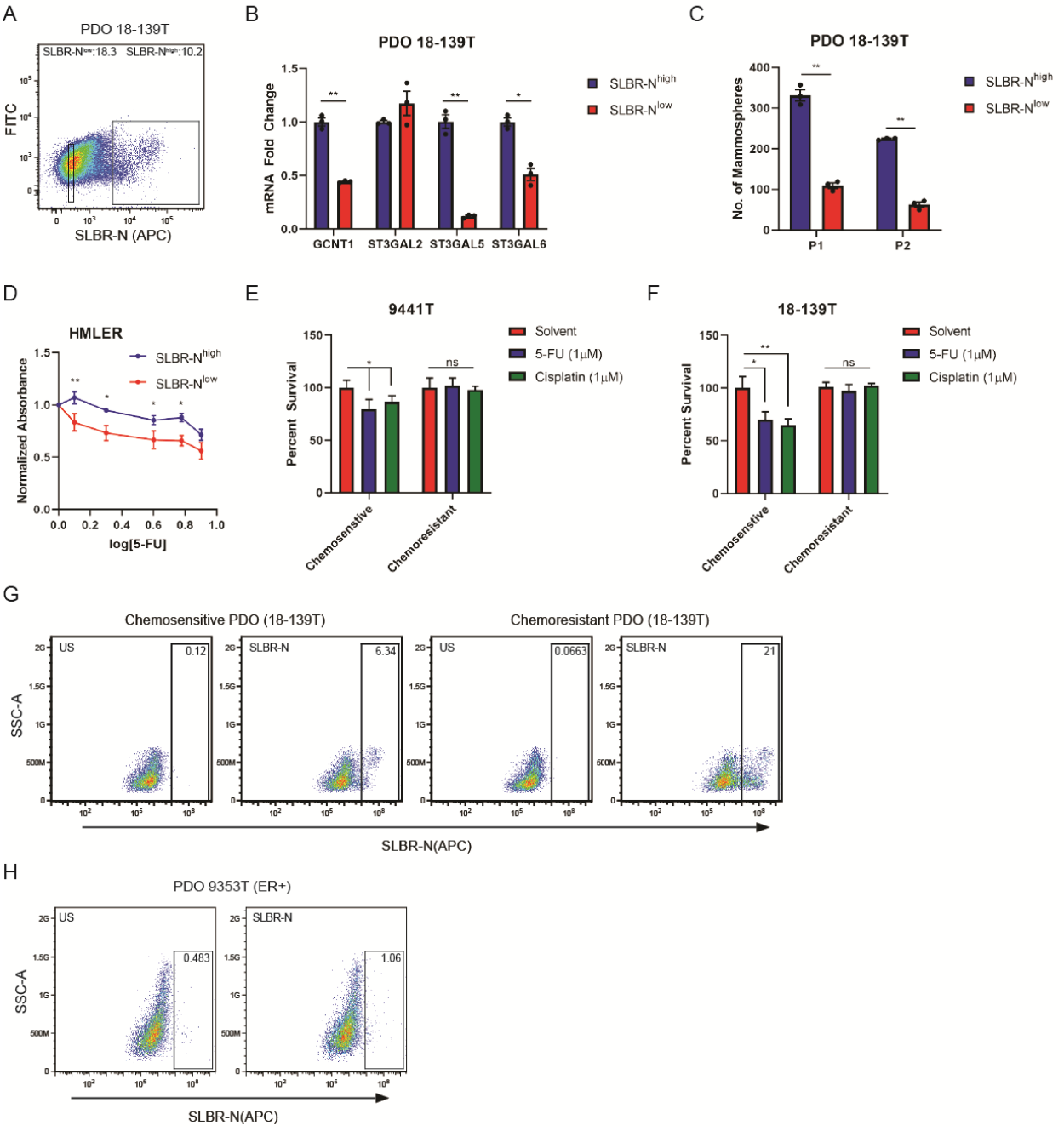

Figure S3: A) TNBC PDO 18-139T was analyzed and sorted into SLBR-N^{high} and SLBR-N^{low} cell populations. Shown is one representative plot of FACS analysis and sorting. **B)** TNBC PDO 18-139T SLBR-N^{high} and SLBR-N^{low} RNA was analyzed by qPCR for GCNT1, ST3GAL2, ST3GAL5, and ST3GAL6. Shown are the means ± SEM of three technical replicates. **C)** TNBC PDO 18-139T SLBR-N^{high} and SLBR-N^{low} cell populations were assayed for self-renewal by serial passage mammosphere formation. Shown are the means ± SEM of three technical replicates. **D)** HMLER SLBR-N^{high} and SLBR-N^{low} cells were treated with 0, 2.5, 5, 10, 25 and 50 μM of 5-FU

for 96 hours and percent surviving cells was quantified. Absorbance was normalized to DMSO control. Shown are the means ± SEM of three independent experiments. **E)** Survival data for the chemosensitive and chemoresistant PDOs (9441T) after treatment with 5-FU (1µM) and cisplatin (1µM). Shown are the means ± SEM of three technical replicates. **G)** Chemosensitive and chemoresitant patient derived organoids (PDO) from TNBC patient tumor (18-139T) were dissociated and incubated with SLBR-N for flow cytometric analysis. Shown is one biological replicate. **H)** Survival data for the chemosensitive and chemoresistant PDOs (18-139T) after treatment with 5-FU (1µM) and cisplatin (1µM). Shown are the means ± SEM of three technical replicates. **I)** Chemosensitive and chemoresitant PDOs from ER+ patient tumor (9353T) were dissociated and incubated with SLBR-N for flow cytometric analysis. Shown is one biological replicate.

Figure S4

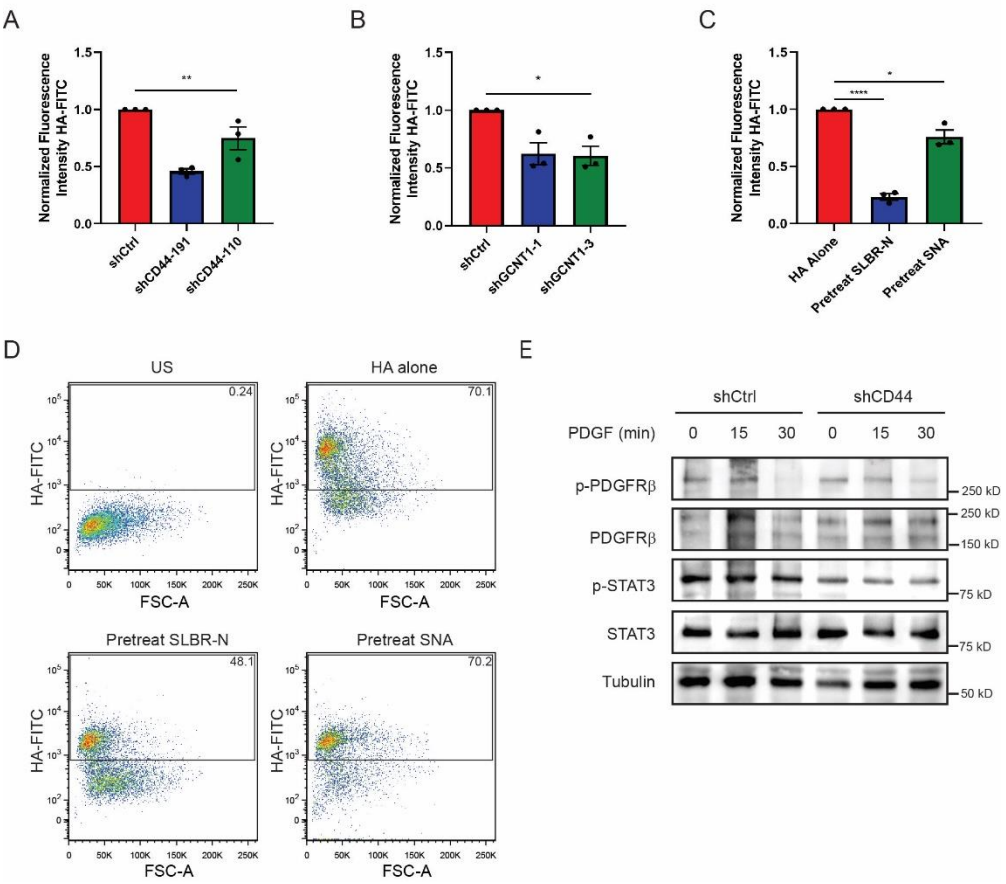

Figure S4: Quantification of HA binding experiments: A) shCD44, B) shGCNT1, C) pretreat SLBR-N, and pretreat SNA. Median fluorescence intensity was normalized to the control. Shown are the means ± SEM of three independent experiments. D) TE3 cells were serum starved for 12 hours and the affinity of the cells for HA was quantified through flow cytometric analysis with HA conjugated to FITC after no pretreatment or pretreating with SLBR-N or SNA. Shown is one replicate of three independent experiments. E) TE3-shCtrl and TE3-CD44 cells were treated with or without 10ng/mL PDGF for 0, 15, or 30 minutes. Isolated protein was assessed by immunoblotting for phospho-PDGFRβ (Tyr1009), total PDGFRβ, phospho-STAT3 (Tyr705), STAT3, and Tubulin. Shown is one replicate of three independent experiments.

Figure S5

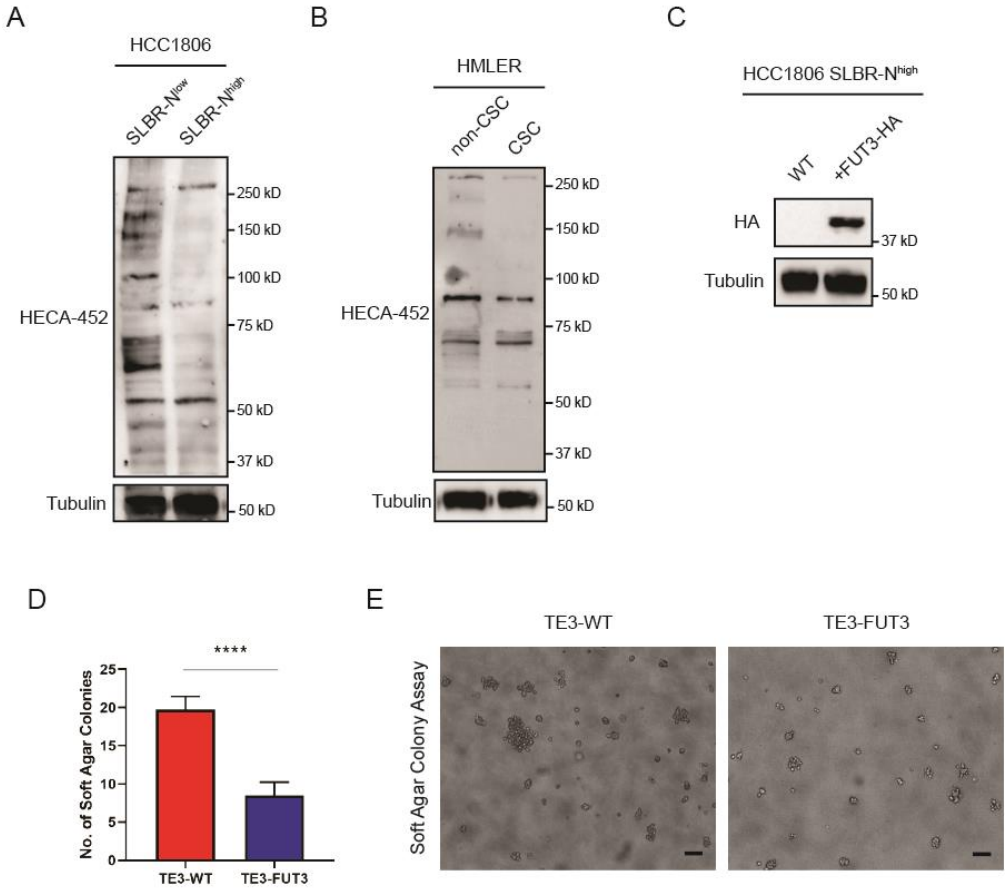

Figure S5: A) HMLER CSC and non-CSC whole cell lysates were assessed for N-linked sLeX expressing glycoproteins by immunoblotting for HECA-452 and Tubulin. Shown is one replicate of three independent experiments. **B)** HCC1806 SLBR-N^{high} and SLBR-N^{low} whole cell lysates were assessed for N-linked sLeX expressing glycoproteins by immunoblotting for HECA-452 and Tubulin. Shown is one replicate of three independent experiments. **C)** HCC1806 SLBR-N^{high} and SLBR-N^{high}+FUT3-HA whole cell lysates were assessed for FUT3 overexpression by immunoblotting for HA and Tubulin. Shown is one replicate of two independent experiments. **D)** Quantitation of soft agar colony formation assay from TE3-WT and TE3-FUT3 cells. Shown is one replicate of three independent experiments. **E)** Representative images of soft agar colony formation assay from TE3-WT and TE3-FUT3 cells. Scale bar = 200µm.

Table S1. Lectin Microarray Information

	Description
1. Sample: Glycan-containing sample (e.g. glycan, glycoprotein, cell lysate etc.)	
Description of Sample	Glycoproteins on the cell membrane extracted from non-cancer stem cells and cancer-stem cells
Sample preparation protocol	Spin cells at 500 × g for 5 mins at 4°C to remove media. If it is a frozen pellet, suspend in 1× PBS (pH 7.4) and spin down at 500 × g for 5 mins at 4°C. Suspend in 1× PBS (pH 7.4) with protease inhibitors (1:100 dilution). Sonicate the solution on ice with a Vial-Tweeter at 70% power (5s on, 10s off, total: 1 min). Spin down any large debris at 500 × g for 5 mins at 4°C. Transfer supernatant to a 3mL ultra-centrifuge tube. Ultra-centrifuge at 100,000 × g for 1 hour at 4°C. Carefully decant the supernatant and resuspend the pellet in 200uL 1× PBS (pH 7.4).
Labeling protocol for sample detection	Samples are labelled with Alexa Fluor 555-NHS (Thermo Fisher).
Two-color reference (if used)	A pooled reference samples are labelled with Alexa Fluor 647-NHS (Thermo Fisher).
Assay protocol	Lectin microarrays are blocked with blocking buffer for one hour at room temperature. Slides are rinsed twice with PBST (0.005%) and once with PBS, then dry the slide using a slide spinner. Each slide was mounted on a 24-well format hybridization cassette (Arrayit), in which each well contains a subarray. To each well, add equal amounts of samples and universal reference, and dilute with PBS and PBST (0.2%) to reach the final volume (150uL). Incubate the slides on an orbital shaker for two hours at room temperature in the dark. After hybridization, wash the arrays with PBST (0.005%) twice for ten minutes, and twice for five minutes. Once finished, remove the slides from the cassette, and immerse the slides in ultrapure water, and dry the slides using a slide spinner.
2. Lectin Library	
General description of the lectin library used in the array	Lectin microarrays are generated in house.

List of lectins and glycan binding proteins, source, concentration and buffer	Please see **Table S2**.
Modification of lectins (e.g. biotin) if any.	N/A
3. Immobilization Surface; e.g., Microarray Slide	
Immobilization surface	Nexterion Slide H Barcoded 3D Hydrogel Coated
Manufacturer	Schott North America
Custom preparation of surface	N/A
4. Array Production	
Description of Arrayer	Nano-Plotter 2.1 piezoelectric printer (GeSim, Germany) with cooled microwell plate holder and cooled printing deck
Lectin deposition	Three replicates of each lectin are printed onto each subarray.
Printing conditions	Dilute lectins to the pre-determined concentrations in the print buffer (final concentration of print buffer: 0.01% Tween-20, 1mM monosaccharide in PBS; Please see **Table S2** for the concentrations of lectins). Load the mixed solution to the microplate. Before printing, check the humidity of the print chamber. The humidity should be kept around 50% during the entire printing. Ensure both microwell plate holder and printing deck are cooled. Adjust the cooling temperature based on ambient temperature and the temperature of the cooled slide deck surface, preventing moisture building up inside the print chamber. Once printing is complete, allow the slides to dry for at least one hour.
Array layout	For each microarray, it contains 24 subarrays (3 columns and 8 rows). In each subarray, triplicates of a lectin are printed, and for a row with five lectins, the spot layout should be 15 columns. The row number depends on how many lectin probes are printed on the arrays (i.e., 110 lectins require 22 rows).
Quality control	Well-characterized glycoproteins including fetuin, asialofetuin and RNase B are used for quality assurances of the printed microarrays.
5. Detector and Data Processing	

Instrument (scanner, flow cytometer)	Fluorescent Slide Scanner Genepix 4300A (Molecular Devices)
Instrument settings	Preview the slide to adjust photomultiplier gain (PMT) for each channel (Alexa Fluor-555: 532nm, Alexa Fluor-647: 635nm) so that the signals are not saturated and within the linear detection range.
Image analysis software	GenePix Pro 7 (Molecular Devices)
Data processing and statistical analysis	Extracted data is processed for quality checks using Grubbs outlier test with $\alpha = 0.05$. Log_2 values of the average signals are median-normalized over the individual subarray in each channel.
6. Lectin Microarray Data Presentation	
Data presentation and interpretation	Hierarchical clustering of the processed data is performed using Pearson Correlation coefficient, and visualized with Multi-experiment Viewer (MeV, v4.8, TM4 Microarray Software Suite). If a lectin's SNR (signal-to-noise ratio) < 3 for more than one third of the total samples, then this lectin is considered as inactive and excluded from the list. *P*-values are calculated using nonparametric statistical tests, which are generated by R (v3.6.1).
7. Data Location	
Data Location	Synapse.org (doi:10.7303/syn26451619)

Table S2. Lectins used in microarrays

Lectin	Species/Origin	Print Conc. (µg/mL)	Rough Specificity /Inhibitory monosaccharide	Vendor/Source
AAL	*Aleuria aurantia*	1000	Fucose	Vector
ACA	*Amaranthus Caudatus*	1000	Gal-β1,3-GalNAc	Vector
AIA	*Artocarpus integrifolia*	500	β1,3-GalNAc	Vector/EY
AMA	*Allium moly*	500	Oligo mannose	EY
Anti-B.G.H2	MAb mouse IgM [A46-B/B10]	undiluted	Blood group H2 antigen	Santa Cruz Biotechnology
Anti-Forssman	MAb Rat IgM [117C9]	undiluted	Forssman Antigen	Abcam
Anti-Lewis A	MAb mouse IgG [7LE]	undiluted	Lewis A	Abcam
Anti-Lewis B	IgM [T218]	undiluted	Lewis B	Sigma
Anti-Lewis X	MAb mouse IgM [P12]	undiluted	Lewis X	Abcam
Anti-Lewis Y	MAb mouse IgM [F3]	undiluted	Lewis Y	Abcam
Anti-MUC5AC human	Mab mouse IgG1 [CLH2]	undiluted	human MUC5AC	Sigma
Anti-MUC5AC mouse	Goat polyclonal to mouse MUC5AC	undiluted	mouse MUC5AC	LSBio
Anti-Mucin 15	Mab mouse IgG1 [H-5]	undiluted	Mucin 15	Santa Cruz Biotechnology
AOL	*Aspergillus oryzae*	1000	Fucose	TCI America
APA	*Abrus precatorius*	500	Gal-β1,3-GalNAc / Lac	EY
ASA	*Allium sativum*	1000	Mannose	EY
Blackbean	*Blackbean crude*	1000	GalNAc	EY
BPA	*Bauhinia purpurea*	500	β-Gal / β-GalNAc	Vector
BR6	*Unknown*	500	Unknown	Gift from Dr. Barbara Bensing
CA	*Colchicum autumnale*	1200	Bi-antennary N-linked glycans	EY
Calsepa	*Calystegia sepium*	1000	Bisecting N-linked glycans	EY
Cholera Toxin	*Vibrio cholerae*	1000	GM1 ganglioside	Sigma
Con A	*Canavalia ensiformis*	1000	Tri-mannose core	EY/Vector
CSA	*Cystisus scoparius*	1000	Terminal GalNAc	EY
DBA	*Dolichos Biflorus*	1000	GalNAc	Vector
diCBM40	engineered NanI from *Clostridium perfringens*	1000	α Sialylation	Generated in house

DSA	*Datura stramonium*	500	LacNAc	EY/Vector
ECA	*Erythrina cristagalli*	1000	LacNAc	Vector
EEL/EEA	*Eunonymus europaeus*	1000	Blood Group B	Vector/EY
GafD	recombinant GafD from *Escherichia coli*	1000	GlcNAc	Generated in house
GNA/GNL	*Galanthus nivalis*	1500	Oligo mannose	Vector/EY
GS-I	*Griffonia simplicifoia-I*	1000	α-Gal / Lac	Vector/EY
GS-II	*Griffonia simplicifoia-II*	1000	GlcNAc	Vector
GS-IB4	*Griffonia simplicifoia-I, isolectin B4*	2000	Gal	Vector
H84T	*Banana lectin*	1000	High mannose	Gift from Dr. David Markovitz
HAA	*Homarus americanus*	1000	Terminal GalNAc	EY
HHL	*Hippeastrum Hybrid*	1500	Oligo/High mannose	Vector
HPA	*Helix pomatia*	1000	Blood Group A	Sigma/EY
LAA	*Laburnum alpinum*	900	GlcNAc	EY
LcH	*Lens Culinaris*	1000	Core Fucose	Vector
LEA/LEL	*Lycopersicon esculentum*	1000	GlcNAc	Vector/EY
Lotus	*Lotus tetragonolobus*	1000	Fucose	Vector
MAL-I	*Maackia amurensis-I*	2000	Sialylation/Sulfation	Vector
MAL-II	*Maackia amurensis-II*	2000	Sialylation/Sulfation	Vector
MNA-G	*Morus nigra Morniga G*	1000	GalNAc	EY
MNA-M	*Morus nigra Morniga M*	1000	Oligo mannose / Gal	EY
MPA/MPL	*Maclura pomifera*	1000	β1,3-GalNAc	Vector
NPA	*Narcissus pseudonarcissus*	1000	Oligo mannose	Vector
PA-I	*Pseudomonas aeruginosa*	1000	Gal	Sigma
PHA-E	*Phaseolus vulgaris Erythroagglutinin*	1000	Bisecting GlcNAc	Vector/EY/Sigma
PHA-L	*Phaseolus vulgaris Leukoagglutinin*	1000	β1,6 Branching N-Link glycans	Vector/EY/Roche
PNA	*Arachis hyogaea*	1000	Gal-β1,3-GalNAc	Vector/EY
PSA	*Pisum sativum*	1000	Core Fucose	Vector
PTA	*Psophocarpus tetragonolobus*	500	Blood Groups	EY

PTL-I	*Psophocarpus tetragonolobus-I*	1500	Blood Group A	Vector
PTL-II	*Psophocarpus tetragonolobus-II*	1000	α2 Fucose	Vector
RCA120	*Ricinus Communis Agglutinin I*	1000	Gal / Lac	Vector
rGRFT	*recombinant Griffithsin*	1000	High mannose	Gift from Dr. Barry O'Keefe
Ricin B Chain	*Ricinus communis*	1000	Gal	Vector
RPA	*Robinia pseudoacacia*	500	Complex N-link glycans	EY
rSVN	*recombinant Scytovirin*	1000	High mannose	Gift from Dr. Barry O'Keefe
SBA	*Glycine max*	1000	LacdiNAc	Vector
SJA	*Sophora japonica*	1000	LacdiNAc	Vector
SK1	*Streptococcus sanguinis SK1*	1800	α2,3 sialylation	Gift from Dr. Barbara Bensing
SK678	*Streptococcus sanguinis SK678*	450	α2,3 sialylation	Gift from Dr. Barbara Bensing
SLBR-B	*Streptococcus gordonii M99*	1000	α2,3 sialylation	Gift from Dr. Barbara Bensing
SLBR-H	*Streptococcus gordonii DL1*	2000	α2,3 sialylation	Gift from Dr. Barbara Bensing
SLBR-N	*Streptococcus gordonii UB10712*	1000	α2,3 sialylation	Gift from Dr. Barbara Bensing
SNA	*Sambucus nigra*	500/1000	α2,6 sialylation	Vector/Sigma
SNA-II	*Sambucus nigra-II*	1000	α2 Fucose /oligo mannose	EY
STA/STL	*Solanus tuberosum*	500	GlcNAc	Vector
TJA-I	*Trichosanthes japonica-I*	1000	α2,6 sialylation	TCI
TJA-II	*Trichosanthes japonica-II*	1000	α2 Fucose	NorthStar Bioproducts/Aniara Diagnostica
TL	*Tulipa sp.*	700	GlcNAc	EY
UDA	*Urtica dioica*	1000	GlcNAc / Oligo mannose	EY
UEA-I	*Ulex europaaeus-I*	1000	α2 Fucose	Vector
UEA-II	*Ulex europaaeus-II*	2000	GlcNAc	Vector
VFA	*Vicia faba*	1000	GlcNAc	EY
VVA	*Vicia villosa*	1000	Terminal GalNAc	Vector/EY
VVA (man)	*Vicia villosa*	500	Mannose	Vector/EY
WFA	*Wisteria floribunda*	1000	GalNAc-β1,4	Vector
WGA	*Triticum vulgare*	1000	GlcNAc	Vector/EY

Table S3. **Mass spectrometry results of SLBR-N enriched glycoproteins expressed in the HMLER CSC population.**

Accession Number	Description	ID	Total #PSM in HMLER CSCs	M.W.
O00425	IF2B3_HUMAN Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	24	64 kDa
O60716-10	CTND1_HUMAN Isoform 2AB of Catenin delta-1	CTNND1	17	101 kDa
P01130-3	LDLR_HUMAN Isoform 3 of Low-density lipoprotein receptor	LDLR	26	77 kDa
P02786	TFR1_HUMAN Transferrin receptor protein 1	TFRC	4	85 kDa
P05556-2	ITB1_HUMAN Isoform 2 of Integrin beta-1	ITGB1	4	87 kDa
P08648	ITA5_HUMAN Integrin alpha-5	ITGA5	7	115 kDa
P16070-10	CD44_HUMAN Isoform 10 of CD44 antigen	CD44	32	53 kDa
P16144-2	ITB4_HUMAN Isoform Beta-4A of Integrin beta-4	ITGB4	2	195 kDa
P23470-2	PTPRG_HUMAN Isoform 2 of Receptor-type tyrosine-protein phosphatase gamma	PTPRG	19	159 kDa
P29317	EPHA2_HUMAN Ephrin type-A receptor 2	EPHA2	11	108 kDa
P35221-2	CTNA1_HUMAN Isoform 2 of Catenin alpha-1	CTNNA1	6	103 kDa
P49327	FAS_HUMAN Fatty acid synthase	FASN	86	273 kDa
P63244	RACK1_HUMAN Receptor of activated protein C kinase 1	RACK1	16	35 kDa
Q08431	MFGM_HUMAN Lactadherin	MFGE8	12	43 kDa
Q13751	LAMB3_HUMAN Laminin subunit beta-3	LAMB3	7	130 kDa
Q6NZI2	CAVN1_HUMAN Caveolae-associated protein 1	CAVIN1	11	43 kDa
Q6YHK3-2	CD109_HUMAN Isoform 2 of CD109 antigen	CD109	4	153 kDa
Q9NZI8	IF2B1_HUMAN Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	13	63 kDa
Q9Y6M1-1	IF2B2_HUMAN Isoform 2 of Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	30	62 kDa