SUPPLEMENTAL MATERIAL

Supplemental Methods

Proteomic Sample Preparation. Human breast tissue biopsies were flash frozen in liquid nitrogen and powderized using a ceramic mortar and pestle. Tissue was dried overnight in a lyophilizer and weighed tissue (approximately 1 mg of each) was homogenized in freshly prepared high-salt buffer (50 mM Tris-HCl, 3 M NaCl, 25 mM EDTA, 0.25% w/v CHAPS, pH 7.5) containing 1x protease inhibitor (Halt Protease Inhibitor, Thermo Scientific) at a concentration of 10 mg/mL. Homogenization took place in a bead beater (Bullet Blender Storm 24, Next Advance, 1 mm glass beads) for 3 min at 4 ºC. Samples were then spun for 20 min 18,000 x g at 4 ºC, and the supernatant removed and stored as the cellular fraction. A fresh aliquot of high-salt buffer was added to the remaining pellet at 10 mg/mL of the starting weight, vortexed at 4 ºC for 15 min, and spun for 15 min. The supernatant was removed and stored as Fraction 2. This high-salt extraction was repeated once more to generate Fraction 3, after which freshly prepared guanidine extraction buffer (6 M guanidinium chloride adjusted to pH 9.0 with NaOH) was added at 10 mg/mL and vortexed for 1 hour at room temperature. The samples were then spun for 15 min, the supernatant removed, and stored as Fraction 4. Fractions 1, 2, & 3 (Cellular) were combined and all fractions were stored at -80 ºC until further analysis. The remaining pellets of each tissue representing insoluble ECM proteins were treated with hydroxylamine.

Hydroxylamine (NH_2OH) Treatment. Following Fraction 4, insoluble protein pellets were treated with freshly prepared hydroxylamine (NH_2OH) buffer (1 M NH_2OH-HCl, 4.5 M guanidine-HCl, 0.2 M K_2CO_3, pH adjusted to 9.0 with NaOH) at 10 mg/mL of the starting tissue weight, as previously described. The samples were prepared as previously described (1). Briefly, pellets were mixed with NH_2OH buffer at 45 ºC with vortexing for 16 hours. Following incubation, the samples were spun for 15 min at 18,000 x g, the supernatant removed, and stored as Fraction 5 at -80 ºC until further proteolytic digestion with trypsin. The final pellet was stored at -80 ºC until further analysis.

Trypsin Digestion. 100 µL of the Cellular fraction (combined fractions 1, 2 and 3) 200 µL of the Fraction 4 & 5 of all samples were subsequently subjected to reduction, alkylation, and enzymatic digestion with trypsin. 100 fmols of each SIL peptide (170 peptides total) were spiked into 100 µL of sample to allow for four injections per sample (50 fmols eQ 1-6 per injection) (2, 3). A filter-aided sample preparation (FASP) approach, as well as C18 cleanup, was performed as previously described (4).

LC-SRM Analysis. Samples were analyzed by LC-SRM and LC-MS/MS as described (3). Briefly, equal volumes from each post-digestion sample were combined and injected every third run and used to monitor technical reproducibility. Skyline was used for method development and to extract the ratio of endogenous light peptides to heavy internal standards from LC-SRM data for protein quantification as described (5). LC–MS/MS data was processed as previously described (4). Limits of detection, quantification, and dynamic range were determined for each peptide as previously described (2). Final fmol values are expressed as fmol/mg where milligrams represent milligrams of starting dry tissue weight.

LC-MS/MS Analysis. Samples were analyzed on an Q Exactive HF mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-nanoLC 1000 system through a nanoelectrospray source. 8 µL of sample was injected. The analytical column (100 µm i.d. × 150 mm fused silica capillary packed in house with 4 µm 80 Å Synergi Hydro C18 resin (Phenomenex; Torrance, CA)) was then switched on-line at 600 nL/min for 10 min to load the sample. The flow rate was adjusted to 350 nL/min, and peptides were separated over a 120-min linear gradient of 2–40% ACN with 0.1% FA. Data acquisition was performed using the instrument supplied Xcalibur™ (version 2.1) software. The mass spectrometer was operated in positive ion mode. Full MS scans were acquired in the Orbitrap mass analyzer over the 300–1800 m/z range with 60,000 resolution at m/z 400. Automatic gain control (AGC) was set at 5E+05 and the ten most intense peaks from each full scan were fragmented via HCD with normalized collision energy of 35. MS2 spectra

were acquired in the Orbitrap mass analyzer with 15,000 resolution. All replicates of each tissue were run sequentially and pre-digested yeast alcohol dehydrogenase standard (nanoLCMS Solutions LLC, Rancho Cordova, CA) was run between tissue groups to monitor drift in analytical performance.

Proteomic Data Analysis. Skyline was used for method development and to extract the ratio of endogenous light peptides to heavy internal standards from targeted LC- SRM data for protein quantification as described (4). Global LC–MS/MS data was processed as previously described (6). Limits of detection, quantification, and dynamic range were determined for each peptide as previously described (6). Principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) were performed using the MetaboAnalyst with sum and range scaling normalizations (7).

Immunohistochemistry, Image Acquisition, and Analysis. Sections were then incubated overnight at 4^{o}C with primary antibodies specific to α-smooth muscle actin (α-SMA, Cell Signaling Technology, clone D4K9N, catalog #: 19245, 1:500), fibroblast activation protein (FAP, Santa Cruz, clone C-20, catalog #: sc-1138, 1:100), lysyl hydroxylase 2 (LH2, Santa Cruz, clone N-15, catalog #: sc-50067, 1:200) and lysyl oxidase (LOX, abcam, catalog #: ab174316, 1:500). Sections were washed with PBS and incubated with Impress anti-rabbit or anti-goat reagent (Vector laboratories, catalog #: MP-7451 and MP7405) prior to development of positive staining signal with Impress DAB reagent (Vector Laboratories, catalog #: SK-4105) and counterstaining with hematoxylin. Stained tissue slides were digitally scanned using Zeiss Axio ScanZ.1 and visualized using Zen 2.6 software (Zeiss). To score tissue composition, total tissue area and regions within this tissue area representing epithelium, stroma and fat were manually defined and presented as a percentage of the total area. For scoring of TDLUs, the total number of TDLUs and the average number of acini per TDLU were counted for each patient sample. To score positive antibody staining, the intraductal lobular regions were manually defined within a tissue section. For α-SMA, LH2 and LOX-positive staining, the stroma within each intraductal lobular region were scored as positive or

negative and presented as a percentage of positive-regions per patient sample. For FAP-positive cells, QuPath 0.1.2 was used to automatically count the percent positive cells within each defined intraductal lobular region and presented as the average percent positive cells in these regions per patient sample. Other primary antibodies included phospho-Histone-H3 (Cell Signaling Technology, catalog#: 9701, 1:200), ZNF217 (Sigma-Aldrich, catalog#: HPA051857, 1:400) and phospho-Akt substrate (RXXS*/T*) (110B7E) (Cell Signaling Technology, catalog#: 9614, 1:200). Brightfield images were acquired using the 40x air (Plan-Apo, 0.95 NA), 20x air (Plan-Apo, 0.75 NA), or 10x air (Plan-Apo 0.45 NA) objectives of a Zeiss AxioImager A1 brightfield microscope or the 4x air (0.13 PhL), 10x air (0.3 PhC) or 40x air (0.6 Ph2) objectives of an Olympus IX81 microscope. Quantification of positive staining was performed by counting the percentage of DAB positive cell nuclei per total nuclei in a defined epithelial ductal region for each human specimen. For mouse specimens, the percentage of high positive (ZNF217, phospho-histone H3) nuclei per epithelial cell was determined using the IHC profiler plugin for ImageJ. For phospho-Akt substrate staining, total high positive staining for epithelial regions was determined using the IHC profiler plugin for ImageJ. At least 3 individual fields were quantified for each specimen and included in an averaged analysis.

Immunofluorescence, Image Acquisition, and Analysis. Primary antibodies used were specific for Collagen12a1 (E15; Santa Cruz Biotechnology, catalog#: sc-68449, 1:200), Lumican (H-90; Santa Cruz Biotechnology, Catalog#: sc-33785, 1:200), and Cytokeratin 8+18 (Fitzgerald, catalog#: 20R-CP004, 1:400). 4′,6-diamidino-2-phenylindole (DAPI) was used to stain cell nuclei. IF staining of mouse organoids cultured in collagen gels was performed using the same procedure as for human tissue sections except that 2% PFA was used for 20mins fixation. Additional antibodies used include: phospho-Akt substrate (RXXS*/T*) (110B7E) (Cell Signaling Technology, catalog#: 9614, 1:200), phospho-Histone-H3 (Cell Signaling Technology, catalog#: 9701, 1:400), and Alexa Fluor 488 Phalloidin (Thermo Fisher Scientific, catalog#: A12379, 1:1000). Images of IF stained sections and MEC organoids were acquired

using a Nikon Eclipse Ti Spinning Disk confocal or an Andor Borealis CSU-W1 spinning disk confocal with 4 laser lines (405, 488, 561, and 640 nm) equipped with either a standard Andor Zyla sCMOS cameras or one for widefield imaging. Quantification of human tissue sections was performed by thresholding images at a fixed intensity and measuring the average integrated density using ImageJ. Quantification of positive staining in mammary epithelial cell spheroids was performed by manually counting total and phospho-Histone H3 positive nuclei and taking an average. For phospho-Akt substrate staining the corrected total cell fluorescence was calculated using Image J (8). At least 4 images of each human tissue specimen or MEC spheroid condition were used for averaged analyses.

Mammography. All mammography was conducted by licensed physicians at the UCSF medical center according to established protocols. Quantitative measurements of the raw mammography images used the automated volumetric density measure developed by Dr. John Shepherd (9). For X-ray imaging of whole breast mastectomies, the mastectomy specimen was placed on a compression station and attenuation images were taken at 25kVp and 39kVp using a Hologic Selenia Digital Mammography system. These images were compared to calibration data taken using a custom breast mimicking phantom and tissue thickness and composition images were generated using the calibration data. The specimen was then transported to a histology laboratory and sliced into sections approximately 5mm in thickness. These sections were labeled and imaged using dual energy absorptiometry (DXA) as described (10). Using the section labels and the measured breast density, breast composition was reconstructed in three dimensions. Regions of high and low density were selected from Egan slices for further analysis.

REFERENCES

1. Barrett AS*, et al.* Hydroxylamine Chemical Digestion for Insoluble Extracellular Matrix Characterization. *J Proteome Res.* 2017;16(11):4177-84.
2. Hill RC, Calle EA, Dzieciatkowska M, Niklason LE, and Hansen KC. Quantification of extracellular matrix proteins from a rat lung scaffold to provide a molecular readout for tissue engineering. *Mol Cell Proteomics.* 2015;14(4):961-73.
3. Johnson TD*, et al.* Quantification of decellularized human myocardial matrix: A comparison of six patients. *Proteomics Clin Appl.* 2016;10(1):75-83.
4. Goddard ET*, et al.* Quantitative extracellular matrix proteomics to study mammary and liver tissue microenvironments. *Int J Biochem Cell Biol.* 2016;81(Pt A):223-32.
5. MacLean B*, et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26(7):966-8.
6. Laklai H*, et al.* Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. *Nat Med.* 2016;22(5):497-505.
7. Xia J, Sinelnikov IV, Han B, and Wishart DS. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res.* 2015;43(W1):W251-7.
8. Burgess A, Vigneron S, Brioudes E, Labbe JC, Lorca T, and Castro A. Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. *Proc Natl Acad Sci U S A.* 2010;107(28):12564-9.
9. Shepherd JA*, et al.* Volume of mammographic density and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2011;20(7):1473-82.
10. Shepherd JA, Herve L, Landau J, Fan B, Kerlikowske K, and Cummings SR. Clinical comparison of a novel breast DXA technique to mammographic density. *Med Phys.* 2006;33(5):1490-8.

Supplemental Tables and Figures

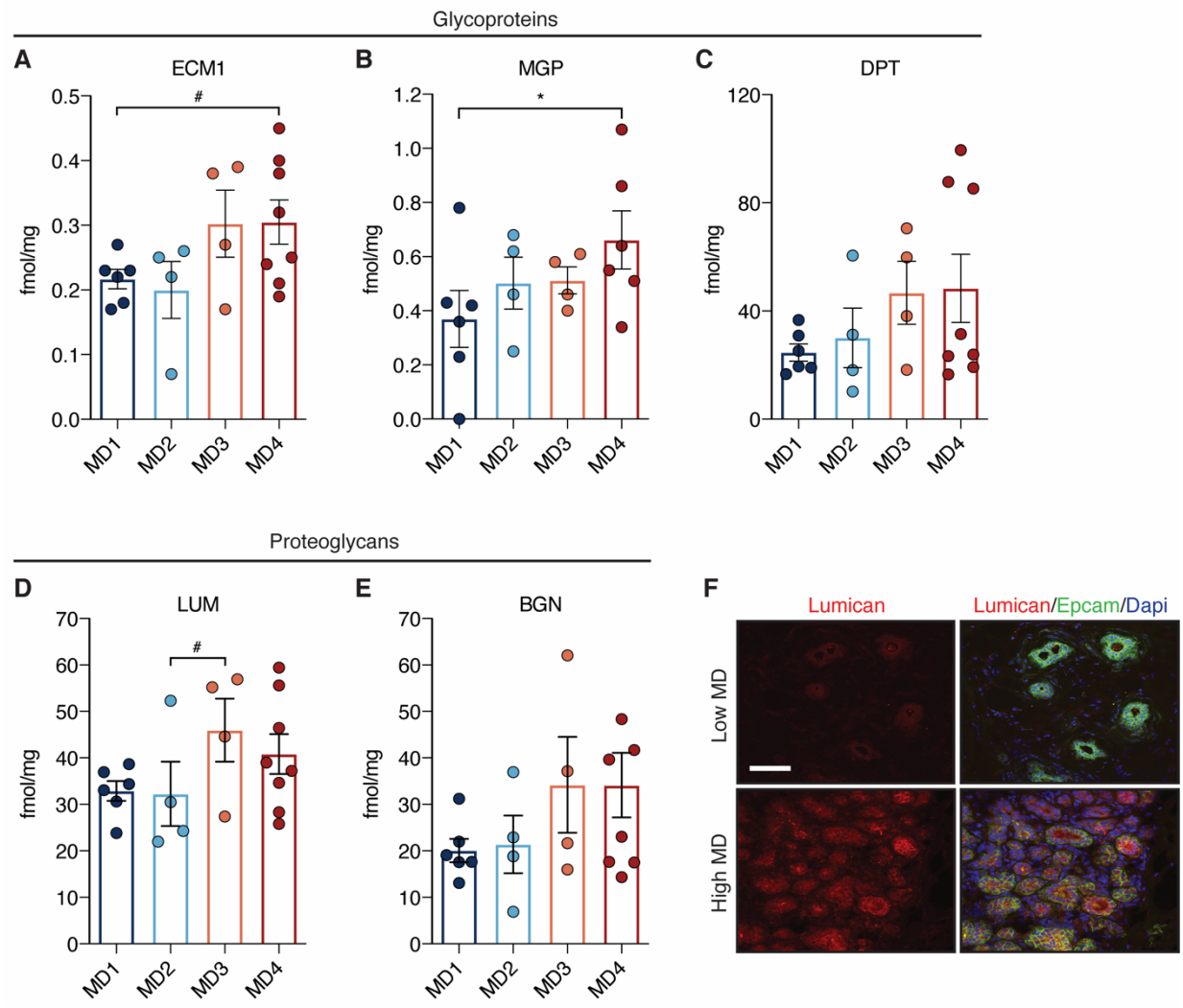
Table S1. Prophylactic mastectomy tissues for proteomic analysis.

#	B#	MD	Age	BRCA	Menopause	Specimen Type
1	B1931	4	32	none	pre	prophylaxis, IDC, DCIS contralateral
2	B1959	4	45	2	pre	prophylaxis
3	B2006	2	36	N/A	pre	prophylaxis, adenocarcinoma
4	B2038	4	42	N/A	pre	IDC same breast
5	B2048	1	40	none	pre	prophylaxis, IDC, contralateral
6	B2058	2	44	none	pre	prophylaxis, DCIS, contralateral
7	B2066	3	47	2	pre	prophylaxis, IDC, DCIS, contralateral
8	B2076	4	34	2	pre	prophylaxis, IDC, DCIS, contralateral
9	B2085	4	37	none	pre	prophylaxis
10	B2158	1	27	1	pre	prophylaxis
11	B2227	4	27	2	pre	prophylaxis
12	B2242	1	38	1	pre	prophylaxis, contralateral MD6
13	B2252	1	40	2	pre	prophylaxis
14	B2300	4	40	none	pre	prophylaxis
15	B2315	1	36	2	pre	prophylaxis
16	B2316	1	25	1	pre	prophylaxis
17	B2325	4	34	2	pre	prophylaxis
18	B2335	2	42	none	pre	prophylaxis
19	B2484	3	41	1	pre	prophylaxis
20	B2509	3	45	2	pre	prophylaxis
21	B2510	3	38	2	pre	prophylaxis
22	B2530	2	38	none	pre	prophylaxis

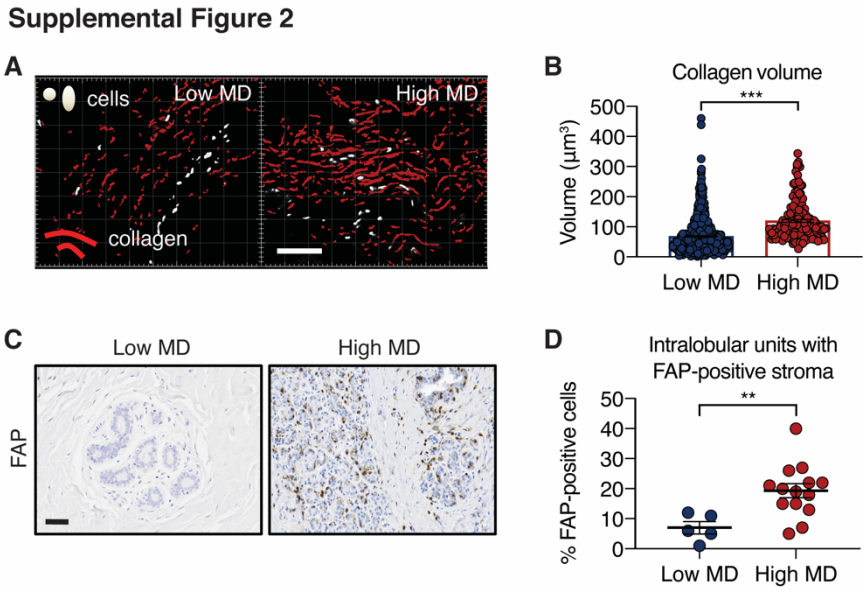
Table S2. Summary of proteomic data derived from human breast specimens.

******(attached as a separate excel file)******

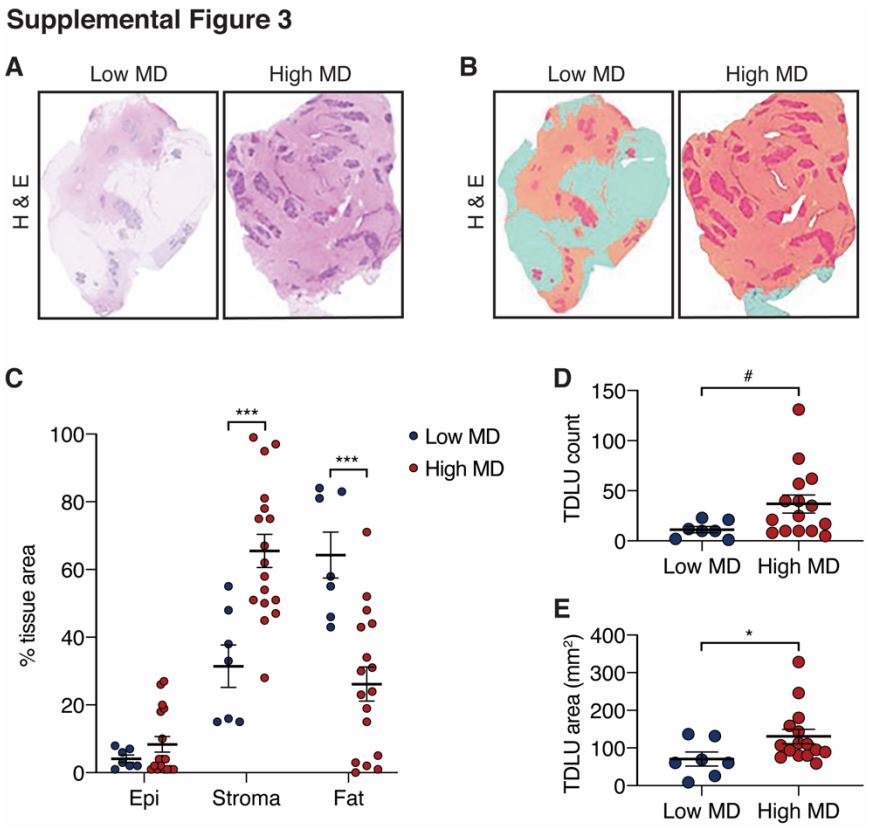
Supplemental Figure 1



Supplemental Figure 1. Expression of several glycoproteins and proteoglycans are elevated in tissues with high breast density. (**A, B and C**) Quantification of measurements for ECM1, MGP1 and DPT from the proteomic analysis of human breast tissues (MD1-MD4; see Figure 1). (**D and E**) The same as in A for the proteoglycans LUM and BGN. (**F**) Immunofluorescence staining of frozen breast tissue sections with antibodies specific for EpCAM (green) and Lumican (LUM; red) to validate an increase in LUM levels in tissues with high MD (Low MD, n=5; High MD, n=5 total). Cell nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI; blue). Scale bar, 50μm. Data are represented as mean ± S.E.M. $^{\#}P<0.1$; $^{*}P<0.05$; Kruskal-Wallis test followed by Dunn's multiple comparisons test (A-E).



Supplemental Figure 2. Tissues with high mammographic density have thicker collagen fibers and increased numbers of intralobular fibroblasts. (**A**) Representative images of Second Harmonic Generation (SHG) Imaging to visualize the collagen of human breast tissues. Red color identifies thicker collagen fibers. Scale bar, 100µm. (**B**) Quantification of collagen volume determined from Imaris image rendering (Low MD, n=5; High MD, n=5). (**C**) Immunohistochemical (IHC) staining of human breast tissues with low and high MD with an antibody specific for fibroblast activated protein (FAP). Scale bar, 100µm. (**D**) Quantification of positive IHC staining for C (Low MD, n=5; High MD, n=14). Data are represented as mean ± S.E.M. ***P*<0.01; ****P*< 0.001, 2-tailed Mann-Whitney U test (B); 2-tailed unpaired *t* test (D)



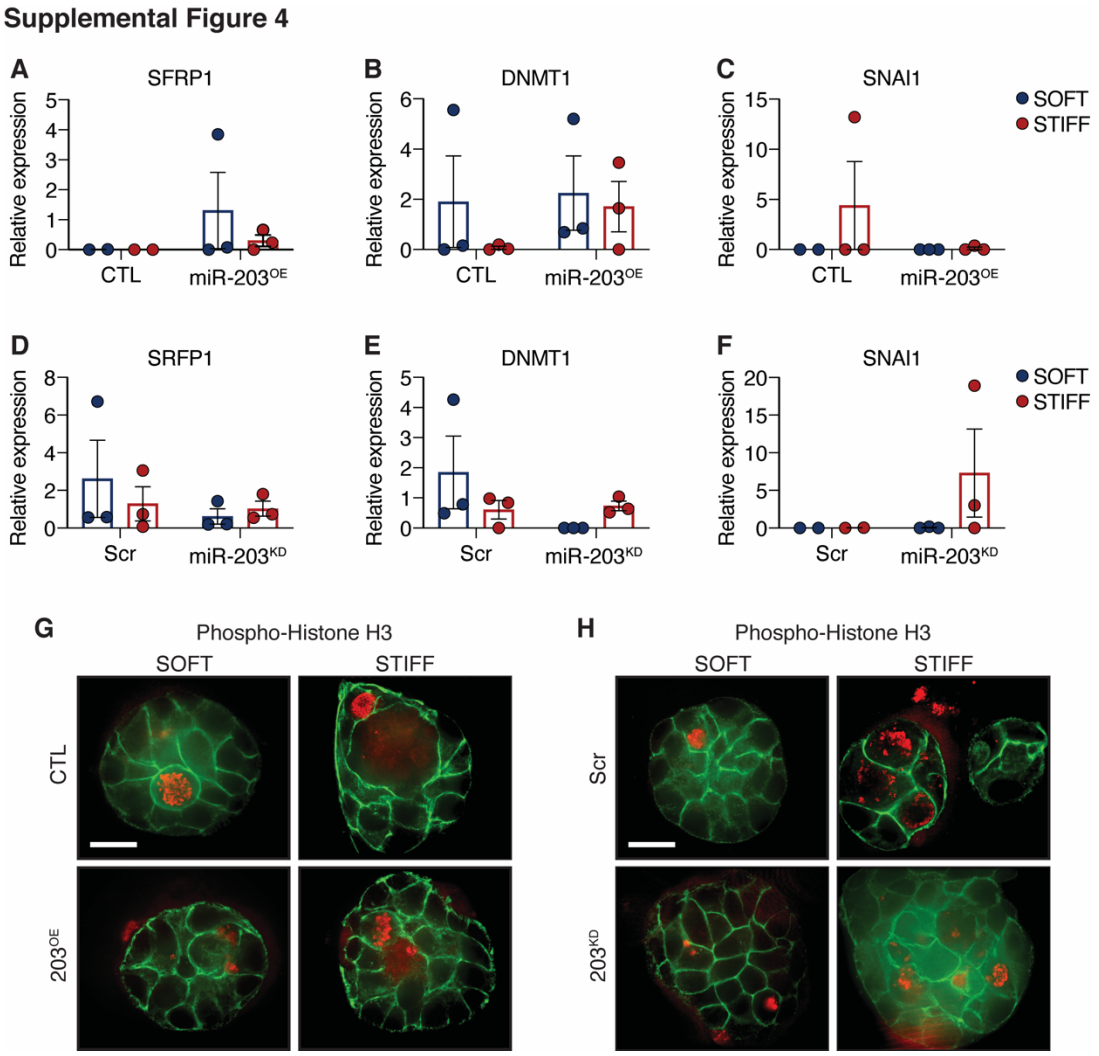
Supplemental Figure 3. Tissues with high mammographic density have greater epithelial and stromal density. (**A**) Whole tissue section scans of Hematoxylin and Eosin (H&E) staining for human breast tissues with low and high MD. (**B**) Selection of regions by training program in QuPath corresponding to epithelium (pink), stroma (orange) and fat (teal). (**C**) Quantification of the average tissue area which corresponds to epithelium (epi), stroma, and fat for human mammary tissues with high and low MD (Low MD, n=7; High MD, n=15 for C-E). (**D**) Quantification of the average number of terminal-ductal lobular units (TDLUs) from the same tissues analyzed in A-C. (**E**) Quantification of average TDLU area using the same tissues analyzed in A-C. Data are represented as mean ± S.E.M. **P*<0.05; ***P*<0.01; ****P*< 0.001, two-way ANOVA with Holm-Sidak's multiple comparisons test (C); 2-tailed Mann Whitney U test (D and E).

Table S3. Short list of stiffness regulated microRNAs with the potential to influence breast cancer risk.

microRNA	Genomic Location	Potential involvement in breast cancer	References (PMID)
miR-29b	7q32.3 1q32.2	Inhibits fibrosis, collagen remodeling, proliferation; promotes migration	30323900; 29422258; 28365400; 23354167; 21359530
miR-144	17q11.2	Inhibits proliferation, migration and invasion, epithelial-to-mesenchymal transition; may inhibit fibrosis	29561704; 27785072; 26097586; 30745456
miR-18a	13q31.3	Candidate biomarker of breast cancer risk; inhibits drug resistance; promotes proliferation and invasion; downregulates tumor suppressors; impairs DNA damage response	27789587; 27338043; 24633304; 21980462
miR-149(*)	2q37.3	Inhibits cell proliferation and metastasis; can have both tumor promoting and suppressing roles	29270025; 24608434; 29344284
miR-203	14q32.33	Promotes differentiation/inhibits stemness; inhibits proliferation, epithelial-to-mesenchymal transition, angiogenesis, invasion and metastasis; inhibits proliferation of ER-positive breast cancer cells	28962215; 27517632; 26194864; 22713668; 22514743; 22393463; 26975850

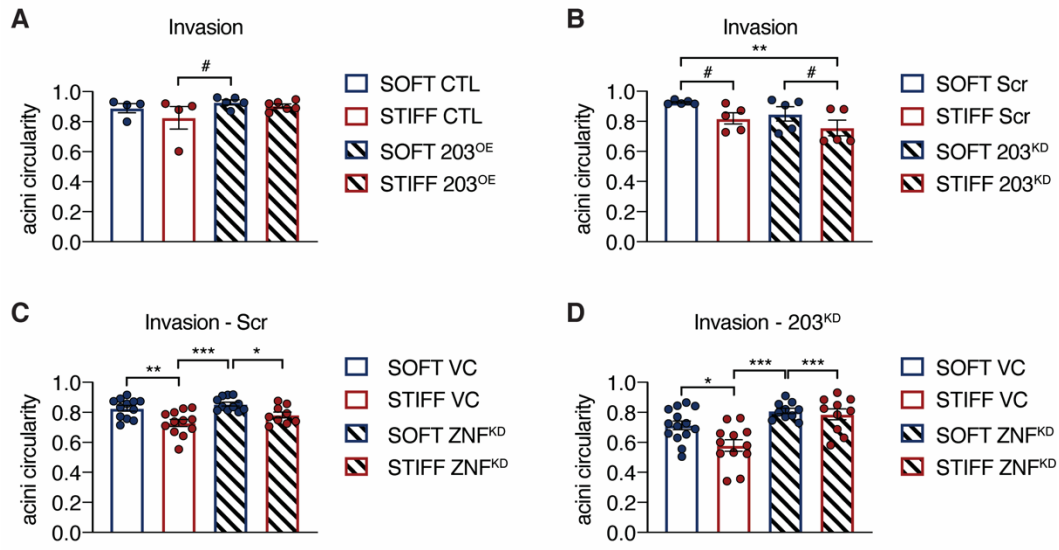
Table S4. miRTarBase compiled list of experimentally validated miR-203 targets.

******(attached as a separate excel file)******



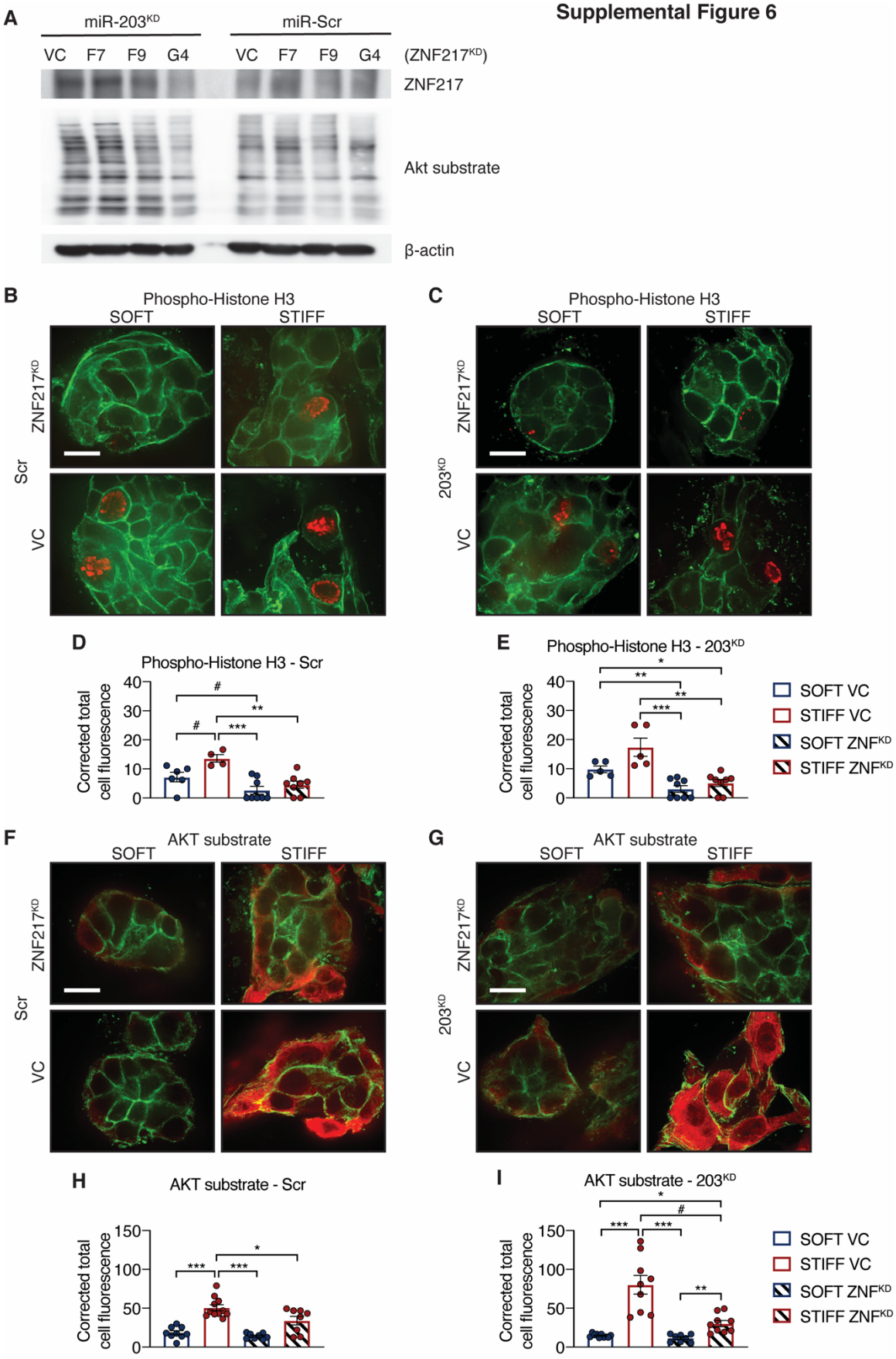
Supplemental Figure 4. microRNA-203 levels regulate mammary epithelial cell proliferation and ZNF217 activity in a stiffness-dependent manner. (**A-C**) qRT-PCR analysis for known ZNF217 regulated transcripts including *SFRP1* (**A**), *DNMT1* (**B**) and *SNAI1* (**C**) using RNA isolated from MCF10A acini expressing a control microRNA (CTL) or miR-203 (203^{OE}) that were matured in basement membrane (BM) extract for 7-10 days before embedding in SOFT (non-cross-linked) or STIFF (L-ribose cross-linked) BM/collagen gels (n=2-3 replicates each). (**D-F**) The same as in A-C using RNA isolated from MCF10A acini overexpressing a scrambled non-targeting antagomir (Scr) or an antagomir targeting miR-203 (203^{KD})(n=2-3 replicates each). (**G**) Representative images of MCF10A acini (as in Figure 6H) fixed and stained by immunofluorescence with antibodies specific to phosphorylated-Histone H3 (Phospho-Histone H3) and phalloidin-488 to stain Actin filaments. (**H**) Representative images of MCF10A acini (as in Figure 6I) stained by immunofluorescence as in G. Scale bar for all panels, 20μm. Quantification of results are presented in Figure 6. Data are represented as mean ± S.E.M.

Supplemental Figure 5



Supplemental Figure 5. microRNA-203 and ZNF217 levels regulate mammary epithelial acini collagen gel invasion. (**A**) MCF10A acini overexpressing a control microRNA (CTL) or miR-203 (203^{OE}) were matured in basement membrane (BM) extract for 7-10 days before embedding in SOFT (non-cross-linked) or STIFF (L-ribose cross-linked) BM/collagen gels. Acini were stained by immunofluorescence using phalloidin-488 to visualize the Actin cytoskeleton and confocal images were acquired for analysis. Acini borders were then traced, and their circularity was measured to determine the extent of acini invasion into the surrounding collagen matrix (n=4 replicates for all panels). (**B**) The same as in A but for MCF10A acini overexpressing a scrambled non-targeting antagomir (Scr) or an antagomir targeting miR-203 (203^{KD}). (**C**) MCF10A Scr cells were transduced with shRNA targeting *ZNF217* and cultured and analyzed as in A and B. (**E**) MCF10A 203^{KD} cells were transduced with shRNA targeting *ZNF217* mRNA and cultured and analyzed as in A and B. Data are represented as mean ± S.E.M. $^{\#}P<0.1$; $^{*}P<0.05$; $^{**}P<0.01$; $^{***}P< 0.001$, Kruskal-Wallis test followed by Dunn's multiple comparisons test (A-D).

Supplemental Figure 6



Supplemental Figure 6. ZNF217 knockdown abrogates matrix stiffness induced and miR-203 dependent increases in mammary epithelial cell proliferation and Akt activity. MCF10A acini overexpressing a scrambled non-targeting antagomir (Scr) or an antagomir targeting miR-203 (203^{KD}) were transduced with 3 different shRNAs DNA constructs targeting *ZNF217* (F7, F9, and G4) or a vector control construct (VC). Protein lysates were prepared and subjected to immunoblot analysis using antibodies specific to ZNF217, phosphorylated-Akt substrate (phospho-Akt substrate) or β-actin. (**B**) Representative images of MCF10A acini that were cultured in SOFT (non-cross-linked) and STIFF (L-ribose cross-linked) basement membrane (BM)/collagen gels and subsequently fixed and stained by immunofluorescence with antibodies specific to phosphorylated-Histone H3 (Phospho-Histone H3) and phalloidin-488 to stain Actin filaments. MCF10A Scr acini were overexpressing an shRNA targeting ZNF217 (G4) or a vector control (VC). (**C**) The same as in B for MCF10A 203^{KD} acini overexpressing an shRNA targeting ZNF217 (G4) or a vector control (VC). Average phospho-Histone H3 positive nuclei over total nuclei were assessed using DAPI nuclear counterstain and plots in (**D**) and (**E**) correspond to the representative images in B and C respectively (n=3 replicates). (**F**) The same as in B showing representative images of MCF10A acini cultured in SOFT and STIFF BM/collagen gels that were subsequently fixed and stained by immunofluorescence with antibodies specific to phospho-Akt substrate and phalloidin-488 to stain Actin filaments. (**G**) Representative images of MCF10A cells prepared as in C and stained by immunofluorescence as in F. Scale bar for all panels, 20μm. Average corrected total cell fluorescence (see methods) for phospho-Akt substrate staining was assessed and plots in (**H**) and (**I**) correspond to the representative images in F and G respectively (n=3 replicates). Data are represented as mean ± S.E.M. $^{\#}P<0.1$; $*P<0.05$; $**P<0.01$; $***P< 0.001$, Kruskal-Wallis test followed by Dunn's multiple comparisons test (D, E, H and I).