Supplementary Figure and Table Legends:

Supplementary Figure S1: Image quantitation. (A) Fibroblasts were placed under proliferative or adipocyte differentiation conditions for 7 days and assessed for fat accumulation by Oil Red O staining. Ten 10X magnification bright field images were taken of each treatment group and the amount of fat accumulation (Oil Red O staining) was quantitated, on a cell-by-cell basis, using nuclear segmentation and color decomposition. This is demonstrated for a bright field image taken at 10X magnification of RMF7 under differentiation conditions and stained with Oil Red O (left panel), where the nuclear membranes are shown in green, cell membranes in yellow, and associated Oil Red O staining in red (right panel). **(B)** Fibroblasts were grown on glass slides for 5 days before their ability to accumulate various matrix proteins (collagen 1A1, fibronectin, osteopontin and tenascin C) was assessed by immunofluorescence. For each of the matrix proteins examined, for each of the fibroblast populations, seven 10X magnification fluorescent images were quantitated on a cell-by-cell basis using nuclear and protein signal segmentation. This is demonstrated for a fluorescent image (left panel) taken at 10X magnification of fibroblasts immunostained for collagen 1A1 (COL1A1), where the unstained background is shown in black, the nuclei in blue and the associated COL1A1 staining in red (right panel). **(C)** Paraffin sections of the #4 mammary glands from 10 week old virgin C57BL/6 mice, either wild type for Cd36 (n=7) or knocked out for Cd36 (n=5), were stained with Masson's Trichrome. Ten 10X magnification bright field images were taken of each gland, and the area of fat cells, on a cell-by-cell basis, and the amount of

matrix accumulation, on an image-by-image basis, was quantitated using nuclear segmentation and color decomposition. This is demonstrated for a bright field image taken at 10X magnification (left panel), where the cytoplasm is shown in red, the collagen in bright green and the fat in blue (right panel). **(D)** Paraffin sections of tissues were stained for CD36 (brown staining). Ten 20X magnification bright field images were taken of each slide. The amount of staining was quantitated for each image, on a cell-by-cell basis, using nuclear segmentation and color decomposition. This is demonstrated for a bright field image, taken at 20X magnification, of cancer-free tissue stained for CD36 (left panel), where the extracellular matrix is shown in green, the nuclei in blue, and associated CD36 staining in red (right panel).

Supplementary Figure S2: HDAFs accumulate less fat than LDAFs. Ten 10X magnification images from each of the 6 treatment groups (3 LDAFs and 3 HDAFs under growth (ctl) or adipocyte differentiation conditions (+PJ2)) stained with Oil Red O were quantitated using nuclear segmentation and color decomposition, on a cell-by-cell basis. Shown are histograms of the average number of red pixels (Oil Red O staining) per cell. This analysis demonstrates that the average amount of fat accumulation per cell is decreased in the HDAFs when compared to the LDAFs, due to a decrease in both the percentage of cells that contain fat and the amount of fat per cell.

Supplementary Figure S3: HDAFs accumulate more COL1A1, FN1 and OPN, but less TNC, than LDAFs. 6 LDAFs and 6 HDAFs were grown on glass slides for 5 days before their ability to accumulate matrix proteins was assessed by immunofluorescence. Shown are representative fluorescent images taken at 10X magnification of each of the 6 LDAFs and 6 HDAFs stained for **(A)** collagen 1A1 (COL1A1), **(B)** fibronectin (FN1), **(C)** oseteopontin (OPN) and **(D)** tenascin C (TNC).

Supplementary Figure S4: Modulation of CD36 expression in RMFs. (A) Quantitation of CD36 QPCR data for RMF8 **(Supplementary Table S2)** infected with either a lentivirus expressing short hairpin RNA to luciferase (shLuc) or a lentivirus expressing short hairpin RNA to CD36 (shCD36). Each bar represents the average of 4 determinations, with error bars showing the standard error of the mean. This analysis confirmed a 4.2-fold decrease in CD36 mRNA expression in shCD36 cells when compared to shLuc cells. **(B)** Quantitation of CD36 QPCR data for RMF2 **(Supplementary Table S2)** infected with either an empty lentivirus (vector) or a lentivirus expressing CD36 (CD36 OE). Each bar represents the average of 3 determinations, with error bars showing the standard error of the mean. This analysis confirmed a 1.7-fold increase in CD36 mRNA in CD36 OE cells when compared to vector cells. * denotes a p value of <0.001 and # denotes a p value of <0.007. p values were calculated using a two-sided t-test.

Supplementary Figure S5: CD36 antibody staining in tissue is specific. (A) Representative bright field images taken at 10X magnification of various disease-free human tissues stained for CD36 as described in Materials and Methods. As reported, ovary, pancreas, liver, thyroid, lymph node, cerebral cortex and placenta tissues were all negative for CD36 staining, while heart and breast tissues were positive for CD36 staining. **(B)** Representative bright field images taken at 10X magnification of breast tissue stained for CD36 as described in Materials and Methods, with (right panel) or without (left panel) primary antibody.

Supplementary Figure S6: Reduced CD36 expression is due to a decrease in both the percentage of CD36-positive cells and the intensity of CD36 staining per cell. (A) Paraffin sections of tissue from 13 low density (LD) and 14 high density (HD) cancer-free women **(Supplementary Table S5)** were stained for CD36. The amount of CD36 staining was quantitated, on a cell-by-cell basis, from ten 20X magnification images from each section. Shown are histograms of the average amount of CD36 staining per cell for the high MD tissues (top panel) and low MD tissues (bottom panel). **(B)** Paraffin sections of tissues from 20 women with IDC **(Supplementary Table S6)** were stained for CD36. The amount of CD36 staining was quantitated, on a cell-by-cell basis, from ten 20X magnification images of the tumor tissue (CA) and the histologically normal adjacent tissue (NA) of each patient. Shown are histograms of the average amount of CD36 staining per cell for the tumor tissues (top panel) and normal adjacent tissues (bottom panel). **(C)** Paraffin sections of tissue from 21 LD and

14 HD ER+ IDC patients **(Supplementary Table S8)** were stained for CD36. The amount of CD36 staining was quantitated, on a cell-by-cell basis, from ten 20X magnification images of the tumor tissue (CA) from each patient. Shown are histograms of the average amount of CD36 staining per cell for the high MD tumor tissues (top panel) and low MD tumor tissues (bottom panel). These analyses demonstrate that the decrease in the average amount of CD36 staining per cell was, in all cases, due to a decrease in both the percentage of cells that stain positive for CD36 and the intensity of CD36 staining per cell.

Supplementary Figure S7: CD36 expression is decreased in multiple cellular compartments of desmoplastic tissues compared to normal adjacent tissues, in all tumor subtypes. (A) Paraffin sections of tissues from 20 women with IDC **(Supplementary Table S6)** were stained for CD36. The amount of CD36 staining was quantitated, on a cell-by-cell basis, from ten 20X magnification bright field images of the tumor tissue (CA) and the histologically normal adjacent tissue (NA) of each patient. Shown are average and standard error of the mean of CD36 signal per cell in the NA and CA tissues of all tumor subtypes pooled together. **(B)** Left-most panels: bright field images taken at 20X magnification of adipocytes in the CA and NA tissue of one woman with an ER+ tumor, stained for CD36. Right panels: bright field images taken at 10X magnification of serial paraffin sections with CA and NA tissue from the same woman, stained for CD36, CD31 and CD68.

Supplementary Figure S8: Breast map from patient BM1. Whole mastectomies, obtained from three women with invasive cancer **(Supplementary Table S7)**, were bread-loafed into ~20 mm thick slices and multiple blocks were taken from each slice for processing and embedding. 5 μm sections of each of these blocks were H&E stained and reviewed by a pathologist to generate histology maps of the entire mastectomy for each woman. Shown is a representative breast map from patient BM1.

Supplementary Table S1: LDAFs and HDAFs sample and patient information. Low density associated fibroblasts (LDAFs) and high density associated fibroblasts (HDAFs) were purified from cancer-free patients. %MD refers to percent mammographic density. Passage indicates the number of times the fibroblasts were passaged in culture before cells were harvested and processed to conduct the microarray experiments presented in **Supplementary Table S3** and the QPCR experiments presented in **Figure 3A**. Un designates unknown information.

Supplementary Table S2: RMFs and CAFs patient and sample information. Carcinoma-associated fibroblasts (CAFs) were purified from women who had undergone surgery for invasive cancer and reduction mammoplasty fibroblasts (RMFs) were purified from disease-free women who had undergone surgery for reduction mammoplasty. Passage indicates the number of times the fibroblasts were passaged in culture before cells were harvested and processed for the

QPCR experiments presented in **Figure 3C**. Un designates unknown information.

Supplementary Table S3: Microarray data from LDAFs and HDAFs. Gene expression was assessed in 7 LDAFs and 7 HDAFs **(Supplementary Table S1)** using Affymetrix HU133 plus 2 chips (GEO GSE38506). The dataset was pre-processed, and differential analysis of gene expression was performed, using the "*affyPLM*" package and a linear model available through the *limma* R library, respectively, from the Bioconductor Project. Fold change refers to the fold difference in gene expression between the LDAFs and HDAFs, where a positive number designates a gene whose expression is higher in HDAFs relative to LDAFs, while a negative number refers to a gene whose expression is lower in HDAFs relative to LDAFs.

Supplementary Table S4: CD36 expression in published microarray datasets. CD36 expression levels were compared between normal breast tissue and invasive ductal carcinoma (IDC) tissues in published microarray datasets.

Supplementary Table S5: Cancer-free patient information. All the patients in this table were diagnosed as cancer-free. Note: Diagnoses listed outside the parentheses are from the slide that was used for quantitation, while the diagnoses within the parentheses are from other slides from the same patient that were not used for quantitation.

Supplementary Table S6: Invasive ductal carcinoma patient information. All patients in this table were diagnosed with invasive ductal carcinoma (IDC) and had tissue blocks containing tumor tissue as well as normal adjacent tissue. ER, PR, and HER2 refer to the estrogen receptor, progesterone receptor and HER2/neu receptor status of the tumors, respectively. ER, PR and HER2 status was assessed by immunohistochemistry (IHC). HER2 status was also assessed by FISH when IHC results were equivocal. Grade refers to the tumor grade. BIRADS refers to the Breast Imaging Reporting and Data System classification for MD: 1 = almost entirely fat, 2 = scattered fibroglandular densities, 3 = heterogeneously dense and 4 = extremely dense.

Supplementary Table S7: Breast maps patient information. All patients in this table were diagnosed with invasive cancer. IDC refers to invasive ductal carcinoma and ILC refers to invasive lobular carcinoma. ER, PR, and HER2 refer to the estrogen receptor, progesterone receptor and HER2/neu receptor status of the tumors, respectively. ER, PR and HER2 status was assessed by immunohistochemistry (IHC). HER2 status was also assessed by FISH when IHC results were equivocal. Stage refers to the tumor stage at the time of surgery.

Supplementary Table S8: Low and high MD invasive ductal carcinoma patient information. All patients in this table were diagnosed with estrogen

receptor positive (ER+), invasive ductal carcinoma (IDC) and had tissue blocks containing tumor tissue. Grade refers to the tumor grade. BIRADS refers to the Breast Imaging Reporting and Data System classification for MD: 1 = almost entirely fat, 2 = scattered fibroglandular densities, 3 = heterogeneously dense and 4 = extremely dense. LD refers to low mammographic density (BIRADS 1) and HD refers to high mammographic density (BIRADS 4).

Supplemental Table S9: CD36 expression is inversely correlated with tumor grade and size. Results of linear regression models used to test the association between CD36 expression levels and clinical characteristics of tumors. In univariate models, CD36 probe 228766 was used as the dependent variable to determine the association with tumor grade, tumor size or age. The coefficients can be interpreted as the approximate change in CD36 per standard deviation in CD36 gene expression level associated with a unit change in the predictor variable.

Supplemental Materials and Methods

Isolation and propagation of human mammary fibroblasts: Fibroblasts were isolated from reduction mammoplasty tissues, low and high mammographic density tissues, and tumor tissues and assessed for purity as previously described (1). Briefly, the tissue specimens were minced and digested with 225 units/ml collagenase type I (Sigma-Aldrich Corp., St. Louis, MO) and 125 units/ml hyaluronidase (Sigma-Aldrich Corp.) in RPMI-1640 with L-glutamine, 25 mM Hepes, 10% FBS and a cocktail of antibiotics and fungicides (Mediatech Inc., Manassas, VA) at 37°C overnight. After several washes, the digest was passed through a 50 μm filter to separate the fibroblasts (flow through fraction) from the epithelial organoids (retained fraction). A portion of the fibroblastic fraction was immediately frozen in DMSO containing medium for future use, while the remaining was placed in RPMI-1640 with L-glutamine, 25 mM Hepes and 10% FBS, a medium that promotes preferential fibroblastic outgrowth. We will refer to this medium as "growth medium" in the following sections. Staining with antibodies to alpha smooth muscle actin (Dako Inc., Carpinteria, CA, #MO851, dilution 1:100), vimentin (Dako Inc. #MO725, dilution 1:100), fibronectin (BD Biosciences, Franklin Lakes, NJ, #610077, dilution 1:100) and E-cadherin (BD Biosciences, #610181, dilution 1:100) was used to characterize and determine the purity of fibroblastic fractions.

Preparation of total RNA, cDNA and biotinylated mRNA for microarray and quantitative polymerase chain reaction (QPCR) analysis: Total RNA was

isolated from either primary human mammary fibroblasts **(Supplementary Tables S1 and S2)** or primary human mammary fibroblasts subjected to adipocyte differentiation, using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen Inc., Valencia, CA). For the Q-PCR experiments, 2 μg of total RNA was used to generate cDNA using TaqMan Reverse Transcription Reagents according to the manufacturer's instructions (Applied Biosystems, Foster City, CA). For the microarray experiments, 1 μg of total RNA from LDAFs and HDAFs **(Supplementary Table S1)** was used to amplify and biotinylate mRNA using the Ambion's MessageAmp II Kit according to the manufacturer's instructions (Ambion, Austin, TX). Briefly, total RNA was first reverse transcribed into double stranded cDNA using oligo(dT) primers, purified, then transcribed and amplified into mRNA using biotinylated and unmodified dNTPs. The amplified and biotinylated mRNA was then fragmented to an average size of 50-100 bases. The quality of the total RNA and amplified mRNA, pre- and post-fragmentation, was analyzed using the Agilent 2100 BioAnalyzer and RNA 6000 Nano Assay according to the manufacturer's instructions (Agilent, Palo Alto, CA).

LDAFs and HDAFs microarray hybridization and analysis: The biotinylated and fragmented mRNA described above was hybridized to Affymetrix HU133 plus 2 chips, which contain approximately 47,000 gene probes (Affymetrix, Santa Clara, CA). The microarray hybridizations and scans were performed by the Gladstone Genomics Core Laboratory (UCSF, San Francisco). The dataset was

pre-processed using the Bioconductor package "*affyPLM*". Processing steps included: (i) background correction; (ii) normalization using the quantile method (2); and (iii) summarization of probe set values using the RMA (Robust Multi-array Average) method, which fits a specified robust linear model to the probe level data (3). Several diagnostic plots were generated for validation. This microarray dataset was deposited with GEO under accession # GSE38506. Differential analysis of gene expression was performed with a linear model available through the *limma* R library from the Bioconductor Project. From this analysis, 20 probe sets were found to be differentially expressed between LDAFs and HDAFs with a p value of <0.05 **(Supplementary Table S3)**.

Q-PCR analysis: The cDNA described above was used to perform Q- PCR (TaqMan) using the standard curve method with primer probe sets for CD36 (Hs00169627_m1) and leptin (Hs00174877_m1) from Applied Biosystems. The expression of beta-D-glucuronidase (GUSB), used as control to normalize for variances in input cDNA, was monitored with a primer probe set from IDT (Coralville, IA). The forward, reverse and primer sequences for GUSB were: 5'-CTCATTTGGAATTTTGCCGATT -3',
5'- CCGAGGAAGATCCCCTTTTTA -3', 5'- FAM-TGAACAGTCACCGACGAGAGTGCTGGTA-TAM -3', respectively. Q-PCR conditions were as follows: 1 cycle at 50°C for 7 minutes; 1 cycle at 95°C for 10 minutes; 50 two-step cycles (95°C for 15 seconds; 60°C for 1 minute); 1 cycle at

72°C for 5 minutes. p values for all QPCR data analysis were calculated using a two-sided t-test.

Adipocyte differentiation assay: Fibroblasts were plated in growth medium at near confluency (4 x 10^4 cells per well of 4 well chamber slides for Oil Red O staining and either 3 x 10^5 cells per 60 mm dish or 7 x 10^5 cells per 100 mm dish for Q-PCR analysis). Fresh growth medium was added the next day and the cells were allowed to recover for another 24 hours before the cells were treated with either growth medium alone, or growth medium supplemented with 10 μM 15-deoxy-$\Delta^{12,14}$-Prostaglandin J2 (Cayman Chemical, Ann Arbor, MI) to induce adipocyte differentiation. Media were replaced every 2 days for the entire length of the experiment.

Lipid accumulation was assayed by Oil Red O staining at the indicated time points. Briefly, cells were rinsed in 1X PBS, fixed in 10% formalin for 10 minutes at room temperature and then rinsed again in 1XPBS. Cells were then dehydrated in 50% aqueous propylene glycol for 3 minutes and 100% propylene glycol for 2 minutes before being stained overnight in 0.5% Oil Red O in 100% propylene glycol. Cells were then destained in 85% aqueous propylene glycol 3 times for 1 minute and then for 10 minutes in running tap water. Cells were finally counterstained with hematoxylin to visualize the nuclei.

To assess adipocyte differentiation, cells were treated for 2 weeks as described above. Total RNA was then collected and cDNA synthesized for Q-PCR analysis. Expression of CD36 and leptin, two genes previously shown to be up-regulated during adipocyte differentiation (4), was monitored.

Matrix accumulation: Fibroblasts were plated in growth medium at near confluency (1.5×10^4 cells per well) into 8 well glass chamber slides. Fresh growth medium was added the next day and cells were allowed to deposit matrix for 4 additional days. Cells were then fixed for 10 minutes at room temperature in paraformaldehyde (Electron Microscopy Sciences, Hatfiled, PA) diluted to 4% in 1XPBS, then further fixed and permeabilized in a methanol gradient (100%, 75%, 50% and 25% diluted in 1XPBS) for 5 minutes each, at -20° C. Cells were blocked in ChemMate Antibody Dilution Buffer (Ventana Medical Systems Inc., Tucson, AZ) for 1 hour at room temperature then incubated with primary antibodies diluted in ChemMate Antibody Dilution Buffer overnight at 4°C. The antibodies used to assess matrix accumulation were against collagen 1A1 (Abcam, Cambridge, UK, #34710, dilution 1:100), fibronectin (BD Biosciences, #610077, dilution 1:100), osteopontin (Abcam, #8448, dilution 1:100) and tenascin C (R&D #MAB2138, 1:100). Cells were washed 3 times in Wash Buffer (5% FBS and 0.1% Fish Gelatin in 1XPBS) for 5 minutes each, at room temperature. Cells were then incubated in the appropriate TRITC-conjugated secondary antibody (SouthernBiotech, Birmingham, AL) diluted 1:500 in ChemMate Antibody Dilution Buffer for 1 hour at room temperature. Cells were

washed 3 times in Wash Buffer for 5 minutes each at room temperature. A coverslip was then mounted on the slide with Prolong Gold with DAPI antifade (Molecular Probes). Images of matrix staining of LDAFs and HDAFs are shown in **Supplementary Figure S3**.

Western blot analysis: Total protein lysates were processed in RIPA Buffer using standard methods. 10 μg of total protein extracts were separated on gradient (4-20%) polyacrylamide gels (Lonza Group Ltd., Basel, Switzerland) and transferred onto Hybond-P membranes (GE Healthcare, Buckinghamshire, UK). Membranes were incubated for 1 hour at room temperature in blocking buffer (5% milk in Tris buffered saline Tween-20 (TBST)) before being probed overnight at 4°C with primary antibodies specific for CD36 (Novus Biologicals, Littleton, CO, NB400-145) and actin (Sigma-Aldrich Inc., A5441), diluted at 1:1,000 and 1:15,000, respectively, in blocking buffer. Membranes were then washed in TBST before being probed for an hour at room temperature with the appropriate species-specific horseradish peroxidase conjugated antibody diluted 1:5,000 in blocking buffer (Chemicon, Temecula, CA). Membranes were washed again before being incubated in SugerSignal West Pico solutions (Pierce Biotechnology Inc., Rockford, IL) and exposed to Hyperfilm ECL (GE Healthcare, Buckinghamshire, UK).

Analysis of CD36 expression in published microarray data: Analysis of CD36 expression in published microarray datasets **(Supplementary Table S4)** was

performed using Oncomine (Compendia Bioscience Inc., Ann Arbor, MI), a Web-based application for assessing target expression data against published cancer types and subtypes.

Short hairpin silencing of CD36 and CD36 over expression: For CD36 silencing, DNA oligos (forward: 5'- CCGGGCAGCAACATTCAAGTTAATTC AAGAGATTAACTTGAATGTTGCTGCTTTTTTG -3' and reverse: 5'- AATTCAAAAAAGCAGCAACATTCAAGTTAATCTCTTGAATTAACTTGAATGTT GCTGC -3') obtained from Invitrogen Corp. (Carlsbad, CA), were annealed and cloned into the lentiviral expression vector pLKO.1-puro digested with AgeI and EcoRI and de-phosphorylated with Calf intestinal Phosphatase (CIP). For CD36 overexpression, CD36 cDNA (Origene Technologies Inc., Rockville, MD) was excised from the pCMV6-XL5 vector and inserted into the lentiviral expression vector pWP1, generously provided by Dr. D. Trono (School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland), digested with PmeI and de-phosphorylated with CIP. The pLKO.1-puro and pWP1 constructs were packaged in 293T cells for viral propagation. Human mammary fibroblasts from reduction mammoplasty tissue (RMFs) were plated at near confluency and transduced 3 times with viral particles suspensions, allowing 1 day of recovery between infections. CD36 levels were decreased in RMF8 and increased in RMF2 **(Supplementary Table S2)**. pLKO.1-puro infected cells were selected with 1 μg/ml puromycin and maintained in growth medium containing this selection. Transduction efficiency of pWP1-infected cells was monitored using

GFP expression, which was driven by the same promoter as CD36 via an IRES bi-cistronic sequence.

Mouse mammary tissue preparation and immunohistochemistry: The #4 mammary glands from either wild type (n=7) or Cd36 knock out (n=5) 10 week old virgin C57BL/6 mice, were surgically removed and fixed into formalin overnight on ice. Glands were then embedded in paraffin using standard protocols.

5 μm sections were cut from the paraffin-embedded glands, de-paraffinized, re-hydrated and stained with Masson's Trichrome using standard protocols. Briefly, sections were fixed in Bouin's for 1 hour at 56^{o}C, extensively washed, then stained with Weigert's Iron Hematoxylin for 2 minutes at room temperature and washed. Sections were then stained with Biebrich Scarlet-Acid Fuchsin Solution [0.9% (w/v) Biebrich Scarlet, 0.1% (w/v) Acid Fuchsin, 1% (v/v) Acetic Acid] for 15 minutes at room temperature, rinsed with distilled water, then destained with Phosphomolybdic-Phosphotungstic Acid Solution [2.5% (w/v) of phosphomolybdic and phosphotungstic acid] for 10-15 minutes at room temperature. Sections were then counterstained with Light Green Solution [2% (w/v) Light Green, 0.8% (v/v) glacial acetic acid] for 5 minutes at room temperature before being dehydrated and cleared through xylene for 2 minutes.

4μm sections were cut from paraffin-embedded tissue blocks. Sections were de-paraffinized and re-hydrated using standard protocols. To reduce non-specific background staining due to endogenous peroxidase, sections were treated with 3% H_2O_2 for 10 minutes at room temperature. For COL1A1 and FN1 antigen retrieval, sections were microwaved for 10 minute in citrate buffer (pH=6.0) and then allowed to cool down for 20 minutes at room temperature. Sections were then incubated for 60 minutes at room temperature with a primary antibody to COL1A1 (Abcam, Ab347100) or FN1 (Epitomics, 1574-1) diluted 1:500 and 1:250, respectively, in Primary Antibody Diluent (Genmed Biotechnologies, South San Francisco, CA). Sections were washed twice then incubated for 30 minutes at room temperature with HRP Polymer (Thermo Scientific Inc.). After 2 washes, sections were developed for 5 minutes at room temperature using diaminobenzidine (DAB) as substrate. The sections were then counterstained in Mayer's hematoxylin to visualize the nuclei.

Human tissue preparation and immunohistochemistry: 4μm sections were cut from paraffin-embedded tissue blocks. Sections were de-paraffinized and re-hydrated using standard protocols. To reduce non-specific background staining due to endogenous peroxidase, sections were treated with 3% H_2O_2 for 10 minutes at room temperature. For CD36 antigen retrieval, sections were microwaved for 1 minute in EDTA buffer (pH=8.0), then allowed to cool down for 20 minutes at room temperature. For CD31 and CD68 antigen retrieval, sections were incubated for 60 minutes in Citrate buffer (pH=6.0) at 80°C, then allowed to

cool down for 20 minutes at room temperature. For CD36, CD31 and CD68 staining, sections were then incubated for 60 minutes at room temperature with a primary antibody to CD36 (Sigma-Aldrich Inc., HPA002018), CD31 (Dako Inc., M0823) or CD68 (Dako Inc., M0876) diluted 1:100, 1:20 and 1:200 respectively, in Antibody Dilution Buffer (Ventana Medical Systems Inc., Tucson, AZ). Sections were washed twice then incubated for 20 minutes at room temperature with Primary Antibody Enhancer (Thermo Scientific Inc., Fremont, CA). After 2 washes, sections were incubated for 30 minutes at room temperature with HRP Polymer (Thermo Scientific Inc.). After 2 washes, sections were developed for 5 minutes at room temperature using diaminobenzidine (DAB) as substrate. The sections were then counterstained in Mayer's hematoxylin to visualize the nuclei. Representative images of CD36 antibody tissue staining, including positive (heart and breast) and negative (ovary, pancreas, liver, thyroid, lymph node, cerebral cortex and placenta) control tissues used to assess staining specificity are shown in **Supplementary Figure S5**.

Microscopy: Ten bright field images of Oil Red O-stained cells and Masson's Trichrome-stained mouse mammary glands were taken at 10X magnification using a Nikon color CCD camera mounted to a Nikon TE2000E motorized inverted microscope. The illumination power and exposure time were maintained at the same level, and all images were automatically corrected for shading and non-uniformities against a blank slide. Seven fluorescent images of cells immunostained for COL1A1, FN1, OPN and TNC were taken at 10X

magnification using a Nikon CoolSNAP HQ^2 CCD camera mounted to a Nikon TE2000E motorized inverted microscope. The illumination power and exposure time were maintained at the same level for all images. The human and mouse tissues stained for CD36, CD31, CD68, COL1A1 and FN1 were scanned at 20X magnification using an Aperio ScanScope Digital Slide Scanner (Aperio Technologies Inc., Vista, CA). Scanned images were viewed using the Aperio's Spectrum Software and 10 regions of interest (ROI) were chosen randomly for quantitation (imaged at 5X or 20X magnification for the human tissues and 10X magnification for the mouse tissues). The non-specific staining observed within the lumen of the mouse mammary ducts with the COL1A1 and FN1 antibodies, due to the presence of "sticky" secretions, was cut out of the images prior to quantitation.

Image quantitation: Four endpoints were subjected to image analysis: (i) expression of CD36 in human mammary tissue sections, (ii) lipid accumulation in cultured human mammary fibroblasts, (iii) matrix and morphometric properties of fat cells in mouse mammary tissue sections, and (iv) expression of matrix proteins in cultured human mammary fibroblasts. In all cases, experimental factors, images, and computed morphometric and quantitative data were imported into BioSig (5). Biosig enables the user to aggregate morphometric analysis performed on a cell-by-cell basis for the entire treatment group to form probabilistic distributions, these distributions being subsequently compared using a Komogrov-Smirnov test.

For endpoints (i) and (ii) described above, the analysis protocol is outlined in our previous publication (6). Briefly, the user specifies regions of positive staining (CD36 or Oil Red O) and hematoxylin (DNA), in a subset of images. This step facilitates construction of a representation through the Gaussian Mixture Models (GMM). The main benefit of this approach to model construction is that the heterogeneity of the staining is captured. The GMM is then used as the data fitness term for segmentation of bright field images using a graph cut formulation. Simultaneously, nuclear regions are detected with elliptic features (7), and the regions between them are tessellated to associate the amount of staining on a cell-by-cell basis. For endpoint (iii), signal detection is also quantified with the method outlined in (6), but the quantification of matrix is performed globally on an image-by-image basis. In these same tissue sections, morphometric features of fat cells are also quantified as follows. First, the nuclear regions are detected using a previously published method (8). Second, images are thresholded to reveal fat cell plasma membrane boundaries. These boundaries are then grouped and used as a reference to delineate fat cells through a cascade of morphometric filters (e.g., aspect ratio, convexity) and size (e.g., area, perimeter) for each connected component. In some instances, membrane boundaries are perceptual, and, as a result, adjacent fat cells are merged; however, such merged cells can be partitioned through convexity analysis (8).

In contrast to previous modes of imaging, samples in endpoint (iv) are imaged through fluorescence microscopy. In this case, the nuclear regions are segmented using the method described in (i) and (ii), the target fluorescence signal being segmented using a previously described method (9), in which regions between neighboring nuclei are tessellated, and the total fluorescent signal on a cell-by-cell basis is quantified.

Human paraffin tissue samples: Samples were identified through unlinked codes in accordance with federal HIPAA guidelines.

For the "Cancer-Free Cohort", histological sections were selected from a registry of 351 women who underwent core biopsies of the breast or lumpectomies at the University of California San Francisco or at California Pacific Medical Center (San Francisco, CA), between January 2, 2006 and April 30, 2009. All selected tissues were diagnosed as benign. Mammographic density was measured on digitized film-screens (acquired upon mammographic examination within 2 years of breast biopsy) by a trained radiologist using custom software as previously described (10). Women with 25-50% MD (n=13) and >70% MD (n=14) were selected for our low and high density cohorts, respectively **(Supplementary Table S5).** Patients with atypia upon pathology review were excluded from the study.

For the “Tumor Subtype Cohort”, histological sections were selected from a list of women who had undergone a lumpectomy or mastectomy surgery at The University of California San Francisco due to the presence of invasive ductal carcinoma (IDC). Eight women had estrogen receptor (ER) positive, progesterone receptor (PR) positive, HER2/neu receptor (HER2) negative tumors (ER+ CA); 6 women had HER2 positive, ER and PR positive or negative tumors (HER2+ CA); 6 women had ER, PR and HER2 (triple) negative tumors (TN CA). Tissue sections contained tumor tissue along with histologically normal adjacent tissue **(Supplementary Table S6).** Mammographic density was measured on digitized film-screens (acquired upon mammographic examination) by a trained radiologist using the Breast Imaging Reporting and Data System (BIRADS) classification: 1 = almost entirely fat, 2 = scattered fibroglandular densities, 3 = heterogeneously dense and 4 = extremely dense.

For the “Breast Map Cohort”, institutional review board approval was obtained to submit whole mastectomy specimens for research, after informed patient consent and procurement of any tissue needed for standard clinical pathological interpretation. Specimens were obtained from women who had undergone a therapeutic total mastectomy at Aurora Health Care (Milwaukee, WI) due to the presence of invasive breast cancer. Residual mastectomy tissues were “bread-loafed” into ~20 mm slices before being fixed in formalin. Tissue blocks were taken from every slice, their location within the mastectomy recorded, and embedded in paraffin using standard protocols. 5 μm sections were cut from

each of the paraffin-embedded tissue blocks, de-paraffinized, re-hydrated and stained with Hematoxylin and Eosin using standard protocols. Each tissue section was reviewed by a pathologist and the histological findings recorded onto a "breast map". Tissue blocks were chosen from three of these breast maps for analysis in this study **(Supplementary Table S7)**. For the tumor (CA) and normal adjacent (NA) samples, tissue blocks were chosen as follows: they contained both invasive cancer tissue and directly adjacent histologically normal tissue. For normal distal (ND) samples, tissue blocks were chosen as follows: they contained no cancer tissue (in situ or invasive) and were at least 2 slices away (>40 mm) from any invasive cancer tissue.

For the "Low and High MD Tumor Cohort", histological sections were selected from a list of women who had undergone a lumpectomy or mastectomy surgery at The University of California San Francisco due to the presence of ER+ IDC. 21 women exhibited low MD (BIRADS = 1) and 14 women exhibited high MD (BIRADS = 4) **(Supplementary Table S8).** Mammographic density was measured on digitized film-screens (acquired upon mammographic examination) by a trained radiologist using the Breast Imaging Reporting and Data System (BIRADS) classification: 1 = almost entirely fat, 2 = scattered fibroglandular densities, 3 = heterogeneously dense and 4 = extremely dense.

Linear regression analyses of published microarray data: The analysis included tumor samples, collected at time of surgery, from 398 invasive ductal

carcinoma (IDC) patients from datasets GSE6532 and GSE9195 (11, 12). Microarray data from each dataset were processed independently. The robust multi-array average (RMA) method was used to adjust for background, normalize signals, and estimate gene expression levels (3). The expression values for each of the five probe sets (206488_s_at, 209554_at, 209555_s_at, 228766_at and 242197_x_at) were centered by the median and then divided by the standard deviation (SD). Linear regression analyses were used to test the associations between the expression of each CD36 probe set and tumor grade, tumor size and age. Univariate models were used to test the association between CD36 as the dependent predictor and tumor grade, tumor size and age as the independent variables. A multivariate linear regression model, with CD36 as the dependent variable and tumor size and grade as the independent variables, was used to determine if the association between CD36 and grade was confounded by the association between CD36 and tumor size. Data from the probe set that was most strongly associated with tumor grade and tumor size (228766_at) are presented in **Supplementary Table S9**. p values were calculated using linear regression models via STATA version 9.

Supplementary References

1. Olumi AF, Grossfeld GD, Hayward SW, Carroll PR, Tlsty TD, Cunha GR. Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. Cancer Res 1999; 59:5002-5011.

2. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003; 19:185-193.
3. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003; 4:249-264.
4. Gregoire FM, Smas CM, Sul HS. Understanding adipocyte differentiation. Physiol Rev 1998; 78:783-809.
5. Parvin B, Yang Q, Fontenay G, Barcellos-Hoff MH. BioSig: an imaging bioinformatics system for phenotypic analysis. IEEE Trans Syst Man Cybern B Cybern 2003; 33:814-824.
6. Chang H, DeFilippis RA, Tlsty TD, Parvin B. Graphical methods for quantifying macromolecules through bright field imaging. Bioinformatics 2009; 25:1070-1075.
7. Yang Q, Parvin B. Harmonic cut and regularized centroid transform for localization of subcellular structures. IEEE Trans Biomed Eng 2003; 50:469-475.
8. Wen Q, Chang H, Parvin B. A Delaunay triangulation approach for segmenting clumps of nuclei. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2009; Boston, MA. 9-12.
9. Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Process 2001; 10:266-277.

10. Prevrhal S, Shepherd JA, Smith-Bindman R, Cummings SR, Kerlikowske K. Accuracy of mammographic breast density analysis: results of formal operator training. Cancer Epidemiol Biomarkers Prev 2002; 11:1389-1393.

11. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC Genomics 2008; 9:239.

12. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007; 25:1239-1246.