DETAILED METHODS

Selection of genes included in shRNA screen and clustering of their SAGE data

Genes were chosen based on their differential expression between the groups of samples (CD44+/CD24- and CD44-/CD24+ cells) shown in Figure 1A. Each gene had a SAGE (serial analysis of gene expression) tag that was ≥ 2-fold higher or lower in a pool of the SAGE libraries created from each CD44+/CD24- cells sample compared to a pool of the SAGE libraries created from each CD44-/CD24+ cells sample with a p-value ≤ 0.05 using the SAGE Digital Gene Expression Displayer of SAGE Genie (http://cgap.nci.nih.gov/SAGE) and also had lentiviral shRNAs available in the TRC library. SAGE Genie uses Bayesian analysis to compute the posterior probability that the underlying frequency of a tag in one collection of tissues exceeds the underlying frequency of the tag in a second collection of tissues. Hierarchical linkage clustering of normalized, log-transformed, median-centered SAGE data for the genes in the screen was carried out using Cluster (1), and results were visualized with MapleTree (developed by L. Simirenko).

Clustering of cell lines microarray data

Before clustering, the previously published microarray data referred to in the legend for Figure 1B were log-transformed and median-centered by cell line. Next, Affymetrix records with expression values >(mean + standard deviation) or <(mean - standard deviation) in < 10% of cell lines were removed, and the remaining data were median-centered by gene. Then, hierarchical linkage clustering was carried out using Cluster using the microarray data for the genes in the shRNA screen. Finally, results were visualized using MapleTree.

Flow cytometry of cell lines

Cells were cultured at 37°C with 5% CO_2 in the media used for the shRNA screen. Harvested cells were double-stained with PE-CD44 (BD Biosciences catalog number 555479) and Alexa 647-CD24 (BioLegend catalog number 311110) antibodies and analyzed by flow cytometry to detect levels of extracellular CD44 and CD24.

qMSP of cell lines and qRT-PCR of Hs 578T to validate *IL6* shRNAs

qMSP (quantitative methylation-specific PCR) and qRT-PCR were performed as previously described (2). The sequences of qRT-PCR primers used are listed in Supplemental Table 10. In brief, the level of methylation or expression of each gene in each sample was assessed in triplicate along with the level of methylation or expression of *ACTB* or *RPL39*, respectively, to which each was compared to determine the qMSP or qRT-PCR reading for each gene of interest. For validation of *IL6* shRNAs, Hs 578T cells were infected one day after plating, maintained in puromycin medium beginning the next day, and harvested three days after infection. The sequences of the *IL6* shRNAs used are listed in Supplemental Table 10.

shRNA screen implementation

In each phase of the screen, each cell line of interest was screened using several steps as follows. First, cells were plated in 384-well plates in growth medium and then incubated for one hour at room temperature. The next day, they were infected with each shRNA of interest or control shRNAs targeting GFP, lacZ, luciferase, RFP, or no genes in quadruplicate (one shRNA per well) in 8 µg/mL polybrene (Sigma catalog number H9268). One or two days later, media replacement was carried out, adding puromycin (to select for infected cells) to two wells with each shRNA. Finally, six days post-infection, cell viability was assessed using CellTiter-Glo (Promega). Specific growth, infection, and selection conditions were used for each cell line as shown in Supplemental Table 9. Conditions were chosen to minimize lentivirus toxicity and maximize infection efficiency. Cell lines were cultured at 37°C with 5% CO_2. When spinning was

used, cells were spun for 30 minutes at 2,250 rpm directly after virus addition. When fresh medium was used, the medium on the cells was removed three hours after virus addition and replaced with new medium. The puromycin addition time refers to how long after virus addition puromycin selection was begun. CellTiter-Glo was used at 10 µL per well for BT-549, Hs 578T, MCF7, and T-47D and 15 µL per well for all other cell lines. Robots used included the MicroFill for cell plating; the Janus for media changes for BT-549, Hs 578T, MCF7, and T-47D; the EP3 for infecting BT-549, Hs 578T, MCF7, and T-47D; and the Tecan Aquarius for media changes for all other cell lines and for infecting all other cell lines.

shRNA screen hits definition

In phase 1, we called shRNAs "hits" based on the four robust z-scores representing their effects on cell viability in either BT-549 and Hs 578T (with and without puromycin each) or in MCF7 and T-47D (with and without puromycin each). shRNAs were designated as hits if they had four robust z-scores below the 95th percentile, targeted genes with two shRNAs with four robust z-scores below the 90th percentile, or had two robust z-scores below the 85th percentile and targeted genes with a ≥ 85-fold expression difference between CD44+/CD24- and CD44-/CD24+ cells. Only shRNAs with infection efficiencies > 0.25 and robust z-score standard deviations below the 99th percentile for both basal-like or luminal cell lines with and without puromycin were considered in this analysis. In phase 2, we considered shRNAs to have "scored" in a cell line if they decreased viability by ≥ 23% (the 80th percentile) compared to control (after averaging effects with and without puromycin). Based on this, we identified 17 shRNAs targeting 15 genes that scored in ≥ 3 basal-like and ≤ 1 luminal lines as "basal-like-specific hits," 13 shRNAs targeting 13 genes that scored in ≥ 5 basal-like and ≤ 4 luminal lines as "basal-like hits," one shRNA that scored in six luminal and four basal-like lines as a "luminal hit," and nine shRNAs targeting nine genes that scored in ≥ 5 basal-like and luminal lines (or four of each) as "non-selective hits." Only shRNAs with infection efficiencies > 0.25 for all cell lines were considered in this analysis.

Immunoblots

Cells for most immunoblots were plated in 6-cm dishes the day before harvesting untreated cells or the day before beginning drug and control treatments (which were carried out as indicated in the figure legends), always using 1,000,000 cells per dish in 5 mL growth medium per dish. The only exceptions were the cells for the blots in Figure 4 and Supplemental Figure 5, which were frozen at -80°C before being used for immunoblotting. Cells were cultured at 37°C with 5% CO_2 in the media used for the shRNA screen. Cells infected with shRNAs had been stably maintained in puromycin medium. The sequences of the *IL6* and *STAT3* shRNAs used are listed in Supplemental Table 10. Cells were harvested in PBS by scraping and then spun at 3,000 rpm at 4°C for five minutes. Next, cells were lysed in buffer containing 50 mM Tris, pH 8, 250 mM NaCl, and 0.5% NP40, supplemented with sodium vanadate and protease inhibitors, for 15 minutes on ice. After lysis, samples were centrifuged at 14,000 rpm for ten minutes, and protein concentration was quantified using Bradford reagent (Bio-Rad Laboratories). Then, 2x sample buffer (containing 0.125 M Tris, pH 6.8, 4% SDS, and 20% glycerol) with 4% beta-mercaptoethanol was added to the lysates, and the mixtures were boiled for five minutes. Next, 30 mg of each was transferred onto and electrophoresed through an 8% polyacrylamide gel. After that, the protein was relocated to nitrocellulose and visualized by Ponceau. Blots were then incubated in 5% milk in TBST (10 mM Tris, pH 8, 150 mM NaCl, and 0.05% Tween) for one hour and then TBST containing antibodies diluted 1:10,000 overnight at 4°C. This was followed by washing with TBST three times, incubation with the appropriate horseradish peroxidase-labeled secondary antibody at 1:10,000 for two hours at room temperature, three more washes with TBST, and developing using Western Blot Chemiluminescence Reagent Plus

(PerkinElmer). For reprobing, blots were washed for one hour in TBST and stripped and reprobed in stripping buffer (0.2% SDS, 64 mM Tris, pH 6.8, and 0.35% beta-mercaptoethanol) at 65°C for 30 minutes. Antibodies used were pStat3 (Cell Signaling catalog number 9131), Stat3 (Santa Cruz catalog number sc-482), pStat1 (Zymed catalog number 33-3400), pStat5 (Cell Signaling catalog number 9359), tubulin (Sigma catalog number T-5168), and pserStat3 (3). Immunoblots were quantitated using ImageJ software.

Cell culture inhibitor experiments

Inhibitors used included JAK Inhibitor I (Calbiochem catalog number 420099), tranylcypromine (Sigma catalog number P8511), 3-(3-pyridinyl)-1-(4-pyridinyl)-2-propen-1-one (generously provided by J. Chesney, University of Louisville), N-(2-Hydroxy-4-nitrophenyl)-N′-(2-bromophenyl)urea (Calbiochem catalog number 559405), 4-methylumbelliferone (Sigma catalog number M1381), 1,4-dimethoxy-2-methylnaphthalene (generously provided by D. Criddle, University of Liverpool), and BSK805 (Novartis). Each inhibitor was dissolved in DMSO. For GI50 calculations, effects of each inhibitor on cell viability compared to DMSO control were assessed in triplicate for each cell line at six to seven concentrations for each inhibitor. Concentrations tested included the following: for JAK Inhibitor I, 0.5, 1, 2, 4, 8, and 16 μM; for tranylcypromine, 125, 250, 500, 1,000, 2,000, and 4,000 μM; for 3-(3-pyridinyl)-1-(4-pyridinyl)-2-propen-1-one, 0.625, 1.25, 2.5, 5, 10, and 20 μM; for N-(2-Hydroxy-4-nitrophenyl)-N′-(2-bromophenyl)urea, 0.75, 1.5, 3, 6, 12, and 24 μM; for 4-methylumbelliferone, 62.5, 125, 250, 500, 1,000, and 2,000 μM; for 1,4-dimethoxy-2-methylnaphthalene, 3.125, 6.25, 12.5, 25, 50, and 100 μM; and for BSK805, 0.00064, 0.0032, 0.016, 0.08, 0.4, 2, and 10 μM. Cells were plated at 500 cells per well in 96-well plates and treated the next day with drug or control. Cells were cultured at 37°C with 5% CO_2 in the media used for the shRNA screen, and cell viability was measured using CellTiter-Glo three days after treatments. The same inhibitor concentrations used for the immunoblots in Figure 6A were used for the luciferase assay and the SAGE-Seq library results in Figure 6C-D and F.

siRNAs experiment to validate screen results

Cells were plated at 5,000 cells per well in 96-well plates and cultured at 37°C with 5% CO_2 in the media used for the shRNA screen. The next day, cells were transfected in triplicate with siGENOME SMARTpools for the genes of interest, “Non-Targeting siRNA #2,” or no siRNAs at 0.1 μM using DharmaFECT 1 (Dharmacon). The sequences of the siRNAs in the SMARTpools are listed in Supplemental Table 10. Cell viability was measured using CellTiter-Glo three days after transfections, with the effects of each siRNAs treatment on each cell line compared to the effects of no siRNAs.

IL6 ELISA of cell lines (including Hs 578T to validate *IL6* shRNAs)

Cells were plated at 15,000 cells per well in 96-well plates in triplicate. Twenty-four hours later, supernatant samples of each well were obtained and used for IL6 ELISA (Thermo Scientific catalog number EH2IL65) following the manufacturer’s instructions. Cells were cultured at 37°C with 5% CO_2 in the media used for the shRNA screen. Cells infected with shRNAs had been stably maintained in puromycin medium. When supernatant samples were obtained, cell viability was also measured using CellTiter-Glo.

Recombinant IL6 rescue experiment

Hs 578T cells were plated at 500 cells per well in 96-well plates. Every 24 hours for three days beginning 24 hours after plating, cells were treated with either rhIL6 (eBioscience catalog number 14-8069-62, 2.5 ng/mL in PBS with 10% FBS) or PBS with 10% FBS alone in triplicate. Cells were cultured at 37°C with 5% CO_2 and puromycin in the medium used in the shRNA

screen. Twenty-four hours after the last treatment, cell viability was measured using CellTiter-Glo.

Caspase activation assay and cell cycle analysis
Hs 578T cells were treated with DMSO control or inhibitors (at the same concentrations as used in the experiments in Figure 6A, C-D, F), transfected with siRNAs or no siRNAs (at 0.1 μM with DharmaFECT 1), or infected with lentiviruses one day after plating. Cells were cultured at 37°C with 5% CO_2 in the medium used for the shRNA screen. Infected cells were maintained in puromycin medium beginning the day after infection. Cells were assayed for caspase activation and analyzed for cell cycle three days after treatment or transfection and six days post-infection. The Apo-ONE Homogeneous Caspase-3/7 Assay (Promega) was carried out in triplicate alongside triplicate measurement of cell viability using CellTiter-Glo. Cell cycle analysis was performed essentially as described previously (4).

Processing of breast cancer samples prior to xenograft studies
Tumors were minced with razor blades and digested with stirring for 3-4 hours at 37°C in DMEM/F12 with 2 mg/mL BSA, 2 mg/mL collagenase type IV, and 2 mg/mL hyaluronidase. After digestion, cells were filtered through 500-micron mesh, washed in DMEM/F12 with 5% FBS, and frozen in DMEM/F12 with 5% FBS and 10% DMSO. Pleural effusions were spun down to collect cells. Cells were washed in DMEM/F12 with 5% FBS and frozen in DMEM/F12 with 5% FBS and 10% DMSO. Cultures of tumor-associated fibroblasts were obtained by plating chopped pieces of tumors on collagen-coated plates in Medium 171 with HuMEC supplement, without BPE, and with 5% FBS. Within two weeks, cultures of robustly growing fibroblasts were established. For mouse injections, pleural effusions were thawed, allowed to recover for 15 minutes at 37°C in Medium 171 with HuMEC supplement, without BPE, and with 5% FBS, and spun down before resuspending for injection. Primary tumor cells were processed analogously, but they were mixed with cultured tumor-associated fibroblasts where indicated.

Signature network
Links between basal-like-specific hits were illustrated using information from primary literature and the Metacore software platform (5).

Stat3 transcriptional activity assay
Hs 578T or MCF7 cells were plated at 5,000 cells per well in 96-well plates and treated the next day with inhibitors or DMSO control in triplicate. Cells were cultured at 37°C with 5% CO_2 in the medium used for the shRNA screen. Cells were transfected the day after treatment with 0.25 μg m67-luc reporter construct (generously provided by J. Bromberg, Memorial Sloan-Kettering Cancer Center) and 0.025 μg pRL-TK (Promega) per well using Lipofectamine 2000 (Invitrogen). Transfection mixes were removed after five hours, upon which treatment with inhibitors or DMSO control was resumed and luciferase activity was measured the next day using the Dual-Luciferase Reporter System (Promega catalog number E1980).

Xenograft studies
For xenograft assays, 2 million SUM159PT, 2 million MDA-MB-231, 3 million MDA-MB-468, or 5 million Hs 578T cells that were suspended in 50% Matrigel (BD Biosciences catalog number 354234) in DMEM/F12, or 2-3x10^5 dissociated, uncultured IDC31 (mixed with an equal amount of cultured tumor-associated fibroblasts), 1x10^5 IDC31-X (previous mouse xenograft of IDC31, mixed with 3x10^5 cultured tumor-associated fibroblasts), 2-3x10^5 PE18, or 2-3x10^5 IDC50-X (previous mouse xenograft of IDC50) cells suspended in 50% Medium 171 and 50% Matrigel (or 3% FBS and 4 mg/mL collagen in Medium 171 gel for IDC50-X), were transplanted into two

contralateral inguinal fat pads following institutionally approved animal surgery procedures. Cells infected with shRNAs had been stably maintained in puromycin medium. The sequences of the *STAT3* shRNAs used are listed in Supplemental Table 10. At time-points and for the durations indicated in the figure legends, animals were subjected to daily gavage with NVP-BSK805 at 2 mg/mouse, freshly dissolved in pH 3, 50 mM citrate buffer or citrate buffer only (control group). Following the period of administration of drug or vehicle, the animals were killed, and tumors were recovered, weighed, and subjected to histological examination. Hematoxylin and eosin (H&E) staining patterns in xenograft tumors derived from SUM159PT, Hs 578T, IDC31-X, and IDC50-X cells were observed with a Nikon Eclipse E800 microscope. Cellularity was quantitated based on H&E staining using ImageJ software. Mice injected with SUM159PT cells containing shSTAT3 #1, shSTAT3 #2, or shGFP control were killed when they had large tumors (> 10 mm in diameter) and clear signs of tumor-related morbidity (loss of grooming, hunched appearance).

Preparation of cells for SAGE-Seq libraries

Hs 578T or MCF7 cells were plated at 3,000,000 cells per P150 and treated the next day with inhibitors or DMSO control or plated at 2,800,000 cells per P150 and transfected the next day with siRNAs (at 0.1 μM with DharmaFECT 1). Cells were cultured at 37°C with 5% CO_2 in the medium used for the shRNA screen, with the transfection mixes removed from transfected cells the day after transfection. Cells were harvested two days post-treatment/transfection, and 1,000,000 cells of each sample were used to prepare SAGE-Seq libraries according to a detailed protocol (http://research4.dfci.harvard.edu/polyaklab/protocols_linkpage.php).

Definition of Stat3 signature genes and clustering of their SAGE-Seq data

SAGE-Seq (serial analysis of gene expression combined with the Illumina/Solexa 1G platform) was performed, and tags were mapped and normalized, essentially as described previously (6). Then, for each tag, fold-changes and Z-test scores for each drug-treated sample compared to matching DMSO control-treated cells, and for each *STAT3* siRNAs-treated sample compared to matching non-targeting siRNAs-treated cells, were calculated. Tag counts used were the sums of all sense SAGE-Seq tags uniquely mapped to the same gene. Fold-changes were calculated as (tag count in listed library)/(tag count in control library) when (tag count in listed library) > (tag count in control library) and (-tag count in control library)/(tag count in listed library) when (tag count in listed library) < (tag count in control library), using a normalized tag count representing a raw tag count of one in that library when a tag count was zero. Z-test scores were calculated as [(tag count in listed library)/(total tags in listed library) - (tag count in control library)/(total tags in control library)]/square root of [(p)(1 - p)/(total tags in listed library) + (p)(1 - p)/(total tags in control library)], where p = [(tag count in listed library) + (tag count in control library)]/[(total tags in listed library) + (total tags in control library)]. Genes were considered to be significantly regulated by *STAT3* siRNAs if they had tags with |fold-change| > 2 and |Z-test score| > 2.33 for *STAT3* versus non-targeting siRNAs-treated cells. For each Hs 578T and MCF7 cells, genes among these that were similarly affected (same direction fold-change) by at least four inhibitors but not by the NQO1 inhibitor in the same cell line were called the Stat3 signature for those cells. The SAGE-Seq data for each set of genes were ordered by hierarchical linkage clustering using Cluster and visualized using MapleTree.

Association of Stat3 signatures with clinical outcome and gene ontology analysis

The microarray data from the two breast cancer data sets (7-9) referred to in Supplemental Figure 5C were processed as described previously (10). In each data set, tumors were considered to have the Hs 578T or MCF7 Stat3 signature if they had average expression values for all genes in the signature downregulated by *STAT3* siRNAs above the 60th percentile and average expression values for all genes in the signature upregulated by *STAT3* siRNAs below

the 40th percentile. Percentiles were calculated separately for groups of only ER+ or only ER- tumors. In each data set, tumors were considered to express the set of genes significantly regulated by *STAT3* siRNAs in Hs 578T cells if they had average expression values for all genes downregulated by *STAT3* siRNAs above the 60th percentile and average expression values for all genes upregulated by *STAT3* siRNAs below the 40th percentile. The microarray data for the genes in the Stat3 signatures were formatted using Cluster and visualized using MapleTree.Gene ontology biological process categories represented by the genes in the Stat3 signatures were determined using the DAVID Functional Annotation Tool (11) with the highest classification stringency and *Homo sapiens* background employed.

Multiple immunofluorescence and immunohistochemistry of tissue microarrays and FFPE xenograft tumors

The triple immunofluorescence for CD44 (Neomarkers), CD24 (Neomarkers), and pStat3 (Cell Signaling) in primary tumors was performed in tissue microarrays (12) or in whole sections of FFPE xenograft tumors. The tissues were deparaffinized in xylene and hydrated in a series of ethanol. After heat-induced antigen retrieval in citrate buffer (pH 6), the samples were blocked with goat serum and sequentially stained with the different primary and secondary antibodies. The sequential staining was optimized to avoid cross-reaction between antibodies and was performed as follows: polyclonal antibody anti-pStat3 (1:25 dilution) and monoclonal (IgG2a) antibody anti-CD44 (1:25 dilution) for one hour at 37°C; goat anti-rabbit Alexa 488-conjugated (Invitrogen, 1:100 dilution) for 30 minutes at room temperature; rabbit anti-mouse IgG2a biotin-conjugated (Invitrogen, 1:100 dilution) for 30 minutes at room temperature; monoclonal (IgM) antibody anti-CD24 (1:25 dilution) for one hour at room temperature; goat anti-mouse IgM Alexa 647-conjugated (Invitrogen, 1:100 dilution) and streptavidin Pacific Blue-conjugated for 30 minutes at room temperature. The samples were washed twice with PBS-Tween 0.05% between each pair of incubations and protected for long-term storage with ProLong Gold mounting medium (Invitrogen). Before image analysis, the samples were stored at -20°C for at least 48 hours. Immunofluorescence images were acquired with a Nikon Ti microscope attached to a Yokogawa spinning-disk confocal unit, 60x plan apo objective, and OrcaER camera controlled by Andor iQ software. Different images from multiple areas of each tumor were acquired, and the expression of pStat3 was scored in 100 cells of each type (CD44+/CD24-, CD44+/CD24+, CD44-/CD24+, and CD44-/CD24-) in the primary tumors or 500 cells (4-6 fields) in the xenograft tumors. The percentage of each cell type within a tumor was scored by counting 200 cells. Immunohistochemistry (for pStat3 only) was carried out using standard procedures of immunohistochemistry and methyl green as a counterstain. Two images from each tumor were acquired, and the expression of pStat3 was scored in the cells in those two fields.

Statistical analyses

All t-tests used in these studies were two-tailed, two-sample equal-variance Student's t-tests. Score tests used here were generalized linear model score tests adjusted for multiple comparisons. T-test, log-rank test, and chi-squared p-values < 0.05 were considered statistically significant. Score test p-values < 0.001 were considered statistically significant.

LEGENDS FOR EXCEL FILES

Supplemental Table 1. Tag counts in CD44+/CD24- and CD44-/CD24+ cell SAGE libraries for genes in shRNA screen. Each SAGE (serial analysis of gene expression) tag shown was ≥ 2-fold higher or lower in a pool of the SAGE libraries created from all samples of CD44+/CD24- cells shown compared to a pool of the SAGE libraries created from all samples of CD44-/CD24+ cells shown with a p-value ≤ 0.05 using the SAGE Digital Gene Expression Displayer of SAGE Genie (http://cgap.nci.nih.gov/SAGE) and corresponds to a gene with lentiviral shRNAs available in the TRC library. The sequence of each tag, its tag counts per 200,000 tags in SAGE libraries created from the indicated samples, and the ID and symbol of the corresponding gene are shown. PE, pleural effusion; ASC, ascites; IDC, invasive ductal carcinoma.

Supplemental Table 3. Viability of breast cell lines when infected with each shRNA in phase 1 of screen. Each shRNA listed had a robust z-score with standard deviation below the 99th percentile for both BT-549 and Hs 578T (for both puro+ and puro- for each) or for both MCF7 and T-47D (for both puro+ and puro- for each) among shRNAs with infection efficiencies > 0.25 in each cell line. The name of each shRNA, the ID and symbol of the targeted gene, and the puro+ and puro- robust z-scores for each cell line, calculated as (average viability - plate median average viability)/(plate median absolute deviation of average viability), are shown. A hyphen indicates that the listed shRNA was not tested in that cell line in phase 1 of the screen.

Supplemental Table 4. Viability of breast cell lines when infected with each shRNA in phase 2 of screen. Each shRNA listed was a hit in phase 1 of the screen and had an infection efficiency > 0.25 in all cell lines in phase 2 of the screen. The name of each shRNA, the ID and symbol of the targeted gene, and the average of the puro+ and puro- percent-of-control values for each cell line, calculated as (100 x average viability)/(median average viability of plate controls), are shown.

Supplemental Table 5. Information about genes targeted by hits identified in phase 2 of shRNA screen. All genes targeted by shRNAs that were identified as hits in phase 2 of the shRNA screen are listed. The type of hit of the corresponding shRNAs, the ID and symbol of each gene, a description of each gene and its map location, and the GO terms assigned to each gene in component, function, and process categories are shown. A hyphen indicates that the corresponding information was not available. GO, gene ontology.

Supplemental Table 6. Fold-changes in SAGE-Seq library tag counts for Stat3 signature genes. All genes in the Stat3 signatures are listed. The ID and symbol of each gene and the fold-changes in its corresponding tag counts in SAGE-Seq (serial analysis of gene expression combined with the Illumina/Solexa 1G platform) libraries created from Hs 578T or MCF7 cells treated with *STAT3* siRNAs (or non-targeting siRNAs as control) or each indicated inhibitor (or no drug as control) are shown. Tag counts used were the sums of all sense SAGE-Seq tags uniquely mapped to the same gene. Each gene listed had |fold-change| > 2 and |Z-test score| > 2.33 for tag counts in the Hs 578T or MCF7 *STAT3* siRNAs SAGE-Seq library and was similarly affected (same direction fold-change) by at least four inhibitors but not by the NQO1 inhibitor in the same cell line. Fold-changes were calculated as (tag count in listed library)/(tag count in control library) when (tag count in listed library) > (tag count in control library) and (-tag count in control library)/(tag count in listed library) when (tag count in listed library) < (tag count in control library), using a normalized tag count representing a raw tag count of one in that library when a tag count was zero. Z-test scores were calculated as [(tag count in listed library)/(total tags in listed library) - (tag count in control library)/(total tags in control library)]/square root of [(p)(1 - p)/(total tags in listed library) + (p)(1 - p)/(total tags in control library)], where p = [(tag count in

listed library) + (tag count in control library)]/[(total tags in listed library) + (total tags in control library)].

Supplemental Table 7. Information about Stat3 signature genes. All genes in the Stat3 signatures are listed. The effect of *STAT3* siRNAs in Hs 578T or MCF7 cells on the expression of each gene ("Down" = decreased, "Up" = increased), the ID and symbol of each gene, a description of each gene and its map location, and the GO terms assigned to each gene in component, function, and process categories are shown. A hyphen indicates that the corresponding information was not available. GO, gene ontology.

Supplemental Table 2. Characteristics of breast cell lines in shRNA screen. The name, subtype (determined based on previously published gene expression data), tissue of origin, and status of ER, PR, HER2, *TP53*, *PIK3CA*, *PTEN*, *BRCA1*, and *SMAD4* are shown for each of the 14 cell lines included in the shRNA screen. A hyphen indicates that the corresponding information has not been determined. ER, estrogen receptor; PR, progesterone receptor; pos, positive; neg, negative; mut, mutant; wt, wild-type; HD, homozygous deletion.

Cell Line	Subtype	Tissue of Origin	ER	PR	HER2	*TP53*	*PIK3CA*	*PTEN*	*BRCA1*	*SMAD4*
BT-474	luminal	primary tumor	pos	pos	pos	mut	wt	wt/pos	wt	-
BT-549	basal-like	primary tumor	neg	neg	neg	mut	wt	822delG (L295X)/null	wt	-
HCC1937	basal-like	primary tumor	neg	neg	neg	mut	wt	HD/null	5382insC (fs>1829X)	-
Hs 578T	basal-like	primary tumor	neg	neg	neg	mut	wt	wt/pos	wt	-
MCF 10A	basal-like	normal breast	neg	neg	neg	neg	-	-	-	-
MCF-12A	basal-like	normal breast	neg	neg	neg	neg	-	-	-	-
MCF7	luminal	pleural effusion	pos	pos	neg	wt	mut	wt/pos	wt	-
MDA-MB-231	basal-like	pleural effusion	neg	neg	neg	mut	wt	wt/pos	wt	-
MDA-MB-453	luminal	pleural effusion	neg	neg	pos	mut	mut	919G>A (E307K)/pos	wt	-
MDA-MB-468	basal-like	pleural effusion	neg	neg	neg	mut	mut	IVS4+1G>T (A71fsX5)/null	wt	HD
SK-BR-3	luminal	pleural effusion	neg	neg	pos	mut	wt	wt/pos	wt	-
SUM159PT	basal-like	primary tumor	neg	neg	neg	mut	mut	wt/pos	wt	-
T-47D	luminal	pleural effusion	pos	pos	neg	mut	mut	wt/pos	wt	-
ZR-75-1	luminal	ascites	pos	pos	pos	wt	wt	323T>G (L108R)	wt	-

Supplemental Table 8. Correlation of Hs 578T Stat3 signature with tumor ER status. The numbers of ER- and ER+ tumors without (-) and with (+) the Hs 578T Stat3 signature for the two cohorts of patients (7-9) analyzed are listed. The percentages of tumors without and with the signature that are ER+ are also indicated, along with chi-squared p-values for the association between the signature and tumor ER status. ER, estrogen receptor.

Cohort	Stat3 Signature	ER-	ER+	% ER+	p-value
286 patients	+	12	34	73.9	0.889
	-	65	175	72.9	
141 patients	+	8	10	55.6	0.060
	-	29	94	76.4	

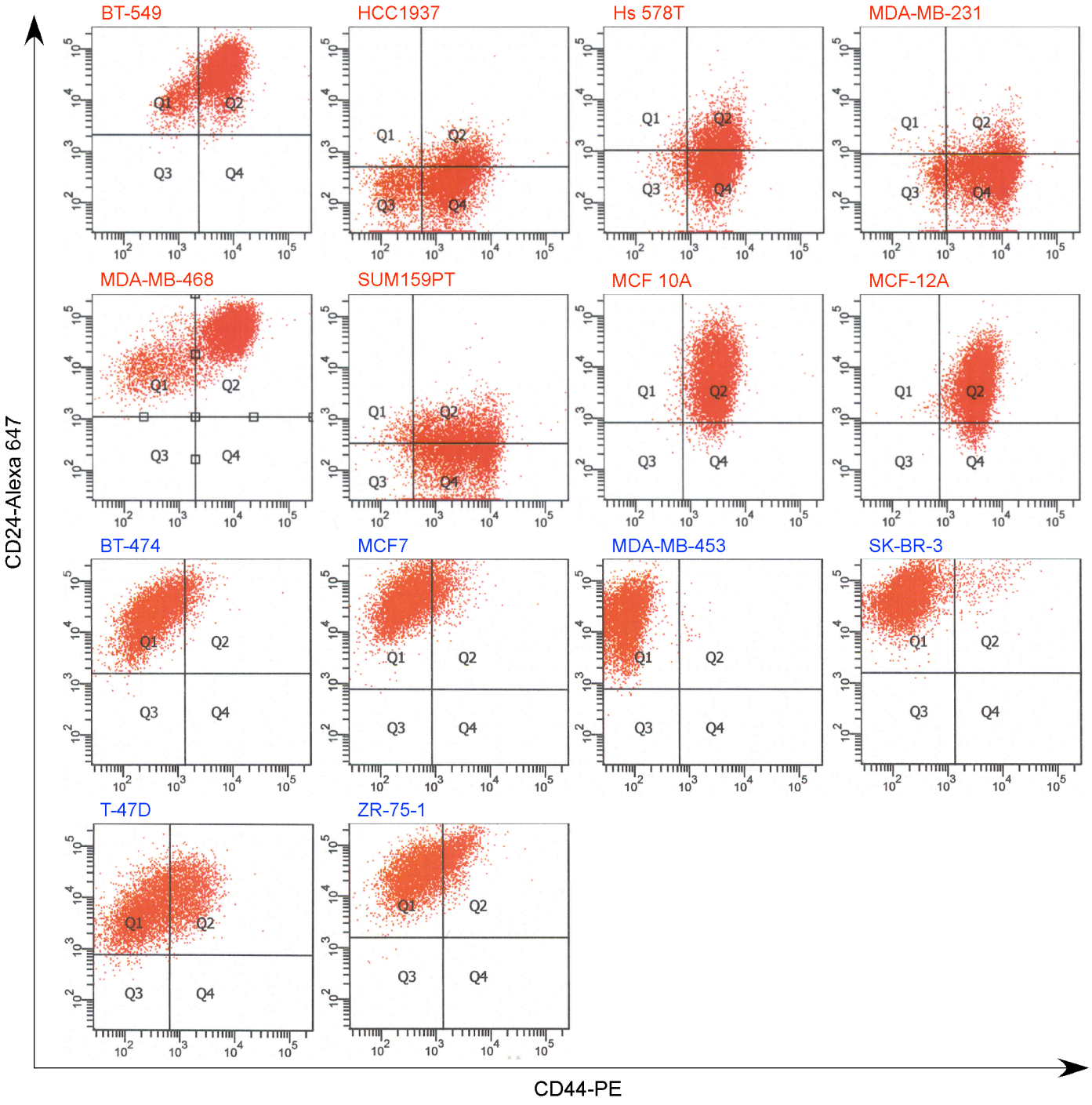
Supplemental Table 9. Growth, infection, and selection conditions used in shRNA screen for each cell line. The name of each breast cell line included in the shRNA screen and the infection and selection conditions used for each are shown. These conditions were chosen in order to maximize infection efficiency and minimize lentivirus toxicity. When spinning was used, cells were spun for 30 minutes at 2,250 rpm directly after virus addition. When fresh medium was used, the medium on the cells was removed three hours after virus addition and replaced with new medium. The puromycin addition time refers to how long after virus addition puromycin selection was begun.

Cell Line	Medium	Cells/Well	Virus/Well (µL)	Spinning	Fresh Medium	Puromycin Addition Time (hours)	Puromycin Concentration (µg/mL)
BT-474	RPMI + 10% FBS + 10 µg/mL insulin	2500	4.0	yes	no	24	2
BT-549		500	2.5	yes	no	24	2
HCC1937		1000	1.5	no	no	24	2
Hs 578T		500	1.9	yes	no	24	2
MCF7		500	4.0	yes	no	24	2
T-47D		500	3.0	yes	no	24	2
ZR-75-1		2000	0.5	yes	no	24	1
MDA-MB-231	McCoy's + 10% FBS	1000	0.5	no	no	24	2
MDA-MB-453		750	2.0	no	yes	48	2
MDA-MB-468		1500	2.0	yes	no	24	2
SK-BR-3		2000	2.0	no	no	24	2
SUM159PT	50% DMEM/F12 + 10% FBS, 50% Medium 171 + HuMEC supp.	1500	1.0	no	no	24	2
MCF 10A	DMEM/F12 + 5% horse serum + 10 µg/mL insulin + 20 ng/mL EGF + 100 ng/mL cholera toxin + 500 ng/mL hydrocortisone	500	0.5	no	yes	48	2
MCF-12A		500	0.5	no	yes	48	2

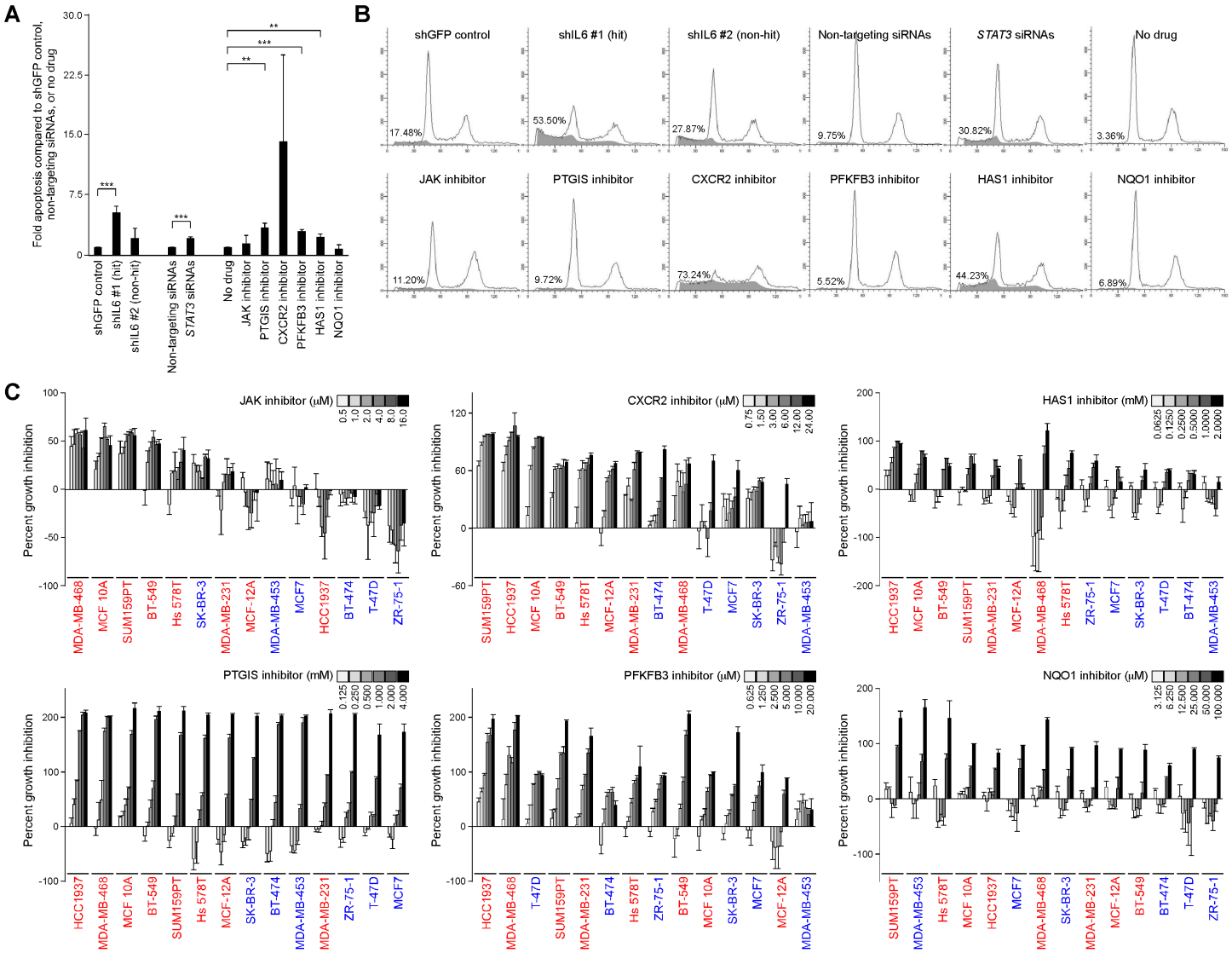
Supplemental Table 10. Sequences of primers, siRNAs, and key shRNAs used in this study. The gene symbol and sequences of each oligo or set of oligos are listed along with the direction (primers), catalog number (siRNAs), or hairpin name (shRNAs) of each.

Oligo Type	Oligo Information	Gene Symbol	Sequences
primer	forward	*IL6*	ACTCACCTCTTCAGAACGAATTG
primer	reverse	*IL6*	AGCCATCTTTGGAAGGTTCAG
siRNAs	M-004691-02-0005	*PTGIS*	GGAAACGGCUGAAGAAUUA; GCGGAGAGCUCGAGAGUAU; GGGCAGGUAUGUCACCGUU; GAAUUCAACCUGCGACGUG
siRNAs	M-007878-03-0005	*CXCL3*	AUCCAAAGUGUGAAUGUAA; UCCAAAGUGUGAAUGUAAG; GAAAAGAUACUGAACAAGG; UGUCCAUUCUGCAGCGUUU
siRNAs	M-007993-02-0005	*IL6*	UAACAAGAGUAACAUGUGU; UGAGGGCUCUUCGGCAAAU; GAACAGAUUUGAGAGUAGU; GAAUCUAGAUGCAAUAACC
siRNAs	M-006763-01-0005	*PFKFB3*	GGACCUAACCCGCUCAUGA; GAGGAUCAGUUGCUAUGAA; AAAGCUACCUGGCGAAAGA; ACAAGUACUAUUACCGCUA
siRNAs	M-009664-00-0005	*HAS1*	GCUACUGGGUGGCCAUGUU; GAUACUGGGUAGCCUUCAA; CAGACACGCUGGUCCAAGU; UGGGUUAUGCUACCAAGUA
siRNAs	M-003146-02-0005	*JAK2*	GAGCAAAGAUCCAAGACUA; GCCAGAAACUUGAAACUUA; GAUCCUGGCAUUAGUAUUA; ACAGAAUGCUGGAACAAUA
siRNAs	M-003544-02-0005	*STAT3*	GGAGAAGCAUCGUGAGUGA; CCACUUUGGUGUUUCAUAA; UCAGGUUGCUGGUCAAAUU; CGUUAUAUAGGAACCGUAA
siRNAs	M-007994-00-0005	*IL6RA*	GAACUCAUCUUUCUACAGA; GAAGUUCUCCUGCCAGUUA; CAACAUGGAUGGUCAAGGA; GGAGACAGCUCUUUCUACA
siRNAs	M-005647-00-0005	*CXCR2*	CAGCAGGCCUUCCUUUGUU; CAUUAUCUAUGCCCUGGUA; GAGGACAUGGGCAACAAUA; GGUGAAGAUCUUAGUAAUU
shRNA	shIL6 #1 (hit)	*IL6*	CCGGCTGGATTCAATGAGGAGACTTCTCGAGAAGTCTCCTCATTGAATCCAGTTTTTG
shRNA	shIL6 #2 (non-hit)	*IL6*	CCGGCATCTCATTCTGCGCAGCTTTCTCGAGAAAGCTGCGCAGAATGAGATGTTTTTG
shRNA	shSTAT3 #1	*STAT3*	CCGGGCACAATCTACGAAGAATCAACTCGAGTTGATTCTTCGTAGATTGTGCTTTTT
shRNA	shSTAT3 #2	*STAT3*	CCGGGCAAAGAATCACATGCCACTTCTCGAGAAGTGGCATGTGATTCTTTGCTTTTT

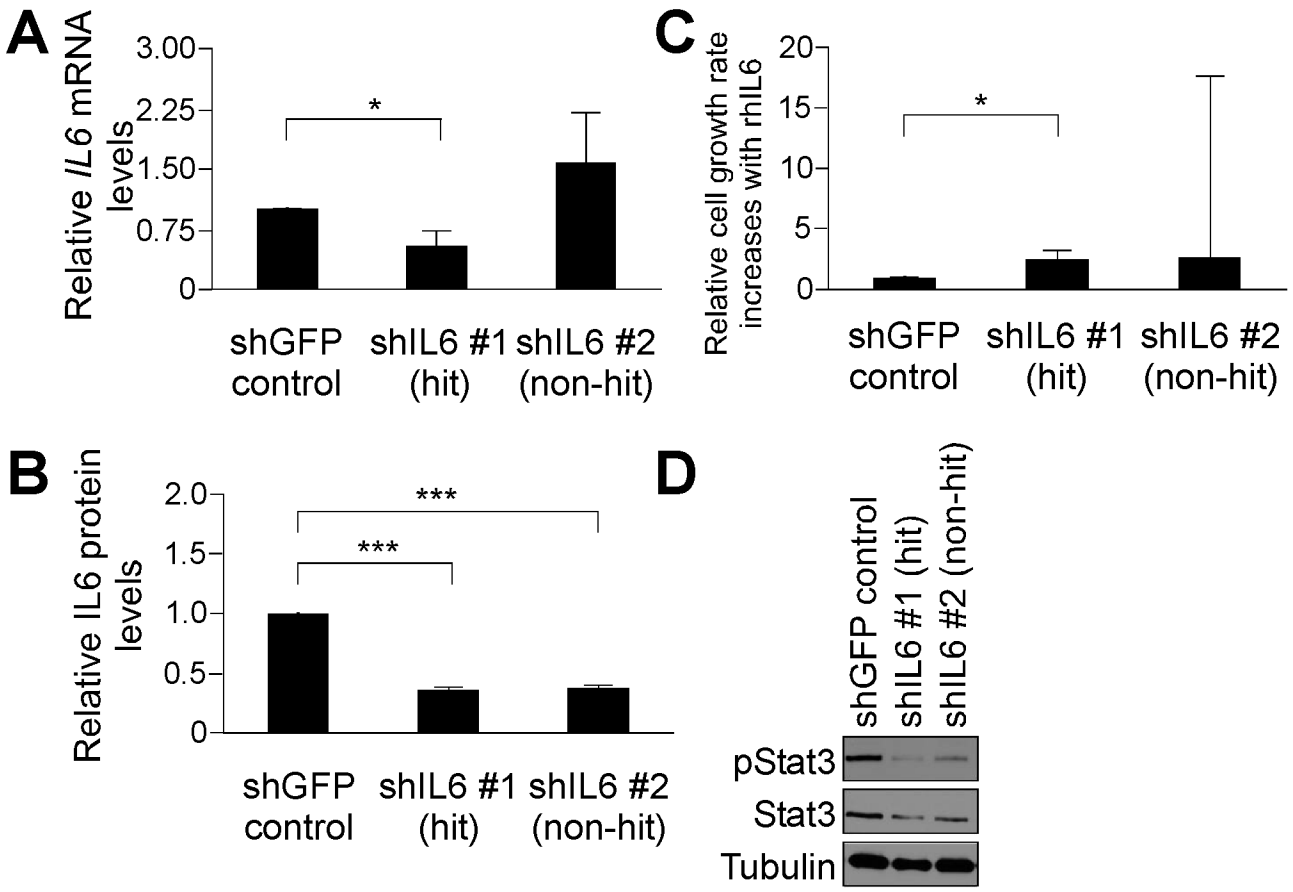
Supplemental Figure 1. Flow cytometry analysis of CD44 and CD24 levels for breast cell lines. Results of double-staining cells with CD44 and CD24 antibodies are shown. When cells were unstained, > 97% fell into quadrant 3 (Q3) for each cell line.



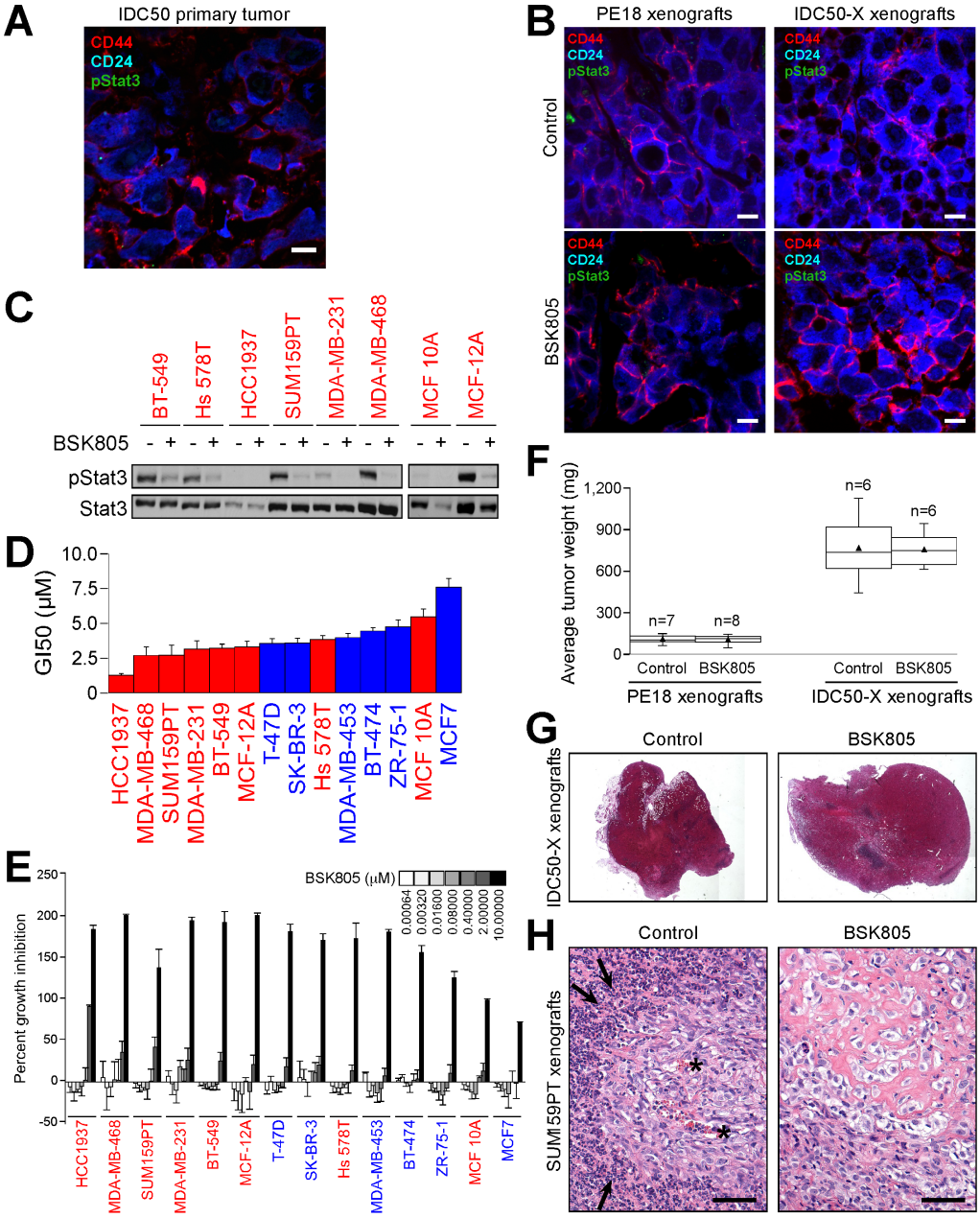
Supplemental Figure 2. Apoptosis and cell growth inhibition induced by *IL6* shRNAs, *STAT3* siRNAs, and the indicated inhibitors. (**A**) Results of Apo-ONE Homogeneous Caspase-3/7 Assay (Promega) using Hs 578T cells adjusted for cell number. *IL6* shRNAs that were and were not hits in the shRNA screen (shIL6 #1 and #2, respectively) do and do not statistically significantly increase apoptosis compared to shGFP control, respectively. Compared to non-targeting siRNAs, *STAT3* siRNAs increase apoptosis and, compared to no drug, the PTGIS, PFKFB3, and HAS1 inhibitors increase apoptosis (all statistically significantly). Compared to no drug, the JAK and CXCR2 inhibitors also increase apoptosis, and the NQO1 inhibitor decreases apoptosis (none statistically significantly). In A and B, infected cells were selected for five days, and treatments with siRNAs and inhibitors were for three days. Error bars (in A and C) show the standard deviations of triplicates. Double/triple asterisks indicate t-test p-values < 0.01/0.001, respectively. (**B**) Results of PI staining and flow cytometry cell cycle analysis using Hs 578T cells. The x-axes indicate DNA content based on PI staining level, and the y-axes indicate cell numbers. shIL6 #1 and #2 (compared to shGFP control), *STAT3* siRNAs (compared to non-targeting siRNAs), and all inhibitors (compared to no drug) increase the percentage of cells in the sub-G1 fraction determined using ModFitLT software (numbers and shaded regions on graphs). (**C**) Percent growth inhibition data for inhibitors in breast cell lines used to prepare the graphs in Figure 2D.



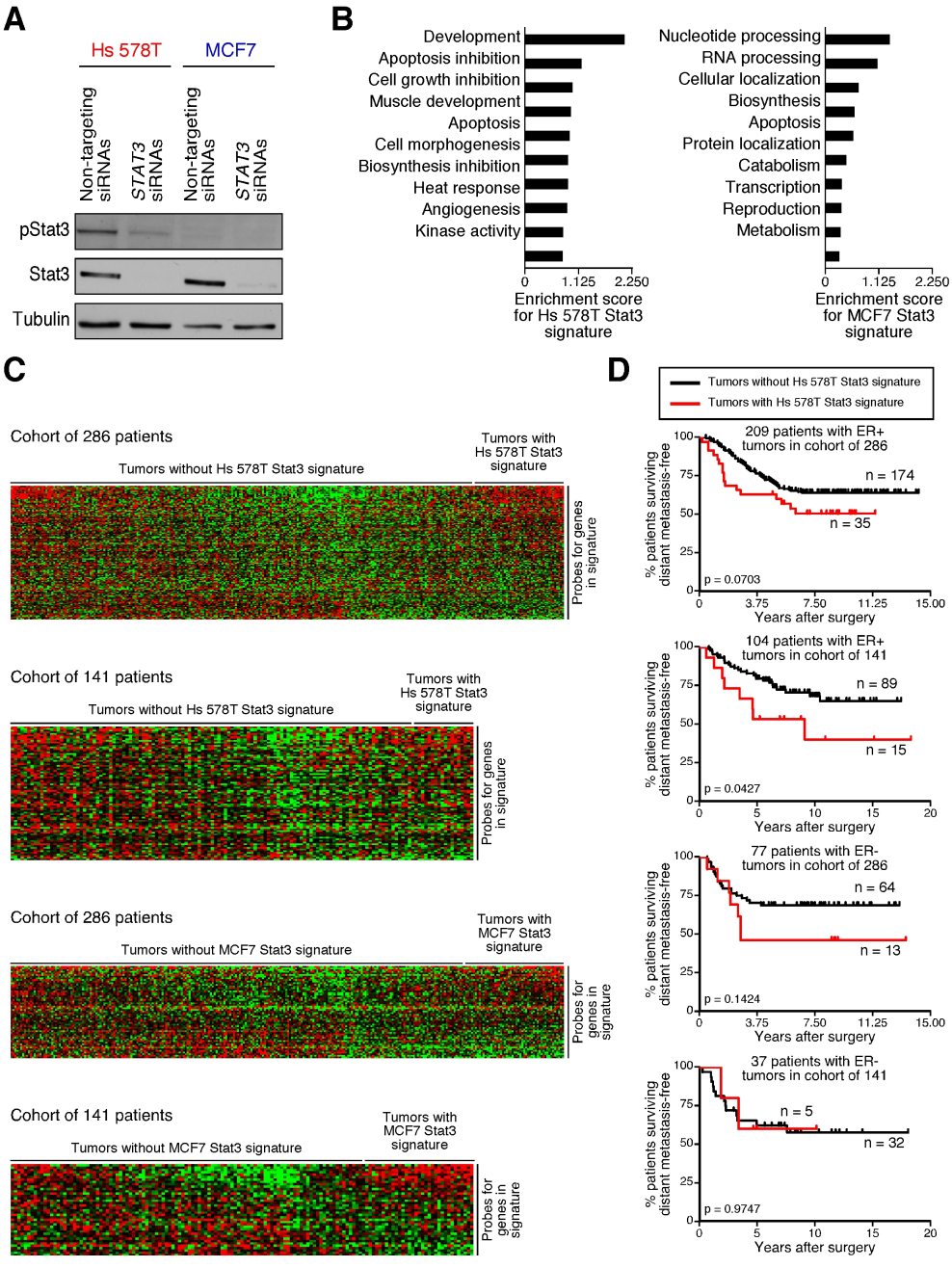
Supplemental Figure 3. Validation of target gene knockdown specifically by the *IL6* shRNA identified as a screening hit using Hs 578T cells. (**A**) qRT-PCR results. *IL6* shRNAs that were and were not hits in the shRNA screen (shIL6 #1 and #2, respectively) do and do not statistically significantly decrease *IL6* mRNA levels compared to shGFP control, respectively. Error bars (in A, B, and C) show the standard deviations of triplicates. Single/triple asterisks (in A, B, and C) indicate t-test p-values < 0.05/0.001, respectively. (**B**) ELISA results. shIL6 #1 and #2 statistically significantly decrease IL6 compared to shGFP control. (**C**) Changes in growth rate observed by cell viability assays after three days of treatment with recombinant human IL6 (rhIL6) compared to with vehicle alone. rhIL6 statistically significantly increases the growth of cells infected with shIL6 #1 but not #2. (**D**) Immunoblots. shIL6 #1 and #2 strongly and moderately decrease pStat3 compared to shGFP control, respectively. Stat3 and tubulin were used as loading controls.



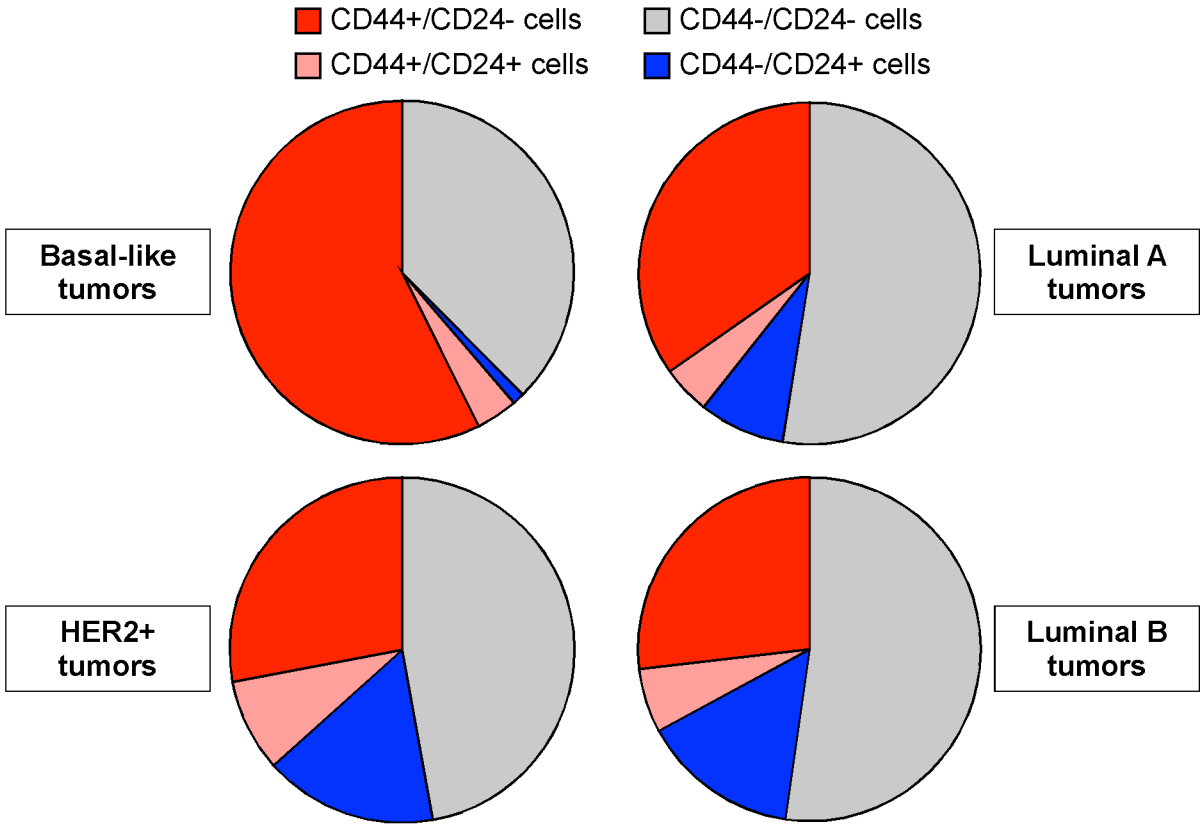
Supplemental Figure 4. Effects of the JAK2 inhibitor BSK805 in breast cell lines and tumors. (**A**) Representative immunofluorescence staining patterns for CD44, CD24, and pStat3 in ER-/PR-/HER2- sarcomatoid breast carcinoma (IDC50). The scale bar represents 10 microns. (**B**) Representative immunofluorescence staining patterns for CD44, CD24, and pStat3 in PE18 and IDC50 mouse xenograft-derived (IDC50-X) xenografts. Each scale bar represents 10 microns. (**C**) Immunoblots. Six-hour treatment with 3 μM BSK805 reduces pStat3 in basal-like breast cells containing it. Stat3 was used as a loading control. (**D**) GI50 (concentration inhibiting cell growth 50%) values for BSK805 in the indicated breast cell lines. Error bars (in D and E) show the standard deviations of triplicates. (**E**) Percent growth inhibition data for BSK805 in breast cell lines used to prepare the graph in D. (**F**) Boxplots showing xenograft tumor weights 63 or 5-17 days after injecting ER-/PR-/HER2+ breast cancer pleural effusion (PE18) or IDC50-X cells, respectively, (after tumors reached palpable size) into two fat pads of "n" mice. Mice were administered daily BSK805 (2 mg/mouse) or vehicle only (Control) for 24 or 6-12 days, respectively, beginning 40 or 0-6 days post-injection, respectively. Triangles mark averages. (**G**) Hematoxylin and eosin (H&E)-stained IDC50-X xenografts. (**H**) Representative H&E staining patterns observed in SUM159PT xenografts showing decreased leukocyte infiltration and angiogenesis with BSK805 treatment. Leukocytes are marked by arrows, and blood vessels are marked by asterisks. Each scale bar represents 50 microns.



Supplemental Figure 5. Stat3 knockdown for SAGE-Seq, Stat3 signatures gene ontology enrichment analysis and expression in breast tumors, and clinical significance of the Hs 578T Stat3 signature in ER-/ER+ tumors. (**A**) Immunoblots of cells used to generate SAGE-Seq libraries. Tubulin was used as a loading control. (**B**) Top ten gene ontology biological process categories represented by the genes in the Hs 578T (left) and MCF7 (right) Stat3 signatures, as determined using the DAVID Functional Annotation Tool with the highest classification stringency and *Homo sapiens* background employed. Genes involved in development are enriched (enrichment score > 2) in the Hs 578T Stat3 signature. (**C**) Expression levels of the genes in the Hs 578T (top) and MCF7 (bottom) Stat3 signatures in previously published microarray data for breast tumors from the two cohorts of patients analyzed. Red and green indicate high and low gene expression, respectively. In each data set, tumors were considered to have the Stat3 signature if they had average expression values for all genes in the signature downregulated by *STAT3* siRNAs above the 60th percentile and average expression values for all genes in the signature upregulated by *STAT3* siRNAs below the 40th percentile. Microarray data were processed as described previously. (**D**) Association of the presence of the Hs 578T Stat3 signature with shorter distant metastasis-free survival in ER+ but not ER- tumors from the two cohorts of breast cancer patients analyzed. Kaplan-Meier curves (for "n" patients with and without the signature) and their corresponding log-rank test p-values are shown. ER, estrogen receptor.



Supplemental Figure 6. Distribution of cell types within breast tumor subtypes. The pie charts shown depict the percentages of all cells observed within each of the four indicated breast tumor subtypes that belong to each of the four indicated cell types in the set of 170 tumors analyzed by immunofluorescence for CD44, CD24, and pStat3 expression. Tumors analyzed included 22 basal-like, 21 HER2+, 104 luminal A, and 12 luminal B. In each sample, 200 cancer cells were assessed for CD44 and CD24 expression.



SUPPLEMENTAL REFERENCES

1. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95:14863-14868.
2. Bloushtain-Qimron, N., Yao, J., Snyder, E.L., Shipitsin, M., Campbell, L.L., Mani, S.A., Hu, M., Chen, H., Ustyansky, V., Antosiewicz, J.E., et al. 2008. Cell type-specific DNA methylation patterns in the human breast. *Proc Natl Acad Sci U S A* 105:14076-14081.
3. Frank, D.A., Mahajan, S., and Ritz, J. 1997. B lymphocytes from patients with chronic lymphocytic leukemia contain signal transducer and activator of transcription (STAT) 1 and STAT3 constitutively phosphorylated on serine residues. *J Clin Invest* 100:3140-3148.
4. Polyak, K., Lee, M.H., Erdjument-Bromage, H., Koff, A., Roberts, J.M., Tempst, P., and Massague, J. 1994. Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. *Cell* 78:59-66.
5. Nikolsky, Y., Ekins, S., Nikolskaya, T., and Bugrim, A. 2005. A novel method for generation of signature networks as biomarkers from complex high throughput data. *Toxicol Lett* 158:20-29.
6. Wu, Z.J., Meyer, C.A., Choudhury, S., Shipitsin, M., Maruyama, R., Bessarabova, M., Nikolskaya, T., Sukumar, S., Schwartzman, A., Liu, J.S., et al. 2010. Gene expression profiling of human breast tissue samples using SAGE-Seq. *Genome Res* 20:1730-1739.
7. Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., et al. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365:671-679.
8. van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347:1999-2009.
9. Chang, H.Y., Nuyten, D.S., Sneddon, J.B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y.D., van't Veer, L.J., Bartelink, H., et al. 2005. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 102:3738-3743.
10. Shipitsin, M., Campbell, L.L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., Nikolskaya, T., Serebryiskaya, T., Beroukhim, R., Hu, M., et al. 2007. Molecular definition of breast tumor heterogeneity. *Cancer Cell* 11:259-273.
11. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4:P3.
12. Park, S.Y., Lee, H.E., Li, H., Shipitsin, M., Gelman, R., and Polyak, K. 2010. Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. *Clin Cancer Res* 16:876-887.