SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Src Activation in Different Subtypes of Breast Cancer, Related to Figure 1

(A) Distribution of SRS scores in different subtypes of breast cancer. Scores were calculated using sum of z-scores across all samples in a dataset for the 159 SRS genes. The scores were then linearly scaled between -1 and 1. The histograms of these scores in different subtypes of breast tumors are plotted. Roughly 90% of ER+ tumors, 50% of HER2+ tumors and 20% of triple-negative (TN) tumors are classified as SRS+.
(B) Distribution of SRS scores in different subtypes of breast cancer as derived by DART algorithm (Jiao et al., 2011).
(C) Western immunoblot showing Src phosphorylation at Y416 in response to ER activation by estradiol in MCF7 and T47D, two representative ER+ cancer cell lines.
(D) Proportion of viable cancer cells after treatment with various concentrations of tamoxifen and dasatinib for 3 days. Cell viability was assayed using Alarmar Blue staining and normalized to untreated controls (n=4).
(E) Effect of Src RNAi on ERBB2 and ERBB3 activation in SKBR3 and HCC1954, two representative HER2+ cancer cell lines. ERBB2 and ERBB3 activation was determined with antibodies against indicated activating phosphorylation sites.
(F) Western immunoblot showing Src phosphorylation at Y215 and Y416 in response to the ERBB3 ligand neuregulin.
(G) Proportion of viable cancer cells after treatment with various concentrations of lapatinib and dasatinib for 3 days. Cell viability was assayed using Alarmar Blue staining and normalized to untreated controls. Expected viability for dual treatment assuming no synergy was calculated as the product of viabilities of single treatments (grey bars). Difference between dual treatment and expected additive values was evaluated by Wilcoxon test (n=4).
(H) Proportion of viable cancer cells after treatment with the indicated drugs for 6 days. The concentrations of trastuzumab and dasatinib are 100ng/ml and 100nM, respectively. Cell viability was assayed using Alarmar Blue staining and normalized to untreated controls. Expected viability for dual treatment assuming no synergy was calculated as the product of viabilities of single treatments (grey bars). Difference between dual treatment and expected additive values was evaluated by Wilcoxon test (n=4).

Figure S2. CAF Signature is Associated with SRS Status and Bone Metastasis Relapse, Related to Figure 2

(A) Unsupervised hierarchical clustering of TN tumors from the TCGA dataset by CAF signature. SRS+ status is indicated by red sticks above the heatmap. *Small arrows* point to CXCL12 and IGF1 gene probes. *Red* and *green bars* left to the heatmap indicate genes that are over- or under- expressed in CAFs, respectively.
(B) Gene Set Enrichment Analysis revealed a strong association between SRS status and CAF enrichment in TCGA dataset.
(C) Scatter plot showing the correlation between SRS scores and CAF signature scores in TN tumor samples, color coded according to the content of αSMA+ cells (refer to Figure 2F). Pearson correlation coefficient and the corresponding p value are shown at the lower right corner. The p value is determined by Student's t-test with t-statistics transformed from the r value by $t = r*sqrt((n-2)/(1-r^2))$.
(D-E) Kaplan-Meier plots showing bone (E) and lung (F) metastasis free survival of patients in each category defined by IHC staining of αSMA. The p value is determined by log-rank test between low groups and medium-high groups. n.s.: no significance.
(F) Box-whisker plot showing the SRS scores by CAF signature status in different subtypes of breast cancer. P values were determined by Wilcoxon test.

Figure S3. Selection of Cancer Cells with Purified Cytokines in vitro, Related to Figure 3

(A) Western immunoblot analysis of Akt activation in response to different concentrations of IGF1 and/or CXCL12, in MDA231 cells.
(B) A representative image of the MDA231-CI cell colonies emerging after selection.
(C) Western immunoblot showing pY416-Src level after immuno-precipitation with anti-Src specific antibody in MDA231 and their in vitro selected derivatives. The specific higher Src activity was seen in MDA231-CI.
(D) Western immunoblot showing Akt and Src activation status in MDA231 cells. Cells were incubated with CXCL12 (30ng/mL) or IGF1 (10ng/mL) for 30 min. Dasatinib (100nM) was added 30 min before addition of CXCL12 and IGF1.
(E) Western immunoblot analysis of Akt activation in response to different doses of IGF1 and CXCL12 combination in MDA231-CI versus MDA231-0 cells.

Figure S4. Src Activity and Bone Metastasis, Related to Figure 4

(A) Lung colonization after tail vein injection of the indicated MDA231 subpopulations. Averaged photon flux values and representative images are shown. “n.s.”: no statistical difference. Error bars = SEM, n = 8-10 in each group.
(B) H&E staining of lung lesions derived from experiments in (A).
(C) Knockdown of Src in MDA231-CI cells with a control shRNA or a Src shRNA (Zhang et al., 2009), as assessed by western immunoblot analysis of Src levels.
(D) Bone metastasis assays of MDA231-CI Src-knockdown cells versus control cells. Bone metastasis development after intracardiac inoculation of the indicated cell lines was monitored by quantitative luciferase bioluminescence. Representative images of mice in each group are shown. n = 8-10 in each group. For quantification of the bioluminescence signal, refer to Figure 4E.
(E) In vitro selection of MDA231 cells following the protocol illustrated in Figure 3D, in media containing 0.2% bovine serum and 10 ng/mL of EGF. pY416-Src levels in the resulting populations were determined by western immunoblotting, with MDA231-0 and MDA231-CI cells as references.
(F) Bone metastasis development after intracardiac inoculation of the indicated cell lines was monitored by quantitative luciferase bioluminescence. Representative images of mice in each group are shown. n = 8-10 in each group. For quantification of the bioluminescence signal, refer to Figure 4G.
(G) Expression of bone osteolytic genes in Src^{high} and Src^{low} single cell clones from MDA231 parental population. Clones that share similar levels of bone osteolytic genes were pooled and tested in bone metastasis assays.
(H) Bone metastasis assays of Src^{high} and Src^{low} single cell clone pools. Bone metastasis development after intracardiac inoculation of the indicated cell lines was monitored by quantitative luciferase bioluminescence. Representative images of mice in each group are shown. n = 8-10 in each group. For quantification of the bioluminescence signal, refer to Figure 4H.
(I) Representative images of mammary tumors formed by three Src^{high} and three Src^{low} GFP+ clones.
(J) Representative images of spontaneous lung metastases in mice harboring the primary tumors shown in (I).
(K) Quantification of GFP+ cancer cells by flow cytometry analysis of femoral bone marrow that was flushed from mice harboring mammary tumors. Representative data are shown for mice harboring tumors generated with different Src^{high} or Src^{low} MDA231 clones.

Figure S5. In vitro Selection on CN34, Related to Figure 5

(A) Western immunoblot showing pY416-Src levels after immuno-precipitation with anti-Src specific antibody in CN34 and their in vitro selected derivatives. The specific higher Src activity was seen in CN34-CI.
(B) MDA231 cells were serially diluted and mixed with $5x10^6$ mouse bone marrow cells. RNA extraction was performed on these cell mixtures, followed by qRT-PCR. Human *B2M* expression was normalized against mouse *Actb* mRNA. Linear regression is plotted.

Figure S6. Production of CXCL12 and IGF1 by Cancer Cells and Mesenchymal Stem Cells (MSCs), Related to Figure 6

(A-B) Production of CXCL12 (A) and IGF1 (B) by breast cancer cell lines (MDA231, CN34 and CN37) and MSCs, cultured in regular media (DMEM supplemented with 10% fetal bovine serum, *D10F*), in MSC maintenance medium (Lonza, Cat. No: PT-3001), or in media conditioned by the indicated cancer cell lines. Conditioned media (C.M.) were collected from confluent monolayers over 2-3 days. MSCs were cultured in C.M. for 30 days, and CXCL12 and IGF1 levels were determined by ELISA.
(C) Expression of *CXCL12* and *IGF1* at transcriptional level by MDA231 and MSCs. MSCs were cultured either in MSC maintenance medium or cancer cell C.M.
(D) Cytokine gene transcript counts in polysome-associated transcriptomic data of MDA231 isolated from mammary tumors in Figures 6A-6C. All genes associated with cytokine activity (GO: 0005125) and growth factor activity (GO: 0008083) are listed. *CXCL12* and *IGF1* are not translated in cancer cells in situ. *IL8* and *TGFB1* are identified as positive references.

Figure S7. CXCR4 and IGF1R Signaling Mediates Pre-selection of Bone Metastatic Cells, Related to Figure 7

(A) MDA231 cells were pre-treated with CXCR4 inhibitor (AMD3100) or IGF1R inhibitor (BMS-754807) at different concentrations for 2 h, and then treated with CXCL12 (300ng/mL) or IGF1 (30ng/mL). Akt activation status was determined by western immunoblotting.
(B) Efficiency of CXCR4/IGF1R double-knockdown (DKD) in MDA231 parental cells. Two different shRNAs for each receptor were applied in DKD1 and DKD2.
(C) pY416-Src levels of cancer cell lines from individual mammary tumors formed by CN37 cells with or without admixed MSCs.
(D) Bone metastatic ability of pooled cancer cell lines derived in (C) .

SUPPLEMENTAL TABLES

Table S1. Bone Metastasis Incidence in Different Subtypes of Breast Cancer, Related to Figure 1

Incidence of relapse to bone and other sites in 615 patients from the EMC-MSK dataset stratified by ER and HER2 status. Incidents were recorded when patients were diagnosed with metastatic disease. Metastases in different sites were diagnosed simultaneously, and the temporal order of different metastases was not traceable.

	ER+	HER2+	Triple-negative
Relapse to bone	105	11	18
Relapse to bone and other sites	35	7	9
Relapse not to bone	33	9	37
Total cases	373	54	188

Table S2. Genes Associated with SRS in Triple-negative Breast Tumors, Related to Figure 1

Gene expression profiles of SRS+ tumors were compared with SRS- tumors within the triple-negative subtype in different data sets. EMC-MSK and three additional small datasets (GSE7390, GSE 4922, and GSE1456) were used as discovery sets. Overexpressed non-SRS genes with fold change > 2 and $p < 0.001$ in EMC-MSK set and $p < 0.05$ in the additional sets score as significant. These SRS-associated genes are validated in TCGA dataset. The false discovery rate was calculated for TCGA dataset using the "*siggene*" package of R. A combined p value was also computed for each gene in the discovery sets using Fisher's combination (see Experimental Procedures). NA: Probe not on the platform.

	Discovery					TCGA validation	
Gene	EMC-MSK p value	GSE7390 p value	GSE4922 p value	GSE1456 p value	Combined p value	p value	FDR
SNED1	1.16E-11	3.21E-02	6.98E-03	4.77E-02	1.10E-12	5.19E-17	0.000E+00
LIMA1	3.67E-11	8.61E-03	9.40E-04	3.20E-04	1.43E-15	5.93E-25	0.000E+00
CX3CR1	1.18E-10	3.81E-03	7.57E-06	3.27E-02	1.65E-15	9.39E-04	1.695E-03
SIDT1	4.19E-10	2.11E-03	2.59E-02	1.55E-08	7.56E-18	4.62E-21	0.000E+00
P2RY5	5.93E-10	2.64E-02	1.67E-04	9.01E-03	2.39E-13	2.92E-08	0.000E+00
PFDN5	6.12E-10	9.68E-03	9.20E-03	2.42E-02	9.66E-12	1.24E-09	0.000E+00
IGFBP4	7.58E-10	7.48E-03	1.71E-05	3.38E-03	4.53E-15	3.37E-08	0.000E+00
SEMA5A	1.21E-09	1.79E-03	3.62E-02	7.51E-03	4.61E-12	6.36E-09	0.000E+00
FAM13B1	1.35E-09	2.87E-03	4.47E-02	5.90E-05	1.11E-13	NA	NA
AHNAK	3.68E-09	3.92E-02	5.03E-05	2.37E-04	2.11E-14	8.29E-11	0.000E+00
TBC1D9	5.05E-09	2.60E-03	1.01E-08	6.15E-07	2.76E-21	8.60E-17	0.000E+00
CIRBP	5.98E-09	6.38E-04	4.07E-03	6.20E-06	1.44E-15	6.79E-12	0.000E+00
NPDC1	8.02E-09	4.85E-02	5.39E-03	3.88E-05	7.49E-13	1.18E-12	0.000E+00
RHOB	8.87E-09	2.89E-02	1.52E-05	3.58E-03	1.47E-13	1.13E-11	0.000E+00
RAI2	1.14E-08	2.42E-02	5.34E-06	6.58E-04	1.24E-14	8.36E-08	0.000E+00
TSC22D3	1.72E-08	5.72E-03	3.02E-02	1.82E-03	3.49E-11	7.60E-18	0.000E+00
XIST	2.30E-08	2.10E-03	4.09E-03	6.34E-04	1.11E-12	NA	NA
DACH1	3.20E-08	2.86E-02	9.89E-08	4.06E-04	5.87E-16	4.84E-09	0.000E+00
C4B	3.77E-08	4.12E-02	5.94E-05	4.18E-05	4.47E-14	NA	NA
PLSCR4	6.04E-08	2.54E-03	4.14E-04	1.65E-02	7.83E-12	2.35E-03	3.867E-03
CBX7	7.00E-08	2.25E-03	9.64E-04	1.19E-05	2.22E-14	2.92E-02	2.910E-02
FBP1	1.28E-07	4.03E-02	9.95E-03	7.27E-05	2.49E-11	5.24E-17	0.000E+00
CXCL12	2.13E-07	4.22E-03	1.30E-02	9.99E-04	7.05E-11	2.22E-07	0.000E+00
C4A	5.62E-07	2.85E-02	4.51E-04	2.58E-04	1.33E-11	NA	NA
AR	8.17E-07	4.98E-03	8.32E-03	6.97E-07	2.39E-13	8.43E-22	0.000E+00
IGF1	1.15E-06	8.12E-03	8.58E-03	8.32E-05	4.24E-11	1.76E-06	2.210E-06
CXCL14	2.31E-06	1.65E-02	2.01E-03	2.63E-04	1.16E-10	4.65E-07	0.000E+00
TMBIM4	2.49E-06	3.86E-04	1.15E-02	9.28E-03	5.07E-10	6.78E-16	0.000E+00
SLC44A4	2.51E-06	2.81E-02	1.40E-04	3.69E-14	1.14E-20	1.52E-15	0.000E+00
FBXL5	4.57E-06	2.06E-02	1.29E-02	8.82E-03	3.25E-08	4.97E-09	0.000E+00
MEG3	6.96E-06	4.59E-02	2.46E-02	4.45E-05	1.53E-09	4.31E-04	1.156E-03

IGF2	1.18E-05	1.58E-02	2.79E-03	4.84E-03	8.91E-09	7.25E-03	1.003E-02
KCNMA1	1.35E-05	7.24E-06	4.36E-02	2.56E-03	6.62E-11	9.69E-07	0.000E+00
ARHGAP6	2.70E-05	9.30E-03	4.99E-02	2.15E-03	7.32E-08	2.46E-05	4.228E-05
KIAA0141	5.96E-05	8.59E-03	4.57E-05	6.49E-06	1.33E-12	9.01E-08	0.000E+00
CRIP1	9.99E-05	3.76E-02	2.53E-02	1.88E-05	6.60E-09	1.18E-09	0.000E+00
C9orf116	1.07E-04	8.60E-03	1.58E-03	5.96E-04	3.43E-09	1.58E-11	0.000E+00
TPSB2	1.61E-04	7.55E-03	7.27E-06	1.05E-02	4.64E-10	NA	NA
EVC	1.62E-04	1.12E-02	3.92E-03	1.75E-03	3.69E-08	1.01E-06	2.210E-06
ITPR1	1.63E-04	2.37E-03	2.08E-04	5.33E-04	2.30E-10	4.23E-13	0.000E+00
SLC24A1	2.51E-04	2.97E-03	4.06E-05	1.23E-03	2.02E-10	2.17E-04	4.831E-04
HOXB2	7.85E-04	2.66E-02	2.25E-04	2.12E-02	2.32E-07	2.57E-09	0.000E+00
GALNT10	9.06E-04	1.27E-02	8.44E-05	1.40E-03	5.15E-09	1.62E-15	0.000E+00

Table S3. Responsive Genes of MDA231 cells to CXCL12 and IGF1, Related to Figure 3

MDA231 cells were treated with CXCL12 and IGF1 at intermediate concentration (30 ng/ml and 10 ng/ml, 30+10) or saturating concentration (300 ng/ml and 100 ng/ml, 300+100). Gene expression profiles were determined by TRAP RNA-seq. NA: no transcript detected. P values were computed by SAM algorithm.

Table S4. Primers for Quantitative PCR, Related to Experimental Procedures

Gene Name	Primer Sequences	
ADAMTS1	TCCGTCATAGAAGATGATGGTTT	GCATGTTAAACACGTGGCCTA
B2M	CTCCGTGGCCTTAGCTGTG	TTTGGAGTACGCTGGATAGCCT
CTGF	CTCCTGCAGGCTAGAGAAGC	GATGCACTTTTTGCCCTTCTT
CXCR4	GGTAGCGGTCCAGACTGATGA	CCTATGCAAGGCAGTCCATGT
FYN	AAGAGCCCCAGAAATTCACA	CGCCAACGATCACAAACTT
IL11	TATCCACTTGAGGGCGATTT	CTGCCCCAGTTACCCAAG
MMP1	CAGAGATGAAGTCCGGTTTTTC	GGGGTATCCGTGTAGCACAT
PTHrP	CTCGGTGGAGGGTCTCAG	TGGATGGACTTCCCCTTGT

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Bioinformatic Analyses

The continuous SRS scores were computed using $S = \sum_{i \in SRS} \sigma_i z_i$, where σ_i denotes the sign of gene i in SRS ($\sigma_i = +1$ for upregulated genes, and $\sigma_i = -1$ for downregulated genes), and z_i denotes normalized expression of gene i (mean zero and unit variance) across all samples in the dataset. S was then re-scaled linearly between -1 and 1. We also implemented DART (Denoising Algorithm based on Relevance Network Topology) (Jiao et al., 2011) to achieve more robust scoring of SRS status.

To compute the combined p value for each gene across all four datasets, we employed Fisher's combination, which is essentially based on chi-square. Specifically, for each gene g, a chi-square value was calculated by $-2 * \sum_{k=1}^{4} \ln P_{g,k}$ with degree of freedom as 2k (=8). A combined p value was then computed using chi-square distribution.

For analysis of stromal components, we used previously reported cancer-associated stromal cell signatures. A combined CAF signature was generated by union of the genes that are significantly over- or under-expressed in myofibroblast and fibroblast-enriched stroma (Allinen et al., 2004). Signatures of other stromal cell types included genes that are over- or under-expressed in tumor-associated leukocytes and myoepithelial cells (Allinen et al., 2004). The expression of CAF signature in individual tumors was either determined by unsupervised hierarchical clustering or by examining the mean expression of the upregulated genes. In unsupervised clustering analyses, patients that constitute the cluster with gene expression pattern matching CAF signatures (with higher expression of over-expressing genes and lower expression of under-expressing genes) were assigned to the CAF+ class. This classification is unambiguous because the expression of the genes comprising the signatures fluctuates as an ensemble. The signatures of tumor-associated leukocytes and myoepithelial cells were analyzed with the same approaches. Survival analyses were performed using the "*survival*" package of R. Log rank tests and Kaplan-Meier plots were computed and plotted by the "*survdiff*" command.

For *CXCL12* and *IGF1* as a two-gene classifier, we defined the tumors with expression values of *CXCL12* and *IGF1* both above the average values across all samples in the dataset as CXCL12/IGF1+. These tumors were then compared to the rest of tumors (CXCL12/IGF1- tumors).The nominal p value of the two-gene classifier was computed using log-rank test. We

also assess the significance of CXCL12 and IGF1 by randomly generating 1000 control gene pairs, repeating the same procedures to obtain 1000 nominal p values, and comparing these control p values to that of CXCL12 and IGF1. These procedures generated an empirically determined p value of 0.035 (CXCL12/IGF1 ranked 35th among the 1000 random gene pairs), and confirmed the significant association of CXCL12 and IGF1 with bone-metastasis free survival.

To evaluate the correlation between the CAF signature and SRS, Gene Set Enrichment Analysis (GSEA, http://www.broadinstitute.org/gsea/) was performed using overexpressed genes in CAF signature as the gene set, SRS status as the phenotype data, TN tumors from EMC-MSK as the dataset, and other parameters at default values.

Principal component analysis (PCA) and GSEA were based on RNA-seq of TRAP samples in the case of (1) MDA231 cells treated with or without CXCL12/IGF1, and (2) of MDA231 tumors supplemented with or without MSCs, or based on microarray transcriptomic data in the case of (3) TN tumors from the EMC-MSK cohort. To be able to combine these datasets in PCA analyses, we first performed Z-transformation for each gene in each individual dataset, using the formula $Zi=(Xi-mean(X))/sd(X)$, where *Zi* represents the z-score of the ith sample, Xi is the expression value of this gene, and mean(X) and sd(X) represent the average value and standard deviation of the ith gene across all samples, respectively. We then combined datasets (1) and (2), or (2) and (3) to generate Figures 6G and 6B-C, respectively. Specifically, for PCA in Figure 6G, we removed genes whose standard deviation < 1 across samples, implemented the analysis using the "prcomp" command of R, and plotted the first two principal components. For PCA in Figure 6B-C, we used a standard deviation of 0.5 as the cutoff, and followed procedures similar to above. For PCA in Figure 3B, the results were plotted using the first three principal components using "*scatter3Dplot*" package of R. We also employed GSEA to examine how dataset (1) correlates with dataset (3) as in Figure 3C. To this end we first isolated genes that are up-regulated by CXCL12 and IGF1, with a fold-change >2 and $p < 0.05$ as criteria (Table S3). The resulting gene sets were subjected to GSEA analysis as described above.

Ingenuity pathway analysis (IPA) was performed using the IPA online toolkit. Differentially expressed genes from (2) were loaded as the input dataset, and results from upstream regulator analysis and functional annotation were exported. P-value was calculated by fisher exact test, on overlap between the differentially expressed gene category and gene categories associated with the regulatory or functional annotations. Z-score is based on the direction of fold change values for genes in the input dataset for which a correlation has been established in the

Ingenuity Knowledge Base. Z-score > 2 indicates “activated/increased”; z-score < -2 indicates “inhibited/decreased”.

Cell Culture

MCF7, T47D, SKBR3, HCC1954, and MDA-MB-231 cells were obtained from ATCC and were cultured under conditions recommended by ATCC. Bone metastasis-derived cells of MDA231 (BoM1833) have been described (Kang et al., 2003). CN34 and CN37 have been described previously (Bos et al., 2009; Padua et al., 2008) and were maintained in M199 medium supplemented with 2.5% FBS, 20 ng/ml of EGF, 10 ng/ml of insulin, 0.5 ug/ml of hydrocortisone, and 100 ng/ml of cholera toxin. Primary human mesenchymal stem cells (Lonza, Cat. No: PT2501) were cultured following the supplier’s instructions. CAFs from a human breast tumor (Asterand, Cat. No: 60266A1) were cultured following the supplier’s instructions. Neuregulin**,** CXCL12 and IGF1 were purchased from R&D Systems, estradiol from Sigma, and Lapatinib from ChemieTek. Alamar blue assays were performed using Resazurin (R&D Systems) following manufacturer’s procedures. CXCL12 and IGF1 ELISA kits were from R&D Systems.

Cell Sorting and Analysis

MDA231 cells were individually sorted by flow cytometry to one well in a 96-well plate to form independent clonal cell lines. pY416-Src level of each clonal line was tested with Pathscan pY416-Src ELISA Kit (Cell Signaling). The parental population was used as a reference. For phospho-flow cytometry, single cell suspensions were fixed in Fix Buffer I (BD Biosciences) for 10 minutes at 37°C, followed by permeablization with Perm Buffer III (BD Biosciences) for 12 h at -20°C. Cells were then stained with Alexa Fluor 647-conjugated pY416-Src antibody (BD Biosciences), and applied to flow cytometry. Data were analyzed with FlowJo.

Cancer Cell-specific Polysome-associated Transcript Profiling

Cultured cells or tumors were lysed and immediately subjected to TRAP as previously described (Heiman et al., 2008) with the following modifications: RNasin Plus RNase inhibitor (Promega, Cat No: N2615) was used as RNase inhibitor, and anti-GFP coated sepharose beads (GE Healthcare) were used for immunoprecipitation. Polysome-associated RNA was purified with RNAqeuous micro kit (Life Technologies). After Ribogreen quantification and quality control of Agilent BioAnalyzer, 500ng RNA (RIN >7) from each sample was used for library construction with TruSeq RNA Sample Prep Kit v2 (Illumina) following the manufacturer’s instructions. Samples were barcoded and run on a Hiseq 2000 in a 75bp/75bp paired-end run, using the TruSeq SBS Kit v3 (Illumina). An average of 44 million paired reads were generated per

sample. Ribosomal RNA reads represented <0.25% of total and the proportion of mRNA bases was approximately 77% on average.

Quantitative PCR and Western Immunoblotting

Total RNA was extracted using PrepEase RNA spin kit (USB). 1μg RNA was subjected to reverse transcription with Transcriptor first strand cDNA synthesis kit (Roche). cDNA corresponding to approximately 5 ng of starting RNA was used for quantitative PCR in quadruplicate. Primers for SYBR Green Mix reaction (Thermo Scientific) are listed in Table S4. Expression data were acquired and analyzed using the ABI 7900HT Fast Real-Time PCR System (Applied Biosystems). Western immunoblotting was performed as previously described (Minn et al., 2005). Antibodies against Src, pY416-Src, pS473-Akt, Akt, pY877-ERBB2, pY1221/2-ERBB2, pY1289-ERBB3, IGF1R, and α-tubulin were purchased from Cell Signaling and used following the manufacturer's instructions.

Generation of Knockdown Cells

Stable shRNA-mediated knockdown was achieved by targeting the following sequences: GGACCTTCCTCGTGCGAGA for c-Src in pSuper-Retro vector (Zhang et al., 2009); ACCAGCTGTTTATGCATAGATA (shRNA1) and ACTGAGTCTGAGTCTTCAAGTT (shRNA2) for CXCR4 in pGIPZ vector; CGGAGAGAACTGTCATTTCTAA (shRNA1) and CGCCGAAGATTTCACAGTCAAA (shRNA2) for IGF1R in pGIPZ vector (Open Biosystems). Virus-infected cells were selected with 5 μg/ml puromycin. Knockdown efficiency was confirmed by quantitative RT-PCR and western immunoblot. CXCR4 surface expression was assessed by flow cytometry with APC-conjugated CXCR4 antibody (BD Biosciences, Cat No: 560936).

Spontaneous Bone Metastasis Assay

In order to record spontaneous metastases at different organ sites from mammary tumors, cancer cells were labelled with turbo-GFP, which generates strong fluorescent signal that can penetrate tissue efficiently. Nine weeks after mammary fat pad injection, different organs from mice including hind limbs, spines, lung and primary tumors were retrieved and examined under a fluorescence dissecting stereomicroscope (Nikon).

Indolent Metastasis Assay

Cell suspensions containing 10^5 cancer cells were injected into the left cardiac ventricle of 7-week-old Cr:NIH-bg-nu-Xid mice. Eight weeks later, mice without detectable bioluminescent signals ($<10^4$ photon per second) were euthanized. Tibiae and femora were immediately

extracted, and minced in PBS. Bone marrow cells were flushed out by repeatedly centrifugation and vortexing. Bone marrow cells were then collected by centrifugation of the suspension, and subjected to RNA extraction and quantitative PCR (see above). Human *β2-microglobulin (B2M)* mRNA expression was used to quantify cancer cell abundance, with mouse *β-actin* mRNA as a control to normalize the data.

Histological Analysis

Hind limb and vertebral bones were excised, fixed in 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA), decalcified with folic acid, embedded in paraffin and subjected to staining with hematoxylin and eosin (H&E, Richard-Allan Scientific Inc.). Mammary tumors were fixed in 4% paraformaldehyde, embedded in OCT compound (Tissue-Tek), sectioned at 8μm, and stained with anti-αSMA (clone 1A4, 1:400 dilution, Dako), anti-CXCL12 (Clone C19, 1:50 dilution, Santa Cruz Biotechnology) and anti-IGF1 antibodies (Clone H-70, 1:50 dilution, Santa Cruz Biotechnology). Corresponding secondary antibodies and ABC detection kit (Vector laboratories) were used for detection following manufacturer's instructions. Tumor tissue microarrays containing TN breast cancer clinical samples were sectioned at 5μm thick, baked for 1h at 56°C, de-paraffinized and treated with 3% hydrogen peroxide for 10 min. Antigen retrieval was performed in citrate buffer (pH6.1, Dako), and stained with anti-αSMA (clone 1A4, 1:400 dilution, Dako). Total immunoreactivity of staining was evaluated and scored by pathologists (Y.H.W. and E.B.) in a blinded fashion.

Statistical Analyses

The sample sizes or numbers of replicates in each experiment have been indicated in the corresponding figure legends. Results are reported as mean ± SEM unless otherwise specified. For other analyses, Wilcoxon tests were used unless noted otherwise.

SUPPLEMNTAL REFERENCES

Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A.*, et al.* (2009). Genes that mediate breast cancer metastasis to the brain. Nature *459*, 1005-1009.

Padua, D., Zhang, X.H., Wang, Q., Nadal, C., Gerald, W.L., Gomis, R.R., and Massague, J. (2008). TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. Cell *133*, 66-77.