Supplementary Figure S1

A. s.c. tumor

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

B. s.c. tumor

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

C. Tumor

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

D. Tumor

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

E. E-cad/ZEB1

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

F. E-cad/FN

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

G. E-cad/ZEB1-APC

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

H. Tumor weight (g)

I. Metastatic index (mm²/g)

J. Lung

K. Liver

L. Lung

<table>
<thead>
<tr>
<th>Condition</th>
<th>No fibro.</th>
<th>+cont. fibro.</th>
<th>+CAFs</th>
</tr>
</thead>
</table>

Tomato
Supplementary Figure S2

A

B

C

D

E

F
Supplementary Figure S5

Summary of the high-throughput screening

<table>
<thead>
<tr>
<th>Inhibitors</th>
<th>Target signal</th>
<th>% inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP1 (analog)</td>
<td>Src, Fyn, Lck</td>
<td>53.1</td>
</tr>
<tr>
<td>ERK inhibitor II</td>
<td>MAPK</td>
<td>56.6</td>
</tr>
<tr>
<td>U-0126</td>
<td>MEK</td>
<td>68.4</td>
</tr>
<tr>
<td>MEK inhibitor I</td>
<td>MEK</td>
<td>53.6</td>
</tr>
<tr>
<td>Gefitinib</td>
<td>EGFR</td>
<td>52.5</td>
</tr>
<tr>
<td>Vandetanib</td>
<td>multi-kinase</td>
<td>50.3</td>
</tr>
</tbody>
</table>

The total number of inhibitors: 358
The number of inhibitors that inhibited CAM6 expression (>50%): 6

B) DCIS\textsuperscript{cont2cy} and DCIS\textsuperscript{CAF2cy}

C) Fold induction

D) Fold induction

E) CAM5 and CAM6

F) DCIS\textsuperscript{cont2cy} and DCIS\textsuperscript{CAF2cy}

G) DCIS\textsuperscript{CAF2cy}

H) Fold induction

I) p-Src, Src, α-tub

J) Metastasis (mm²)

K) Signal intensity

L) Cont. vector, Active Src
Supplementary Figure S7
Supplementary Figure S9

A

CAM6

DCIS<sub>cnt2cy</sub>  59%  86%  79%  74.77

DCIS<sub>CAF2cy</sub>  59%  82%  72%  71.07

DCIS+H₂O  68%  88%  78%  78.23

DCIS+5AZ  62%  79%  71%  70.98

CAM5

DCIS<sub>cnt2cy</sub>  3%  3%  3.32

DCIS<sub>CAF2cy</sub>  3%  3%  2.98

E-cad

DCIS<sub>cnt2cy</sub>  0%  0%  0%  0

DCIS<sub>CAF2cy</sub>  0%  0%  0%  0
SUPPLEMENTARY FIGURE LEGENDS

Figure S1, Related to Figure 1. CAF-induced breast carcinoma cell clusters with the E\textsuperscript{hi} and E/M states, collective invasion and metastasis

A. Appearance of 21-day-old tdTomato\textsuperscript{+} DCIS tumor xenografts subcutaneously (s.c.) developed with no fibroblasts (No fibro.), control fibroblasts (+cont. fibro.) or CAFs (+CAFs) in mice. Note the decreased number of intact acini (arrows) in tumors admixed with CAFs.

B. Detection of tdTomato\textsuperscript{+} DCIS cells and GFP\textsuperscript{+} fibroblasts (arrowheads) in 21-day-old subcutaneous tumor xenografts.

C. Immunohistochemistry of sections prepared from subcutaneous DCIS tumors developed with no fibroblasts, control fibroblasts or CAFs using human-specific anti-vimentin (Vim) antibody. Note the presence of the injected human vimentin-positive control fibroblasts (arrowheads, middle) and CAFs (arrowheads, right) in 21-day-old tumor xenografts.

D. Immunohistochemistry of sections prepared from 21-day-old DCIS tumors subcutaneously developed with no fibroblasts, control fibroblasts or CAFs using antibodies for human-specific anti-vimentin (Vim) and anti-fibronectin (FN). The positively stained tumor cells (arrows) and stromal cells (arrowheads) are shown.

E. Immunofluorescence of sections prepared from 21-day-old subcutaneous DCIS tumors developed with control fibroblasts or CAFs using anti-E-cadherin (E-cad) (Abcam, Cat#ab40772) and -ZEB1 (Santa Cruz, Cat#sc-515797) antibodies. The E\textsuperscript{hi} (simple arrow) and E/M (triangular arrow) cancer cells, as well as nuclear ZEB1\textsuperscript{+} stromal cells (arrowheads), are shown.

F. Immunofluorescence of sections prepared from DCIS tumors admixed with CAFs using the indicated antibodies. E-cad\textsuperscript{FN\textsuperscript{+}} and E-cad\textsuperscript{Vim\textsuperscript{+}} tumor cells (arrows) and FN\textsuperscript{+} stromal cells (arrowheads) are shown.

G. Flow cytometry of the cell suspension dissociated from 30-day-old tumor xenografts raised by tdTomato\textsuperscript{+} DCIS cells admixed with CAFs using anti-E-cad and -ZEB1 antibodies. E\textsuperscript{hi}, E/M and E-cad\textsuperscript{hi}ZEB1\textsuperscript{hi} cell populations are marked.

H. Tumor weight measured at 30 days after subcutaneous injection of the indicated cells into mice.

I. Lung metastatic indices evaluated by nodule volume at 60 days after subcutaneous injection of the indicated cells into mice.

J. Immunohistochemistry of sections prepared from the lungs at 60 days after subcutaneous injection of DCIS cells admixed with CAFs into mice, using antibodies for human-specific anti-vimentin (Vim). Note the presence of vimentin-positive DCIS cells.
(arrows), but not CAFs.

K and L. Appearance of tdTomato+ metastatic nodules in the indicated organs dissected from mice at 60 days after subcutaneous injection of tdTomato+ DCIS cells with CAFs (K) or control fibroblasts (L). Note the tdTomato+ metastatic nodule (arrow) in the liver of the animal from the CAF group, while there are no tdTomato+ cells in the indicated organs in the control fibroblast group. The image (L-lung) is also shown, as in Fig. 1G.

Data information: Star indicates intact acinar structure of DCIS cells (B and D). Scale bars, 1 mm (A, K and L), 30 µm (B, C, D and J) and 10 µm (E, F). Asterisk indicates a significant difference relative to No fibro. and +cont. fibro. groups (H and I). Student’s t test (H) and Wilcoxon rank sum test (I). The horizontal line represents the mean value (H and I).
Figure S2, Related to Figure 2. Highly invasive and metastatic breast cancer cells generated by intra-tumoral CAFs

A. Immunofluorescence of sections prepared from 5-day-old tumor organoids formed by DCIS<sup>CAF<sub>2cy</sub></sup> and DCIS<sup>cnt2cy</sup> using anti-E-cad and -ZEB1 antibodies. E<sup>Hi</sup> (simple arrow) and E/M (triangular arrow) tumor cells are shown.

B. Immunostaining of sections prepared from 21-day-old subcutaneous tumors generated by the indicated cells using anti-fibronectin (FN) antibody and human-specific anti-vimentin (Vim) antibody. Positively stained tumor cells (arrows) and stromal cells (arrowheads) are shown. Star indicates intact acini in DCIS tumors.

C. Tumor weight measured at 30 days after subcutaneous injection of the indicated cells into mice. The horizontal line represents the mean value. Asterisk indicates a significant difference relative to DCIS<sup>alone2cy</sup> using Mann-Whitney U test.

D. Appearance of lung metastatic nodules (arrows) at 30 days after intravenous injection of the indicated cancer cells into recipient mice.

E. H&E staining of the lung sections prepared from mice at 30 days after intravenous injection of the indicated cancer cells.

F. Immunofluorescence of the indicated cells using anti-E-cad and -ZEB1 antibodies. E-cad<sup>+</sup> cancer cells (simple arrows) and nuclear ZEB1<sup>+</sup> cancer cells (triangular arrows) are also shown.

Data information: Scale bars, 10 µm (A and F), 30 µm (B), 1 mm (E) and 5 mm (D).
Figure S3, Related to Figure 3. The E$^\text{hi}$ state mediated by E-cad, CAM5 and CAM6 expression in DCIS$^{\text{CAF2cy}}$

A. Kaplan-Meier survival analysis for lung metastasis-free survival (MFS) and distant metastasis-free survival (DMFS) using the CAF-induced metastasis signature (CIMS) in the indicated human breast cancer patient cohorts.

B. Immunofluorescence (left) and real-time PCR (right) of the indicated cells measuring E-cad expression. E-cad$^+$ cancer cells (arrows) are shown. Scale bar, 30 µm.

C. Real-time PCR of the indicated cells using primers specific for CAM5 and CAM6.

D. Positive linear correlations between E-cad, CAM5 and CAM6 mRNA expressions in the GSE14333 cohort and the human TCGA breast cancer cohort.

E. Signal intensity differences between CAM5/E-cad, CAM6/E-cad and CAM5/CAM6 on the indicated cells expressing various shRNAs evaluated by in situ PLA. The horizontal line represents the mean value.

Data information: Asterisk indicates a significant difference relative to the CIMS$^-$ group (A), DCIS$^{\text{cnt2cy}}$ (B), the control group (C) and GFP-shRNA-expressing DICS$^{\text{CAF2cy}}$ (E). Student’s t-test (B and C), Mann-Whitney U test (E) and Cox proportional hazards regression test (A). Error bars, SE.
Figure S4, Related to Figure 4. The Ehi state required for invasive and metastatic abilities in DCIS\textsuperscript{CAF2cy}

A-C. Measurement of tumor weight at 30 days after subcutaneous injection of the indicated cells into recipient mice.

D. Cell-cell adhesion of GFP-labelled DCIS\textsuperscript{alone2cy}, DCIS\textsuperscript{cnt2cy} or DCIS\textsuperscript{CAF2cy}, seeded as adherent cells, on the indicated GFP-negative layer cells. The number of GFP-positive adherent cells was quantified in each group.

Data information: Asterisk indicates a significant difference relative to GFP-shRNA-expressing DCIS\textsuperscript{CAF2cy} (A-C) and the adherent GFP-positive DCIS\textsuperscript{CAF2cy} seeded on DCIS\textsuperscript{CAF2cy} layer cells (D). Student’s t-test (A-C) and Mann-Whitney U test (D). Error bars, SE. The horizontal line represents the mean value (A-C).
Figure S5, Related to Figure 6. Src activation mediates the $E^{hi}$ and E/M states and metastatic ability in DCIS$^{CAF2cy}$

A. Summary of high-throughput screening identifying 6 compounds that significantly (>50%) inhibited CAM6 mRNA expression in DCIS$^{CAF2cy}$.

B. Immunostaining of frozen sections prepared from the indicated tumors using an anti-phosphorylated Src (p-Src, Tyr 416) antibody. p-Src$^+$ cancer cells (simple arrows) are shown.

C. Real-time PCR of DCIS$^{cn2cy}$ and DCIS$^{CAF2cy}$ expressing the indicated shRNA using primers specific for the Src gene.

D. Real-time PCR of the indicated cells treated with or without saracatinib (Sara) for 24 hr measuring CAM5, CAM6 and E-cad expressions.

E. Real-time PCR of DCIS$^{cn2cy}$ and DCIS$^{CAF2cy}$ expressing the indicated shRNA using primers specific for CAM5, CAM6 and E-cad genes. Immunofluorescence of the indicated cells using anti-E-cad antibody (right). E-cad$^+$ cancer cells (simple arrows) are also shown.

F. In situ PLA of DCIS$^{mt2cy}$ and DCIS$^{CAF2cy}$ expressing the indicated shRNA using anti-Src and -E-cad antibodies. The signal (arrow) detected by in situ PLA is indicated (upper) and the signal intensity is also evaluated in the above-described cells (lower).

G. Immunofluorescence of DCIS$^{CAF2cy}$ using the indicated antibodies.

H. Real-time PCR of the indicated cells measuring ZEB1 expression.

I. Immunoblotting of DCIS cells expressing the control empty vector (Cont. vector) or constitutively active Src mutant (Active Src) using the indicated antibodies.

J. Measurement of nodule volume in the lungs at 30 days after intravenous injection of the indicated cells into mice.

Data information: Asterisk indicates a significant difference relative to DCIS$^{CAF2cy}$ expressing GFP-shRNA (C, E, F and H), DCIS$^{CAF2cy}$ treated without saracatinib (D) and DCIS cells expressing the control vector (J). Student’s t-test (C-E and H) and Mann-Whitney U test (F and J). Error bars, SE. The horizontal line represents the mean value (F and J). Scale bars, 30 µm (B and F), 10 µm (E and G).
Figure S6, Related to Figure 7. Stromal SDF-1 and TGF-β mediate the formation of invasive and metastatic breast tumor clusters with Ehi and E/M states via Src activation

A. Real-time PCR of DCIScnt2cy and DCISCAF2cy using the indicated primers. The DCIScnt2cy and DCISCAF2cy were extracted from tumor xenografts generated by DCIS cells admixed with control fibroblasts and CAFs expressing GFP or TβRII ecto, respectively.

B. Real-time PCR of DCIScnt1cy and DCISCAF1cy measuring E-cad expression. The DCIScnt1cy and DCISCAF1cy were extracted from 30-day-old tumor xenografts generated by DCIS cells expressing the indicated shRNAs admixed with control fibroblasts and CAFs, respectively.

C. Immunofluorescence of DCIS cells treated with recombinant SDF-1 (100 ng/mL) and/or TGF-β1 (10 ng/mL) for 48 hr using anti-E-cad or ZEB1 antibody. E-cad+ cancer cells (simple arrows) and nuclear ZEB1+ cancer cells (triangular arrows) are shown. Scale bars, 10 µm.

D. Real-time PCR of MCF-7-ras cells treated with recombinant SDF-1 and/or TGF-β1 for 24 hr to measure the indicated gene expressions.

Data information: Asterisk indicates a significant difference relative to DCISCAF2cy extracted from tumor with CAFs expressing GFP (A) and GFP-shRNA-expressing DCISCAF1cy (B), and between the depicted groups (D). Student’s t-test (A, B and D). n.s.: not significant (D). Error bars, SE.
Figure S7, Related to Figure 8. CAF-induced CTC clusters, tumor emboli and metastatic colonization

A. Immunohistochemistry of lung sections prepared from mice at 60 days after subcutaneous injection of DCIS\textsuperscript{CAF2cy} using anti-\(\alpha\)-SMA antibody. Tumor embolus (asterisk in a broken circle) is indicated. \(\alpha\)-SMA-positive cells (arrows) are detected in smooth muscle cell layers in the blood vessel and bronchus (star).

B. Immunostaining of lung sections prepared from mice at 60 days after subcutaneous injection of DCIS cells admixed with CAFs, using the indicated antibodies. Note that DCIS cells show highly positive staining for E-cad, CAM5, CAM6 and p-Src (arrows), in contrast to negative staining for ZEB1.

Data information: Scale bars, 100 \(\mu\)m (A) and 30 \(\mu\)m (B).
Figure S8, Related to Figure 9. The $E^{hi}$ and E/M states in DCIS$^{\text{CAF2cy}}$ are associated with poor outcomes in Her2$^+\text{ER}^-\text{PR}^-$ breast cancer patients

A. Kaplan-Meier survival analysis for high (red line) and low (black line) expression levels of the indicated genes in the described breast cancer patients. The rectangle indicates data shown in Fig. 9D. RFS: relapse-free survival, OS: overall survival. The hazard ratio (HR) is also shown. n.s.: not significant. Asterisk indicates a significant difference relative to the group with lower expression using the Cox proportional hazards regression test.
Figure S9 No significant changes in DNA methylation status in the E-cad, CAM5 and CAM6 gene promoter regions in DCISCAF2cy

A. Analysis of DNA methylation in the indicated DCIS cells by pyrosequencing using the promoter regions of the CAM6, CAM5 and E-cad genes. Percent methylation is depicted above each CpG site in representative pyrograms and the averaged percent methylation is shown on the right side. Note slightly decreased percent methylation in the CAM6 gene in DCISCAF2cy (71.07%) relative to that in DCIScm2cy (74.77%). Exposure of DCIS cells to 20 μM 5-azacytidine (5AZ) for 72 hr also decreases the percent methylation in the CAM6 gene relative to the control H2O treatment. In contrast, methylation is barely detectable in the examined CpG sites in the CAM5 and E-cad genes in both DCISCAF2cy and DCIScm2cy.
Table S1, related to Fig 3. CIMS

Table S2, related to Fig 9. Associations between CEACAM6-positivity in breast cancer tissues and patient characteristics (n=257)

Table S3, related to Fig 9. Associations between CEACAM5-positivity in breast cancer tissues and patient characteristics (n=257)

Table S4, related to Fig 9. Associations between E-cadherin-positivity in breast cancer tissues and patient characteristics (n=257)

Table S5, related to Fig 9. Associations between ZEB1-positivity in breast cancer tissues and patient characteristics (n=257)

Table S6, related to Materials and Methods. Oligonucleotide sequences.

Table S7, related to Materials and Methods. Key Resources (Antibodies, Recombinant proteins and Software and Algorithms).