SUPPLEMENTAL DATA

Experimental Procedures

Cells and Cell Culture

Human breast carcinoma cell lines, MDA-MB-231 and MCF7, were purchased from American Type Tissue Culture Collection (ATCC). 231BoM-1833, 231BrM-2a, CN34, CN34-BoM2d, CN34-BrM2c and MCF7-BoM2d cell lines were kindly provided by Dr. Joan Massagué (Memorial Sloan-Kettering Cancer Center) (1-3). Luciferase-labeled cells were generated by infecting the lentivirus carrying the firefly luciferase gene. The immortalized mouse bone microvascular endothelial cell (mBMEC) was a generous gift from Dr. Isaiah J. Fidler (M.D. Anderson Cancer Center) (4). MCF10A and MCF10DCIS.com cells were purchased from ATCC and Asterand, respectively. MDA-MB-231, its variant cells, MCF7 and MCF-BoM2d cells were cultured in DMEM medium supplemented with 10% FBS and antibiotics. CN34 and its variant cells were cultured in Medium199 supplemented with 2.5% FBS, 10 µg/ml insulin, 0.5 µg/ml hydrocortisone, 20 ng/ml EGF, 100 ng/ml cholera toxin and antibiotics. MCF10DCIS.com cells were cultured in RPMI-1640 medium supplemented with 10% FBS and antibiotics. MCF10A cells were cultured in MEGM mammary epithelial cell growth medium (Lonza). mBMEC was maintained at 8% CO_2 at 33 °C in DMEM with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate, 1% non-essential amino acids and 1% vitamin mixture. Bone marrow stromal fibroblast cell lines HS5 and HS27A, and osteoblast cell line, hFOB1.19, were purchased from ATCC. Bone marrow derived human mescenchymal stem cells, BM-hMSC, were isolated for enrichment of plastic adherent cells from unprocessed bone marrow (Lonza) which was depleted of red blood cells. To confirm the purity of the cells, the lack of CD34 and CD45 and expression of CD44 and CD29 were examined. Finally, multipotency of the cells to differentiate into adipogenic and osteogenic lineages were also confirmed. BM-hMSC was maintained in minimum essential medium supplemented with 20% FBS and antibiotics. HS5 cell line was cultured in DMEM medium supplemented with 10% FBS and antibiotics. HS27A was cultured in RPMI-1640 medium supplemented with 10% FBS and antibiotics. hFOB1.19 was cultured in DMEM medium without phenol red supplemented with 10% FBS and antibiotics at 34 °C, and it was cultured at 39°C for 24 hours to be differentiated into mature phenotype before using. Knockdown of HAS2 was achieved by infecting the lentivirus carrying shRNA to HAS2 gene (Open Biosystems). The efficiency of the knockdown was confirmed by quantitative PCR. CSCs were isolated by MACS system (Miltenyi Biotec) using antibodies to CD24 (Stem cell technologies), CD44 (Biolegend) and ESA (GeneTex).

Gene-expression microarray profiling

RNA was extracted from isolated CSCs using the RNeasy mini kit (Qiagen) followed by DNase treatment and re-purified using the RNA cleanup kit (Qiagen). Labeling and hybridization of the samples to Human gene 1.0ST chip (Affymetrix) which contains more than 750,000 probes were performed by the CFG Microarray Core Facility (New York). Normalization of the chip was performed using RMA algorism in Expression consol software (Affymetrix). These expression data were submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE25976. Clustering and its visualization were performed using Cluster and TreeView softwares. Gene set enrichment analysis was performed using the package “PGSEA” in the R environment (http://cran.r-project.org). Gene sets were adopted from the previous publication (5).

Kinexus antibody microarray profiling

All procedures were carried out by following the manufacturer’s protocols. Briefly, isolated CSCs were lysed by lysis buffer with protease inhibitors and phosphatase inhibitors, and the lysates were centrifuged and the

supernatants were collected. The concentration of the protein in the supernatants was determined by Bradford method. Proteins were then labeled with Cy5 and hybridized to KAM-1.2PN chip. This service was provided by Kinexus. We combined data from the chips by adjusting the total intensities of each spot followed by analyzing the image using Cluster3.0 and TreeView softwares.

Cohort data analysis for breast cancer patinets

From GEO and ArrayExpress data, 10 breast cancer microarray cohort data that contain the information of patient survival statuses were chosen (Table S1). Prognosis value of each marker was determined for overall or metastasis (relapse)-free survival statuses by comparing patient groups with three different thresholds (upper 25% vs lower 75%, upper 50% vs lower 50% or upper 75% vs lower 25%). Prognosis value of a combination of various genes as a signature in breast cancer microarray cohorts was also determined as follows (6). Briefly, a Pearson correlation coefficient for the correlation between the expression data for each patient and the average expression of the gene signature (Table S5) was calculated. Patients were grouped according to the correlation values, with 0 or the average correlation coefficient used as the threshold and the prognosis value of the patients was evaluated.

Real-Time PCR and Western blotting

qRT-PCR was performed using DNA Engine Opticon3 (MJ Research) and the SYBR Green qPCR Kit (Fermentas). Primers used in this study were summarized in Table S4. Western blotting was carried out using the HAS2 antibody (Santacruz) as previously described (7).

Reporter assay and chromatin immunoprecipitation

For reporter assay, HAS2 promoter region (1000 bp upstream from the transcriptional start site) was inserted into the pGL4 firefly luciferase reporter plasmid (Promega). The reporter plasmid was co-transfected with phRG-TK *Renilla* luciferase as an internal control (Promega) into MB231 cells using FuGENE HD reagent (Roche). Luciferase activities were then measured by using Dual-luciferase reporter assay system (Promega) and luminometer (Berthold Detection Systems). Chromatin immunoprecipitation was performed with MB231 cells as described previously (8) using the primers listed in Table S4.

Particle exclusion assay

Pericellular HA matrices were visualized by adding glutaraldehyde-fixed sheep erythrocytes (Sigma) into the cell culture of CSC. Briefly, isolated CSCs were seeded in 24 well plates in RPMI-1640 with 2% serum and incubated for 12 h. The medium was removed and 10^8 red blood cells were added followed by taking photo images under phase-contrast microscopy.

Cell adhesion assay and transmigration assay

mBMEC cells were grown to confluency in 24-well plates and the monolayers were washed with PBS. Isolated CSCs were resuspended in medium containing 10% FBS, and 10^5 cells were plated in each well of the mBMEC monolyer. They were allowed to adhere to the monolayer for 30 min. Plates were then washed three times and cells were lysed with 50 µl Passive lysis buffer (Promega). Firefly luciferase activity was measured using a luminometer. Isolated CSCs were also labeled with Cell tracker green (Invitrogen) for 20 min. The cells were plated in each well of the mBMEC culture and they were allowed to adhere to the monolayer for 30 min. Cells were then washed three times and the cell number was counted under fluorescent microscope. For transmigration assay, 500 cells of mBMEC were seeded into a trans-well insert, and they were allowed to grow to confluence for 2 days. Isolated CSCs were labeled with Cell tracker green for 20 min and 10^5 cells were

seeded into the trans-well insert containing RPMI-1640 medium with 2% serum. The bottom chamber of trans-wells was also filled with RPMI-1640 with 10% serum. After 48 hours, labeled cells that were transmigrated through the endothelial cells were counted under the fluorescent microscope.

In vivo tumor cell survival assay

CSCs were labeled with Cell tracker green and 10^6 cells were intravenously injected through the tail vein into nu/nu mice. After 48 hrs, rhodamine-lectin (100 mg/kg) was injected through i.v. Ten minutes later, mice were sacrificed and the lungs were removed and fixed with 10% formalin. The tissue was embedded in OTC buffer (Sakura Finetek), solidified at -80°C and sectioned (15 μm) at -20°C. The fluorescent image was visualized by confocal microscopy and the z-staked pictures were combined and cell numbers were counted.

Preparation of tumor associated-macrophage

Primary human monocytes were purchased from Astarte Biologics and human monocyte cell line THP1 was obtained from ATCC. To generate conditioned medium of 231BoM cells, cells were cultured in T75 flasks and then medium was changed to RPMI-1640 or AIM-V without serum when the cells reached 80% confluency. After 24hrs incubation, supernatant was collected and filtered. To induce differentiation of monocytes to TAM, primary monocytes were cultured for 5-7 days in AIM-V medium with 25% of conditioned medium of 231BoM, 10% FBS, 20 ng/ml IL-13 (Miltenyi Biotec) and 20 ng/ml IL-4 (Miltenyi Biotec). TAMs were also generated from THP1 cells that were cultured for 2 days in RPMI medium with 50% of conditioned medium of 231BoM, 320 ng/ml PMA (Sigma), 10% FBS, 20 ng/ml IL-13 and 20 ng/ml IL-4. The differentiation of monocytes to TAMs was confirmed by FACS using antibodies to CD163 and CD206 (eBiosciences).

Conditioned medium experiment

Two hundred thousand cells of THP1 were inoculated in a 12-well plate with RPMI-1640 medium containing 50% of conditioned medium of 231BoM, 320 ng/ml PMA, 10% FBS, 20 ng/ml IL-13, and 20 ng/ml IL-4. After 48 hrs, medium was changed to RPMI-1640 with 10% FBS, and 10^5 cells of CSC prepared from 231BoM or 231BoM/shHAS2 were added on top of the THP1-TAM monolayer and they were incubated 24 hrs. Medium was then changed to RPMI-1640 without serum and the plate was further incubated for 24 hours, and the supernatants were collected. These medium were designated as primary conditioned medium. To generate the secondary conditioned medium, 2 x 10^5 cells of stromal cell lines (HS5, HS27A, hFOB1.19 and BM-hMSC) were cultured in 12-well plate in RPMI-1640 with 10% FBS for 24 hours. The medium was then changed to the primary conditioned media or RPMI-1640 with PDGF-BB (100 ng/ml), and they were further incubated for 24 hours. After the incubation, medium were changed to RPMI-1640 without serum, and the cells were further incubated for 24 hours. The medium were then collected, filtered and designated as the secondary conditioned medium. Two-thousand cells of CSC were seeded in a 96-well plate with the secondary conditioned medium with 1% FBS or RPMI-1640 with FGF7 or FGF9 (100 ng/ml) (RayBiotech) for 72 hours. After the incubation, cell proliferation was measured by the MTS dye method (Promega). The rate of cell proliferation was normalized by the result of control group containing CSC in RPMI-1640 medium with 1% FBS. The growth of CSC population was measured by treating CSCs with the secondary conditioned medium, FGF7 or FGF9 followed by FACS analysis. One-hundred thousand cells of 231BoM, CN34-BoM2d and MCF7-BoM2d were seeded in 24-well plate in the secondary conditioned medium with 10% FBS for 6 days. After the incubation, stem cell population was measured by FACS using CD24-FTIC (eBiosciences), CD44-APC (Biolegend) and ESA-PE (eBiosciences) antibodies. For sphere forming assay, 500 cells of MCF7-BoM2d were seeded in a 96-well ultralow plate (Corning) in the secondary conditioned medium with 1% FBS for 8 days. After the incubation, the number of spheres was counted under a microscope.

Cytokine and growth factor antibody array

Primary and secondary conditioned medium were collected as described above. Analysis of the expressed cytokines and growth factors was performed using Human cytokine and growth factor antibody arrays (AAH-CYT-5 and AAH-GF-1, RayBiotech) by following the manufacturer's protocols. The luminescence signal intensity was quantified and normalized using ImageJ software.

Immunocytochemistry

Four-hundred-thousand cells of THP1 were seeded on a cover glass which was incubated in RPMI-1640 with 50% of conditioned medium of 231BoM, 320 ng/ml PMA, 10% FBS, 20 ng/ml IL-13 and 20 ng/ml IL-4 for 2 days. Medium was then changed to RPMI-1640 with 1% FBS, and 4 x 10^5 cells of CSCs from 231BoM or 231BoM/shHAS2 were added to the THP1-TAM monolayer followed by further incubation for 48 hrs. The medium were removed and cells were fixed with cold 70% ethanol. Cells were then incubated in PBS with 2% BSA for 2 hours followed by incubation with primary antibodies for CD68 (GeneTex) and PDGF-BB (Santacruz). After washing with PBS, the secondary antibodies (anti-mouse-alexa568 and anti-rabbit-alexa488, Invitrogen) were added, and cells were then treated with ProLong antifade reagent with DAPI (Invitrogen). The fluorescent images were taken by confocal microscopy.

Animal Experiments

For orthotopic tumor growth, CSCs carrying the luciferase gene were mixed with Matrigel (BD Biosciences) at 1:1 ratio in 100 µl volume and they were injected into the fourth mammary gland of nu/nu mice (3-4 weeks). Tumor growth was then monitored using Xenogen bioimager (Caliper) by measuring photon flux after 15 minutes of intraperitoneal injection of 150 mg/kg of luciferin. For experimental metastasis assay, nu/nu mice (7-8 weeks) were anesthetized by intraperitoneal injection of 120 mg/kg ketamine and 6 mg/kg xylazine. Five hundred thousand cells of CSCs in 100 µl of PBS were then injected into the left cardiac ventricle of the mice. To confirm a successful injection, we immediately monitored photon flux from whole body of the mice. The metastatic growth of tumors in the bone was monitored and quantified by measuring luminescence at tibia using Xenogen bioimager. To examine the effect of bone microenvironment on tumor growth, 1 x 10^4 cells of CSC with or without 2 x 10^4 cells of TAM were co-injected directly into tibial bone and the mice were monitored for 8 weeks. Osteolytic lesions were visualized by X-ray radiography by exposing to an X-ray at 35 kV for 10 sec using Faxitron Instrument Model MX-20 (Faxitron Corp.). For 4-MU treatment, mice were gavaged once a day with 4-MU (400 mg/kg) which was suspended in water with 1% arabic gum. The treatment began one day before injection of tumor cells.

Immunohistochemistry

Immunohistochemical analysis was carried out for paraffin-embedded, surgically resected specimens of breast cancer using polyclonal antibody to HAS2 (Santacruz). Briefly, the sections were deparaffinized, rehydrated and heated at 100 °C for 20 min in 25 mM sodium citrate buffer (pH 6.0) for antigen exposure. They were treated with 3% H_2O_2 to block endogenous peroxidase activity and then incubated with primary antibody (1:200 dilution) for 16 h at 4 °C. After washing in PBS/0.1% Tween-20, the sections were treated with horseradish peroxidase-conjugated rabbit-specific IgG (Dako Corp.). The sections were washed extensively, and DAB substrate chromogen solution was applied followed by counterstaining with hematoxylin. Results of the immunohistochemistry for HAS2 were judged based on the intensity of staining, and the grading was done by two independent persons without prior knowledge of the grade, stage, or patient survival information. The Southern Illinois University Institutional Review Board approved obtaining human specimens for this study.

Supplemental References

1. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An emryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 2008;40:499-507.
2. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 2003;3:537-49.
3. Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. Nature 2009;459:1005-9.
4. Kim MY, Oskarsson T, Acharyya S, Nguyen DX, Zhang XH, Norton L, et al. Tumor self-seeding by circulating cancer cells. Cell 2009;139:1315-26.
5. Langley RR, Ramirez KM, Tsan RZ, Van Arsdall M, Nilsson MB, Fidler IJ. Tissue-specific microvascular endothelial cell lines from H-2K(b)-tsA58 mice for studies of angiogenesis and metastasis. Cancer Res 2003;63:2971-6.
6. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217-26.
7. Iiizumi M, Bandyopadhyay S, Pai SK, Watabe M, Hirota S, Hosobe S, et al. RhoC promotes metastasis via activation of the Pyk2 pathway in prostate cancer. Cancer Res 2008:68:7613-20.
8. Furuta E, Pai SK, Zhan R, Bandyopadhyay S, Watabe M, Mo YY, et al. Fatty acid synthase gene is up-regulated by hypoxia via activation of Akt and sterol regulatory element binding protein-1. Cancer Res 2008;68:1003-11.
9. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529-41.
10. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.
11. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci USA 2005;102:13550-5.
12. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 2007;13:3207-14.
13. Pawitan Y, Bjöhle J, Amler L, Borg AL, Egyhazi S, Hall P, et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 2005;7:R953-64.
14. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006;439:353-7.
15. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-72.
16. 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;25:1239-46.
17. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.
18. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature 2005;436:518-24.

Supplemental Figures

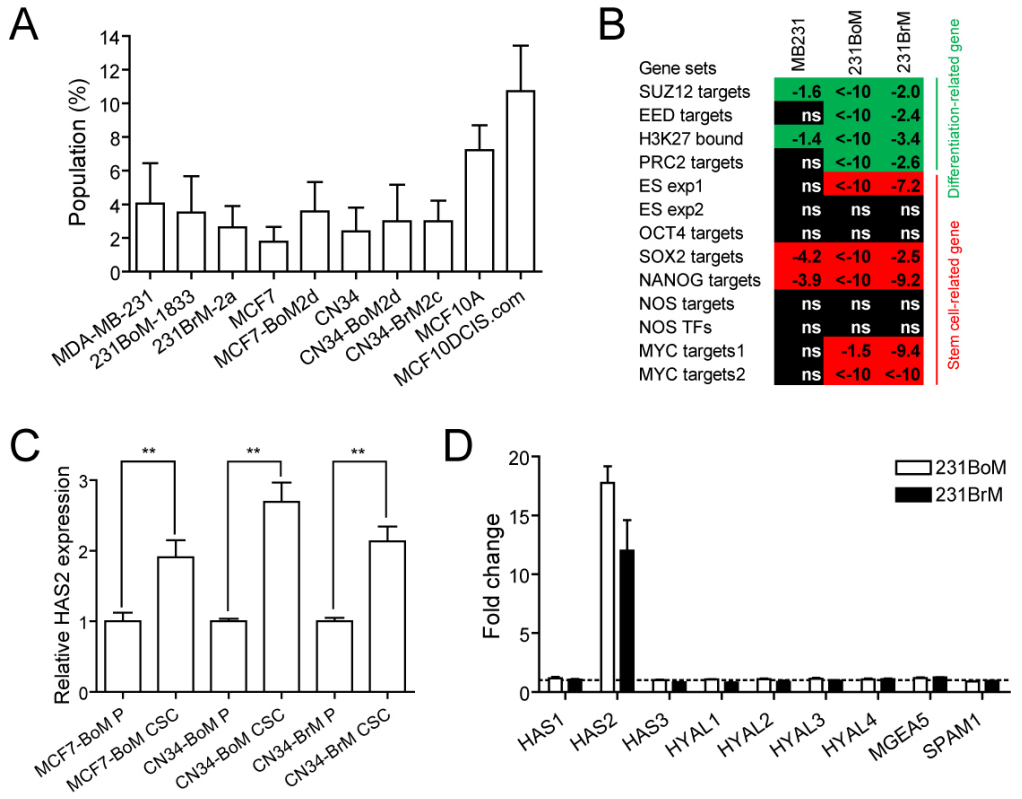

Figure S1 Cancer stem-like cell populations of breast cancer cell lines

(A) CSCs from breast cancer cell lines were isolated by MACS using CD24, CD44 and EpCAM antibodies. The percentages of isolated cell populations were plotted. Data are represented as mean ± SD (n=4-26). (B) Expression profile analysis was done for CSCs isolated from MB231, 231BoM and 231BrM using Affymetrix human gene 1.0ST chip, and gene-sets enrichment analysis was performed for the genes related to stem cell and differentiation that were expressed in CSCs or unsorted parental cells. Columns represent individual samples and sample names are indicated above. Rows represented individual gene set (5). Red, gene-set enrichment for overexpression; green, gene-set enrichment for underexpression; black, no significant enrichment. Numbers indicated log value of the P values for gene-set enrichment significance by comparing the stem cells with the unsorted parental cells. (C) *HAS2* gene expression in metastatic cell lines was measured by qRT-PCR. P; Parental cells, CSC; stem cell populations of the cell lines. Data are represented as mean ± SEM (n=3). ** indicates P<0.001. (D) Fold changes of expression of hyaluronan synthases and hyaluronidases in CSCs. The expression level of CSCs from MB231 was set as 1. Data are represented as mean ± SEM (n=2).

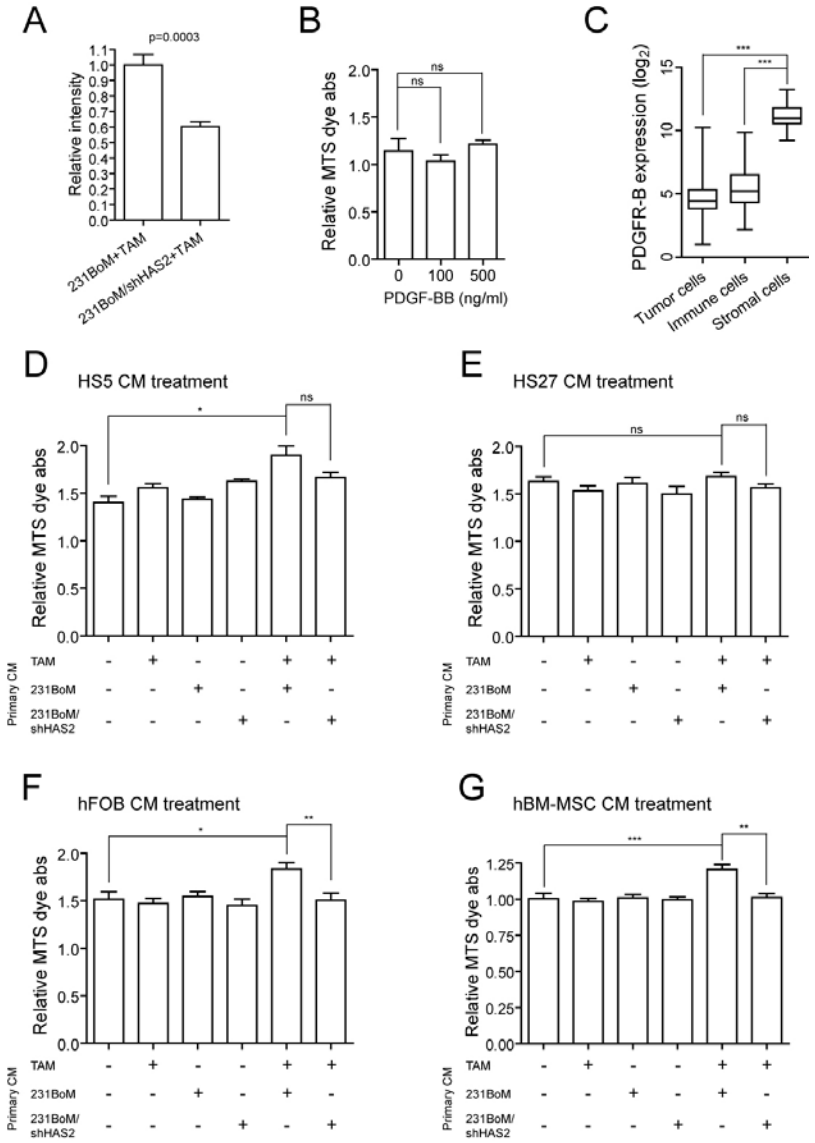

Figure S2 TAMs derived PDGF-BB stimulates stromal cells

(A) Relative intensity of PDGF-BB expression in immunofluorescence data in Fig. 4D. Six individual pictures were taken and intensity of PDGF-BB expression was measured by using ImageJ software. Data are represented as mean ± SEM (n=6). (B) CSCs isolated from 231BoM were treated with the indicated dose of PDGF-BB followed by measuring their growth and survival by MTS assay. Data are represented as mean ± SEM (n=3). (C) Expression levels of PDGFR-B (PDGF receptor beta) in tumor, immune and stromal cells were evaluated. The expression data of PDGFR-B were obtained from the GEO site, and they were normalized and plotted. "Tumor cells" indicates breast cancer cell lines. "Immune cells" indicates various immune cells from human blood. "Stromal cells" indicates compilation of data for fibroblasts, osteoblasts and mescenchymal stem cells. P-value is calculated by Kruskal-Wallis test. *** indicates P<0.0001. (D-E) The effect of secondary conditioned medium on the growth of CSCs was measured by the MTS assay. The primary conditioned medium (Primary CM) was first prepared by co-culturing TAMs and CSCs isolated from 231BoM or 231BoM carrying shRNA-HAS2. To prepare the secondary conditioned medium, HS5, HS27A, hFOB1.19 or hBM-MSC was treated with each primary CM for 24 hrs. Cells were washed and further incubated with serum free medium for another 24 hrs, followed by collecting the secondary CM. CSCs were prepared from 231BoM and they were treated with the secondary CM prepared from HS5 (D), HS27A (E), hFOB1.19 (F) or hBM-MSC (G) with 1% FBS followed by MTS assay (n=3). *** indicates P<0.0001, ** indicates P<0.001, * indicates P<0.05, and ns indicates not significant. Data are represented as mean ± SEM.

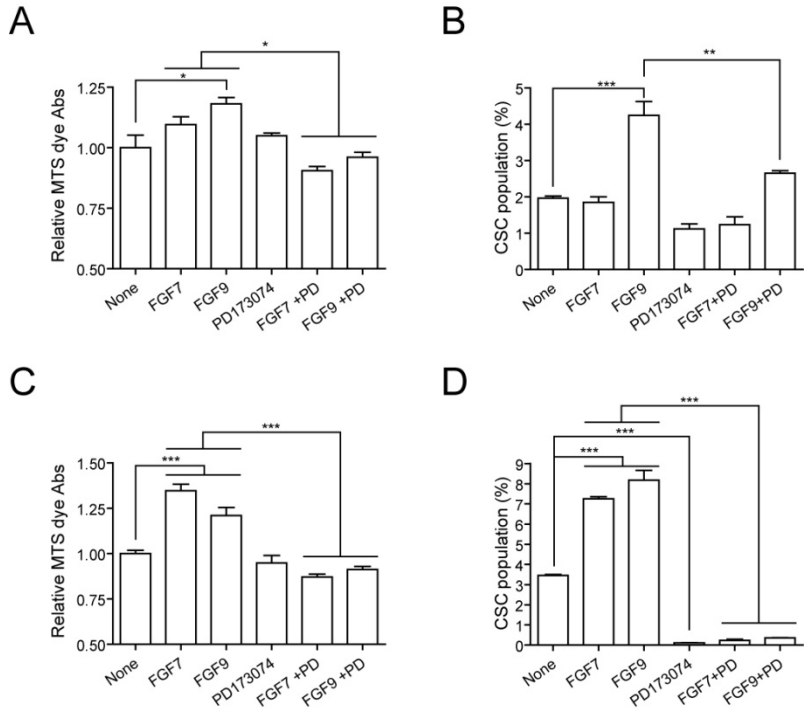

Figure S3 Effect of FGF7 and FGF9 on the cell proliferation and CSC population

(A) The effect of FGF9 and FGF7 on the growth of CSCs from CN34-BoM2d cells was measured by MTS assay. CSCs were prepared from CN34-BoM2d cells and they were treated with the combination of FGFs and PD173074 with 1% FBS followed by MTS assay (n=3). (B) The effect of FGF9 and FGF7 on the CSC population in CN34-BoM2d cells was measured by FACS. Parental CN34-BoM2d cells were treated by the combination of FGFs and PD173074 in culture medium with 10% FBS for 6 days (n=3). (C) The effect of FGF9 and FGF7 on the growth of CSCs from MCF7-BoM2d cells was measured by MTS assay (n=3). (D) The effect of FGF9 and FGF7 on the CSC population in MCF7-BoM2d cells was measured by FACS (n=3). *** indicates $P<0.0001$, ** indicates $P<0.001$, * indicates $P<0.05$. Data are represented as mean ± SEM.

Table S1. Patient cohort data used in this study

Cohort name	GEO ID	Microarray chip	Reference
Chin	NA*	Affymetrix human genome U133A	(9)
VV (van de Vijver)	NA	NKI-AVL homo sapiens 18K cDNA array	(10)
Miller	GSE3494	Affymetrix human genome U133A&B	(11)
Desmedt	GSE7390	Affymetrix human genome U133A	(12)
Pawitan	GSE1456	Affymetrix human genome U133A&B	(13)
Bild	GSE3143	Affymetrix human genome U95v2	(14)
Sotiriou	GSE2990	Affymetrix human genome U133A	(15)
Loi	GSE6532	Affymetrix human genome U133A&B&U133 plus 2.0	(16)
Wang	GSE2034	Affymetrix human genome U133A	(17)
Minn	GSE2603	Affymetrix human genome U133A	(18)

* ArrayExpress ID; E-TABM-158

Table S2. Survival analysis for the data represented in Figure 1B

Thirteen genes that were up- or down-regulated in CSCs from metastatic variants were colored as red and green in their background, respectively. P-values (P<0.05) of poor or good prognosis value were also colored as red and green letters, respectively.

Over-all survival

GENE	Prognosis		Miller cohort				VV cohort				Desmedt cohort				Chin cohort			
NAME	Good	Poor	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis
MMP1	0	5	0.685	0.013	0.007	Poor	0.156	0.034	0.082	Poor	0.150	0.005	0.002	Poor	0.040	0.178	0.036	Poor
SERPINB2	1	1	0.742	0.761	0.659	None	0.998	0.913	0.354	None	0.032	0.037	0.695	Poor	0.450	0.285	0.533	None
EHF	0	0	0.275	0.786	0.095	None	0.744	0.405	0.238	None	0.880	0.465	0.778	None	0.272	0.759	0.441	None
CHRDL1	1	0	0.325	0.331	0.269	None	NA	NA	NA	NA	0.245	0.492	0.264	None	0.487	0.910	0.127	None
SPANXB1	1	2	0.009	0.236	0.396	Good	NA	NA	NA	NA	0.036	0.012	0.014	Poor	0.255	0.468	0.499	None
HAS2	0	2	0.178	0.063	0.219	None	0.734	0.879	0.186	None	0.310	0.782	0.167	None	0.224	0.829	0.027	Poor
CCBE1	0	0	0.697	0.779	0.254	None	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ANXA8L1	0	0	0.907	0.906	0.861	None	NA	NA	NA	NA	0.894	0.821	0.755	None	0.986	0.850	0.542	None
CRISPLD2	1	0	0.040	0.007	0.247	Good	NA	NA	NA	NA	0.754	0.226	0.783	None	0.757	0.173	0.522	None
UCA1	0	0	0.530	0.734	0.189	None	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ODZ2	2	0	0.682	0.360	0.084	None	0.850	0.001	0.007	Good	NA	NA	NA	NA	NA	NA	NA	NA
LCP1	0	1	0.159	0.844	0.551	None	0.457	0.629	0.027	Poor	0.184	0.755	0.497	None	0.873	0.188	0.326	None
MPZL2	0	1	0.297	0.962	0.288	None	NA	NA	NA	NA	0.394	0.200	0.064	None	0.167	0.565	0.193	None

GENE	Prognosis		Pawitan cohort				Bild cohort				Sotiriou cohort				Loi cohort			
NAME	Good	Poor	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis
MMP1	0	5	0.039	0.007	0.004	Poor	0.441	0.180	0.606	None	0.920	0.323	0.335	None	0.995	0.193	0.571	None
SERPINB2	1	1	0.549	0.871	0.454	None	0.457	0.702	0.809	None	0.668	0.318	0.006	Good	0.252	0.529	0.186	None
EHF	0	0	0.095	0.364	0.683	None	NA	NA	NA	NA	0.101	0.756	0.424	None	0.153	0.754	0.786	None
CHRDL1	1	0	0.070	0.001	0.004	Good	0.226	0.412	0.866	None	0.392	0.309	0.424	None	0.800	0.097	0.378	None
SPANXB1	1	2	0.348	0.374	0.040	Poor	NA	NA	NA	NA	0.323	0.488	0.472	None	0.294	0.681	0.988	None
HAS2	0	2	0.209	0.785	0.485	None	0.671	0.040	0.017	Poor	0.057	0.519	0.172	None	0.296	0.327	0.741	None
CCBE1	0	0	0.860	0.943	0.495	None	NA	NA	NA	NA	NA	NA	NA	NA	0.716	0.478	0.955	None
ANXA8L1	0	0	0.603	0.357	0.901	None	0.353	0.317	0.623	None	0.231	0.789	0.193	None	0.205	0.238	0.077	None
CRISPLD2	1	0	0.090	0.156	0.107	None	NA	NA	NA	NA	0.545	0.540	0.316	None	0.293	0.951	0.497	None
UCA1	0	0	0.052	0.357	0.693	None	NA	NA	NA	NA	NA	NA	NA	NA	0.887	0.734	0.359	None
ODZ2	2	0	0.012	0.434	0.711	Good	NA	NA	NA	NA	NA	NA	NA	NA	0.814	0.865	0.760	None
LCP1	0	1	0.282	0.368	0.840	None	0.675	0.249	0.594	None	0.304	0.824	0.202	None	0.853	0.297	0.301	None
MPZL2	0	1	0.734	0.872	0.011	Poor	0.940	0.782	0.773	None	0.180	0.360	0.382	None	0.137	0.556	0.156	None

Metastasis-free survival

GENE	Prognosis		Wang cohort				VV cohort				Desmedt cohort				Chin cohort			
NAME	Good	Poor	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis
MMP1	0	5	0.005	0.020	0.088	Poor	0.219	0.095	0.291	None	0.023	0.000	0.001	Poor	0.023	0.011	0.008	Poor
SERPINB2	1	1	0.476	0.315	0.115	None	0.035	0.070	0.369	Good	0.049	0.090	0.693	Poor	0.646	0.245	0.584	none
EHF	2	0	0.111	0.005	0.315	Good	0.988	0.669	0.794	None	0.226	0.993	0.565	None	0.968	0.158	0.105	None
CHRDL1	1	0	0.500	0.590	0.297	None	NA	NA	NA	NA	0.141	0.579	0.114	None	0.930	0.361	0.074	None
SPANXB1	1	3	0.506	0.677	0.950	None	NA	NA	NA	NA	0.041	0.006	0.002	Poor	0.989	0.418	0.831	None
HAS2	0	3	0.037	0.087	0.017	Poor	0.224	0.226	0.125	None	0.5534	0.7146	0.8833	None	0.209	0.221	0.232	None
CCBE1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ANXA8L1	0	0	0.082	0.851	0.872	None	NA	NA	NA	NA	0.733	0.873	0.309	None	0.565	0.250	0.937	None
CRISPLD2	1	0	0.561	0.147	0.797	None	NA	NA	NA	NA	0.674	0.523	0.756	None	0.264	0.654	0.838	None
UCA1	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ODZ2	1	0	NA	NA	NA	NA	0.688	0.002	0.022	Good	NA	NA	NA	NA	NA	NA	NA	NA
LCP1	1	0	0.594	0.055	0.049	Good	0.400	0.653	0.349	None	0.367	0.710	0.199	None	0.431	0.344	0.412	None
MPZL2	0	1	0.900	0.994	0.889	None	NA	NA	NA	NA	0.585	0.324	0.120	None	0.822	0.967	0.376	None

GENE	Prognosis		Pawitan cohort				Minn cohort				Sotiriou cohort				Loi cohort			
NAME	Good	Poor	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis	25%vs75%	50%vs50%	75%vs25%	Prognosis
MMP1	0	5	0.034	0.004	0.002	Poor	0.447	0.948	0.500	None	0.799	0.930	0.718	None	0.523	0.033	0.191	Poor
SERPINB2	1	1	0.972	0.841	0.604	None	0.830	0.203	0.960	None	0.779	0.716	0.066	None	0.226	0.059	0.064	None
EHF	2	0	0.005	0.086	0.711	Good	0.222	0.063	0.252	None	0.325	0.875	0.974	None	0.561	0.177	0.944	None
CHRDL1	1	0	0.339	0.003	0.005	Good	0.556	0.856	0.857	None	0.857	0.137	0.570	None	0.927	0.765	0.127	None
SPANXB1	1	3	0.246	0.385	0.019	Poor	0.049	0.403	0.599	Good	0.492	0.564	0.415	None	0.214	0.055	0.017	Poor
HAS2	0	3	0.075	0.234	0.108	None	0.414	0.341	0.025	Poor	0.298	0.985	0.084	None	0.029	0.258	0.992	Poor
CCBE1	0	0	0.609	0.760	0.097	None	NA	NA	NA	NA	NA	NA	NA	NA	0.509	0.781	0.525	None
ANXA8L1	0	0	0.731	0.586	0.633	None	0.263	0.534	0.373	None	0.315	0.471	0.724	None	0.462	0.301	0.240	None
CRISPLD2	1	0	0.033	0.029	0.046	Good	0.993	0.649	0.492	None	0.302	0.221	0.123	Good	0.898	0.370	0.819	None
UCA1	0	0	0.709	0.768	0.972	None	NA	NA	NA	NA	NA	NA	NA	NA	0.914	0.965	0.323	None
ODZ2	1	0	0.235	0.725	0.947	None	NA	NA	NA	NA	NA	NA	NA	NA	0.500	0.361	0.598	None
LCP1	1	0	0.175	0.145	0.461	None	0.488	0.199	0.067	None	0.512	0.629	0.088	None	0.698	0.938	0.251	None
MPZL2	0	1	0.434	0.791	0.043	Poor	0.798	0.599	0.068	None	0.633	0.552	0.933	None	0.835	0.063	0.429	None

Table S3. Phosphoproteins that were up-regulated in CSCs from 231BoM and 231BrM

Eighty-nine phosphoproteins that were up-regulated in both stem cells from 231BoM and 231BrM at least 1.5-fold higher or lower than that from MB231 were listed.

Antibody codes	231BoM	231BrM	Target protein name	Phospho sites	Full target protein name
PK105	5.65	4.41	S6Ka (p70/p85 S6Ka)	T421+S424	p70/p85 ribosomal protein-serine S6 kinase alpha
PN058	4.02	3.97	Pax2	S394	Paired box protein 2
PK007-2	4.96	1.89	CDK1/2	Y15	Cyclin-dependent protein-serine kinase 1/2
PK002	3.94	2.85	AMPKa1/2	T174/T172	5'-AMP-activated protein kinase subunit alpha 1/2
PK060-2	3.47	2.94	p38a MAPK	T180+Y182	Mitogen-activated protein-serine kinase p38 alpha
PN075	2.98	3.13	Smad1/5/9	S463+S465/S463+S465/S465+S467	SMA- and mothers against decapentaplegic homologs 1/5/9
PN080	2.77	2.11	STAT2	Y690	Signal transducer and activator of transcription 2
PK004	2.27	2.51	Btk	Y223	Bruton's agammaglobulinemia tyrosine kinase
PN040-1	1.94	2.66	Hsp27	S15	Heat shock 27 kDa protein beta 1 (HspB1)
PN001	2.79	1.55	4E-BP1	S65	Eukaryotic translation initiation factor 4E binding protein 1 (PHAS1)
PK052	1.77	2.55	MEK4 (MAP2K4)	S257+T261	MAPK/ERK protein-serine kinase 4 (MKK4)
PN097	1.71	2.49	eNos	T495	Nitric-oxide synthase, endothelial
PK073	1.87	2.29	PKCa	S657	Protein-serine kinase C alpha
PK024	1.77	2.27	FAK	S910	Focal adhesion protein-tyrosine kinase
PN020	1.80	2.13	Cofilin 2	S3	Cofilin 2
PN003-PN004	1.61	2.24	Adducin a	S726	Adducin alpha (ADD1)
PK093-1	2.31	1.52	PKCm (PKD)	S910	Protein-serine kinase C mu (Protein kinase D)
PN036	1.97	1.61	Histone H2A.X	S139	Histone H2A variant X
PN059	1.85	1.53	Paxillin 1	Y31	Paxillin 1
PN052	1.66	1.72	MYPT1	T696	Myosin phosphatase target 1
PK018-1	-1.80	-1.66	FAK	Y576	Focal adhesion protein-tyrosine kinase
PN044	-1.51	-1.98	Integrin b1	S785	Integrin beta 1 (fibronectin receptor beta subunit, CD29 antigen)
PN014	-2.11	-1.57	BRCA1	S1497	Breast cancer type 1 susceptibility

					protein
PN121	-1.59	-2.09	Tau	T521	Microtubule-associated protein tau
PN076	-1.51	-2.18	Smad2	S465+S467	SMA- and mothers against decapentaplegic homolog 2
PN079	-2.13	-1.73	STAT1	Y701	Signal transducer and activator of transcription 1
PK028-PK029-1	-2.17	-2.10	GSK3a	Y279/ Y216	Glycogen synthase-serine kinase 3 alpha
PK084	-2.29	-2.03	PKCg	T674	Protein-serine kinase C gamma
PK128	-2.07	-2.41	MEK4 (MAP2K4)	S80	MAPK/ERK protein-serine kinase 4 (MKK4)
PN100	-3.34	-1.61	Histone H3	T11	Histone H3.3
PK086	-2.26	-3.42	PKCh	S674	Protein-serine kinase C eta
PK083	-1.74	-3.97	PKCg	T655	Protein-serine kinase C gamma
PK060-4	-2.17	-3.64	p38a MAPK	T180+Y182	Mitogen-activated protein-serine kinase p38 alpha
PN057-3	-2.84	-2.97	p53	S392	Tumor suppressor protein p53 (antigenNY-CO-13)
PP004	-3.21	-2.77	SHP2	S576	Protein-tyrosine phosphatase 1D (SHPTP2, Syp, PTP2C)
PK079	-3.58	-3.13	PKCd	S645	Protein-serine kinase C delta
PK104	-6.20	-2.52	S6Ka (p70/p85 S6Ka)	T229	p70/p85 ribosomal protein-serine S6 kinase alpha
PK106	-7.24	-1.78	S6Ka (p70/p85 S6Ka)	T389	p70/p85 ribosomal protein-serine S6 kinase alpha

Table S4. Primers in this study

Name	Sequence
ACTB F	TGAGACCTTCAACACCCCAGCCATG
ACTB R	CGTAGATGGGCACAGTGTGGGTG
HAS2 F	GCCTCATCTGTGGAGATGGT
HAS2 R	TCCCAGAGGTCCACTAATGC
MMP1 F	GAGCTCAACTTCCGGGTAGA
MMP1 R	CCCAAAAGCGTGTGACAGTA
CRISPLD2 F	CCATGTGCAGTCCTGGTATG
CRISPLD2 R	TCCCCAGACAGTCATCTTCC
ODZ2 F	TACTGCAGCTGGAAATGTGC
ODZ2 R	CCTGGTGGGACTTCTTGTGT
PDGFB F	AAGCACACGCATGACAAGAC
PDGFB R	GACGGACGAGGGAAACAATA
HAS2 ChIP F1	GGGACCAGATTGAGCAAAAA
HAS2 ChIP R1	TCTAAAAAGATCGCGCTGGT
HAS2 ChIP F2	TCAGTCATCAGCAGGCTTGT
HAS2 ChIP R2	TGGCTTAACTGGCTTTTGCT
HAS2 ChIP F0	ATGTTTTGTGCTGTGGGATG
HAS2 ChIP R0	CCATCCCTTTCAGGTTTTGA

Table S5. A set of the signature genes used for the analysis in Figure 6D

Up-regulated genes		Down-regulated genes	
MMP1	NSBP1	GMFG	COL4A1
EHF	PTPN22	ITGB3	ANXA8
SERPINB2	TLR2	TGFBI	ANXA8L2
IL24	ZNF521	ARHGAP28	EFEMP1
ESM1	PLAT	SLC46A3	SCNN1A
CHRDL1	SPANXC	LOC349196	SLPI
HAS2	MEF2C	GPR65	UCA1
RAB38	SPANXA2	LOC100130876	LCP1
ROBO1		PON2	ANXA8L1
DOCK10		PPME1	CRISPLD2
SPANXB1		TAGLN	ODZ2
SEMA3A		OLR1	MPZL2
CCBE1		SCNN1G	
TRY6		IFIT1	
ARMCX2		PDK4	