MolecularMedicine

Functional Heterogeneity within the CD44 High Human Breast Cancer Stem Cell–Like Compartment Reveals a Gene Signature Predictive of Distant Metastasis

Rikke Leth-Larsen,[1] Mikkel G Terp,[1] Anne G Christensen,[1] Daniel Elias,[1] Thorsten Kühlwein,[1] Ole N Jensen,[2] Ole W Petersen,[3] and Henrik J Ditzel[1,4]

[1]Department of Cancer and Inflammation Research, Institute of Molecular Medicine, and [2]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark; [3]Department of Cellular and Molecular Medicine, Danish Stem Cell Centre, and Centre for Biological Disease Analysis, University of Copenhagen, Copenhagen, Denmark; and [4]Department of Oncology, Odense University Hospital, Odense, Denmark

The $CD44^{hi}$ compartment in human breast cancer is enriched in tumor-initiating cells; however, the functional heterogeneity within this subpopulation remains poorly defined. We used a triple-negative breast cancer cell line with a known bilineage phenotype to isolate and clone $CD44^{hi}$ single cells that exhibited mesenchymal/basal B and luminal/basal A features, respectively. Herein, we demonstrate in this and other triple-negative breast cancer cell lines that, rather than $CD44^{hi}/CD24^{-}$ mesenchymal-like basal B cells, the $CD44^{hi}/CD24^{lo}$ epithelioid basal A cells retained classic cancer stem cell features, such as tumor-initiating capacity *in vivo*, mammosphere formation and resistance to standard chemotherapy. These results complement previous findings using oncogene-transformed normal mammary cells showing that only cell clones with a mesenchymal phenotype exhibit cancer stem cell features. Further, we performed comparative quantitative proteomic and gene array analyses of these cells and identified potential novel markers of breast cancer cells with tumor-initiating features, such as lipolysis-stimulated lipoprotein receptor (LSR), RAB25, S100A14 and mucin 1 (MUC1), as well as a novel 31-gene signature capable of predicting distant metastasis in cohorts of estrogen receptor–negative human breast cancers. These findings strongly favor functional heterogeneity in the breast cancer cell compartment and hold promise for further refinements of prognostic marker profiling. Our work confirms that, in addition to cancer stem cells with mesenchymal-like morphology, those tumor-initiating cells with epithelial-like morphology should also be the focus of drug development.

Online address: http://www.molmed.org
doi: 10.2119/molmed.2012.00091

INTRODUCTION

Recent studies have suggested that only a small subpopulation of cells within solid tumors are capable of self-renewal and of generating tumors in mice that resemble the tumor of origin, which has led to a new model of tumorigenesis called the cancer stem cell (CSC) concept. CSCs give rise to the differentiated progeny that constitute the bulk of the tumor and thus contributes to the tumor's cellular heterogeneity (1–5). CSCs are thought to be responsible for metastatic potential and possess innate resistance mechanisms against chemotherapy- and radiation-induced cancer cell death, while the bulk of a tumor lacks these capacities (6–9). The resistance mechanisms may cause these cells to survive and become the source of later tumor recurrence (10–14), highlighting the need for therapeutic strategies that specifically target pathways central to these CSCs, such as Notch, Hedgehog and Wnt (15). However, such approaches are not straightforward, since they may harm the normal stem cell population and are challenged by the presence of plasticity between populations of cancer cells (16–18).

Advances have been made in identifying and enriching tumor-initiating cells with CSC features in several cancer types, including breast, brain and colon, but additional protein markers specific for, or associated with, this cell population are needed (9,19–21). In breast cancer, Al-Hajj *et al.* (9) demonstrated that cells derived from breast cancer metastases with a $CD44^{hi}/CD24^{-/lo}/lineage^{-}/$ epithelial cell adhesion $(ESA)^{+}$ pheno-

Address correspondence to *Henrik J Ditzel, Department of Cancer and Inflammation Research, Institute of Molecular Medicine, University of Southern Denmark, J.B. Winsloewsvej 25.3, 5000 Odense C, Denmark. Phone: +45-65503781; Fax: +45-65503922; E-mail: hditzel@health.sdu.dk.*
Submitted February 28, 2012; Accepted for publication June 7, 2012; Epub www.molmed.org) ahead of print June 8, 2012.

The Feinstein Institute for Medical Research North Shore LIJ

type were enriched for cells exhibiting CSC characteristics (9), and recent clinical studies showed that residual breast tumor cell populations, upon neoadjuvant chemotherapy, were enriched by $CD44^{hi}/CD24^{-/lo}$ defined breast CSCs and that these cells had increased expression of mesenchymal markers (22). More recently, a few other markers preferentially expressed on CSC-like cells were reported, including CXCR1 and aldehyde dehydrogenase (ALDH), as assessed by the ALDEFLOUR™ assay (23–25). $CD44^{hi}/CD24^{-/lo}$ cells are most common in basal-like subtypes of breast cancer and are very low in $HER2^{+}$ tumors (26).

It has been proposed that CSCs are linked to the epithelial-to-mesenchymal transition (EMT), wherein epithelial cells transform into a more invasive and motile mesenchymal-like phenotype, which drives tumor cell dissemination (27,28). Subsequently, the disseminated mesenchymal-like tumor cells must undergo the reverse transition, mesenchymal-to-epithelial transition (MET), at the site of metastases, which recapitulates the pathology of the corresponding primary tumors (29,30). Therefore, phenotypic plasticity between epithelial-like and mesenchymal-like cells in the formation of CSCs has major implications for therapeutic approaches (31). In this regard, a recent transcriptional analysis of basal-like breast cancer cell lines identified two major subclasses (basal A and basal B clusters), in which basal B appeared mesenchymal-like, whereas basal A may have either luminal-like or basal-like morphology (32).

To address these issues in more depth, we generated a panel of isogenic single-cell clones with either mesenchymal-like $CD44^{hi}/CD24^{-}$ or epithelial-like $CD44^{hi}/CD24^{lo}$ phenotypes from a triple-negative (estrogen receptor $[ER]^{-}$, progesterone receptor $[PR]^{-}$, $HER2^{-}$) cell line. This breast cancer subtype was chosen because these patients display high levels of $CD44^{hi}/CD24^{-/lo}$ cells and have a poor prognosis. No targeted therapy is currently available for this subtype, limiting medical treatment to chemotherapy. Somewhat surprisingly, our results showed that the cells that exhibited CSC-like characteristics, such as $CD44^{hi}/CD24^{-/lo}$ expression, the formation of mammospheres, initiation of tumors in mice and resistance to chemotherapy, displayed epithelial-like morphology and characteristics corresponding to the basal A subtype, indicating that this CSC-like population should also be targeted by novel treatments. Importantly, the tumors established from the epithelial-like single-cell clone with CSC-like features contained both cancer cells with epithelial-like, basal A characteristics, as well as cancer cells with mesenchymal-like, basal B characteristics, indicative of EMT in the process of tumor formation. Our transcriptomic and proteomic analyses provide a list of new biomarkers of the mesenchymal-like $CD44^{hi}/CD24^{-}$ subpopulations as well as epithelial-like $CD44^{hi}/CD24^{lo}$ subpopulations. The significance of selected biomarkers was confirmed in other cell lines of the basal A and basal B phenotypes, respectively. Finally, a 31-gene breast cancer signature capable of predicting recurrence of ER^{-} patients is presented, suggesting that tumors derived from patients with poor outcome contained more cells with CSC-like features. Interestingly, no prognostic signature for ER^{+} patients could be identified, suggesting that CSC-associated genes differ among subtypes of breast cancer.

MATERIALS AND METHODS

Cell Lines

HMT3909 and clones thereof originate from a biopsy of human primary infiltrating ductal breast carcinoma containing both medullary and intraductal components (33). HMT3909S13 and clones thereof, including A4 and G4, were derived from HMT3909S8 described by Petersen *et al.* (33). HMT3909S13-derived cell lines were cultured in PureCol-coated (Advanced BioMatrix Inc., San Diego, CA, USA) flasks in serum-free medium (mammary epithelial basal medium [MEBM]; Lonza Group Ltd, Basel, Switzerland) with supplements (MEGM™ Mammary Epithelial Cell Growth Medium): SingleQuots Kit (Clonetics™; Lonza), 10 ng/mL fibroblast growth factor (FGF), 20 ng/mL epidermal growth factor (EGF; PeproTech, Rocky Hill, NJ, USA), 2% B27 (Gibco; Life Technologies Corporation, Grand Island, NY, USA) and 4 μg/mL heparin (Sigma-Aldrich, St. Louis, MO, USA). MDA-MB-231, MDA-MB-468, T47D, MCF7 and Hs578T cell lines were obtained from ATCC and cultured in Dulbecco's modified Eagle medium, 10% fetal bovine serum, 1% penicillin/streptomycin and 1% nonessential amino acids (Gibco; Life Technologies Corporation). Cells were harvested with Accutase (EMD Millipore, Billerica, MA, USA). HMT3909S13 cells treated with chemotherapeutics were cultured as described above, but 24 h after seeding, these cells were subjected to paclitaxel, doxorubicin, methotrexate and salinomycin for 10 d. The medium and chemotherapeutics were replaced every third day. A4 and G4 were transfected with the luciferase gene Luc2 using Lentifire, a ready-to-use lentivirus-based delivery system (In Vivo Imaging Solutions, Fort Collins, CO, USA). Clones of HMT3909S13 were authenticated by G-band karyotyping (Supplementary Figure S1), and all cell lines were frequently tested for mycoplasma.

Antibodies

Flow cytometry included the following: phycoerythrin (PE)-conjugated anti-human CD44 (isotype control PE-conjugated IgG2b; BD Pharmingen; BD, Albertslund, Denmark), Alexa Fluor 647–conjugated anti-human CD24 (isotype control Alexa Fluor 647-conjugated IgG2a; Biolegend, San Diego, CA, USA), anti-human lipolysis-stimulated lipoprotein receptor (LSR) (Abcam, Cambridge, UK) and anti-human RAB25 (Abnova, Taipei City, Taiwan). Immunocytochemistry (ICC)/immunohistochemistry (IHC) included the following: anti-human CK5/6, CK7/8, CK17, CK18, CK19,

HER2, vimentin, mucin 1 (MUC1), ESA, PR (Dako, Glostrup, Denmark), CK14, EGF receptor (EGFR), ER (Novacastra; Leica Microsystems GmbH, Wetzlar, Germany), CK6A (Novus Biologicals, Littleton, CO, USA), E-cadherin (Abcam), integrin α6, ZEB1, LSR (Atlas Antibodies, Stockholm, Sweden), S100A14 (Protein Tech Group, Chicago, IL, USA), CD24 (Neo Markers, Fremont, CA, USA), integrin β1 (Biogenex, Fremont, CA, USA), fluorescein isothiocyanate (FITC)-conjugated goat anti-rabbit IgG (Pierce; Thermo Fisher Scientific, Rockford, IL, USA), and FITC-conjugated goat anti-mouse (Zymed; Life Technologies Corporation). Western blotting included the following: ZEB1 (Atlas Antibodies), S100A14 (Protein Tech Group) and goat anti-rabbit IgG (Dako).

Flow Cytometry

Cells were resuspended in HEPES buffer with protease inhibitors (Roche Diagnostics, Hvidovre, Denmark). For intracellular staining, cells were fixed in ethanol, permeabilized in 0.2% saponin and blocked with 10% normal goat serum. Samples were incubated with fluorochrome-conjugated antibodies followed by flow cytometry (FACSCalibur; BD Biosciences, San Jose, CA, USA) or fluorescence-activated cell sorting (FACSDiva; BD Biosciences). Mean fluorescence intensity is defined as geometric mean value of the fluorescence.

Mammosphere Assay

Dilution series of cells were seeded on ultra-low attachment plates in serum-free medium and incubated for 18 d. Clones were trypsinized into single-cell suspensions, diluted and seeded into new wells for mammosphere formation four times successively to assess the ability to self-renew.

Immunohistochemistry and Cytochemistry

Tissue or cells were fixed in formalin and were paraffin embedded. Antigen retrieval was performed by microwave boiling in T-EG buffer/TRS buffer (Dako). Primary antibody detection was either horseradish peroxidase (HRP)-conjugated EnVision™ polymer (K4001; Dako), PowerVision Poly-HRP (Novocastra; Leica Microsystems) or EnVison FLEX (K8010; Dako) with 3,3′-diaminobenzidine (DAB) (K3468; Dako) as chromogen. Microscopy of tissues was performed on a Leica DMLB microscope (100×/nummerical aperture [NA] 1.25; Leica Microsystems) using LasV3.6 acquisition software.

Animals and Surgery

Orthotopic transplantation of tumor cells was achieved by injection into surgically exposed mammary fat pads of anesthetized NOG or CB-17 severe combined immunodeficienc (SCID) female mice aged 8 wks (Taconic, Ejby, Denmark). Before inoculation, cells were resuspended in extracellular matrix from the Engelbreth-Holm-Swarm sarcoma (Sigma-Aldrich) and culture medium. Endpoint was a tumor size of 1.2 cm or tumor growth for 6 months. The mice were euthanized by cervical dislocation, and tumors were fixed in formalin and paraffin embedded. For evaluation of metastatic potential of A4 and G4 cells, 10^6 Luc2-transfected cells were injected into the tail veins of NOG mice. Relative quantification of metastasis was performed every second week for 8 wks by using whole-body bioluminescent imaging (IVIS spectrum; Caliper Life Sciences, Hopkinton, MA, USA). Mice were injected with 150 mg D-luciferin/kg body weight, gas anesthetized with isoflurane and imaged. Data analysis was performed using Caliper Life Sciences Living Image (version 4.0). All animal procedures were performed in accordance with a protocol approved by the Experimental Animal Committee, The Danish Ministry of Justice.

Proliferation Inhibition Assay

Twenty-four hours after seeding, cells were subjected to chemotherapeutics (two-fold dilutions of the drug from a concentration wherein the most sensitive of the cell lines had a 70–90% reduction in relative proliferation) (0.2–40 μmol/L paclitaxel, 0.05–0.8 μmol/L doxorubicin, 0.07–75 μmol/L methotrexate, 0.13–20 μmol/L salinomycin) for 96 h, followed by replacement of MEBM without phenol-red (Gibco; Life Technologies Corporation) containing 10% 5 mg/mL MTT solution (Sigma-Aldrich). After 4 h of incubation, 100 μL MTT solvent was added to dissolve the insoluble formazan crystals. The amount of formazan was colorimetrically measured at 590 nmol/L. *P* values are based on two-sided *t* tests of triplicates.

Quantitative Proteomics

HMT3909S13 was subjected to "stable isotope labeling by amino acid in cell culture" (SILAC) to allow quantitative comparison of proteins/peptides by mass spectrometry (35,36). HMT3909S13 was propagated as described above, but in custom-made MEBM without L-arginine, L-lysine and L-glutamine (Lonza) supplemented with L-glutamine and either light ($^{12}C_6$) or heavy ($^{13}C_6$) lysine and arginine ($^{12}C_6$ [Sigma-Aldrich]; $^{13}C_6$ [Cambridge Isotope Laboratories, Andover, MA, USA]). Cells ($^{12}C_6$ and $^{13}C_6$) were mixed 1:1 and labeled with antibodies for flow cytometry sorting into $CD44^{hi}/CD24^{-}$ and $CD44^{hi}/CD24^{lo}$. Sorted cells were lysed in a hypotonic buffer (10 mmol/L Tris-base, 1.5 mmol/L $MgCl_2$, 10 mmol/L NaCl, pH 6.8) supplemented with protease inhibitors (Roche) followed by sedimentation (300*g*, 5 min), resuspension in a gradient buffer (0.25 mol/L sucrose, 10 mmol/L HEPES, 100 mmol/L succinic acid, 1 mmol/L EDTA, 2 mmol/L $CaCl_2$, 2 mmol/L $MgCl_2$, pH 7.4), and homogenization using a motor-driven Potter homogenizer (Sartorius; Fisher Scientific GmbH, Schwerte, Germany). After enzymatic digestion with lysyl endoproteinase and trypsin, peptides were analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS; Q-TOF Micro Tandem Mass Spectrometer; Waters Corporation, Milford, MA, USA) using a 2-h stepped linear gradient (37). All relevant mass spectrometry spectra were validated manually to ensure high signal-to-

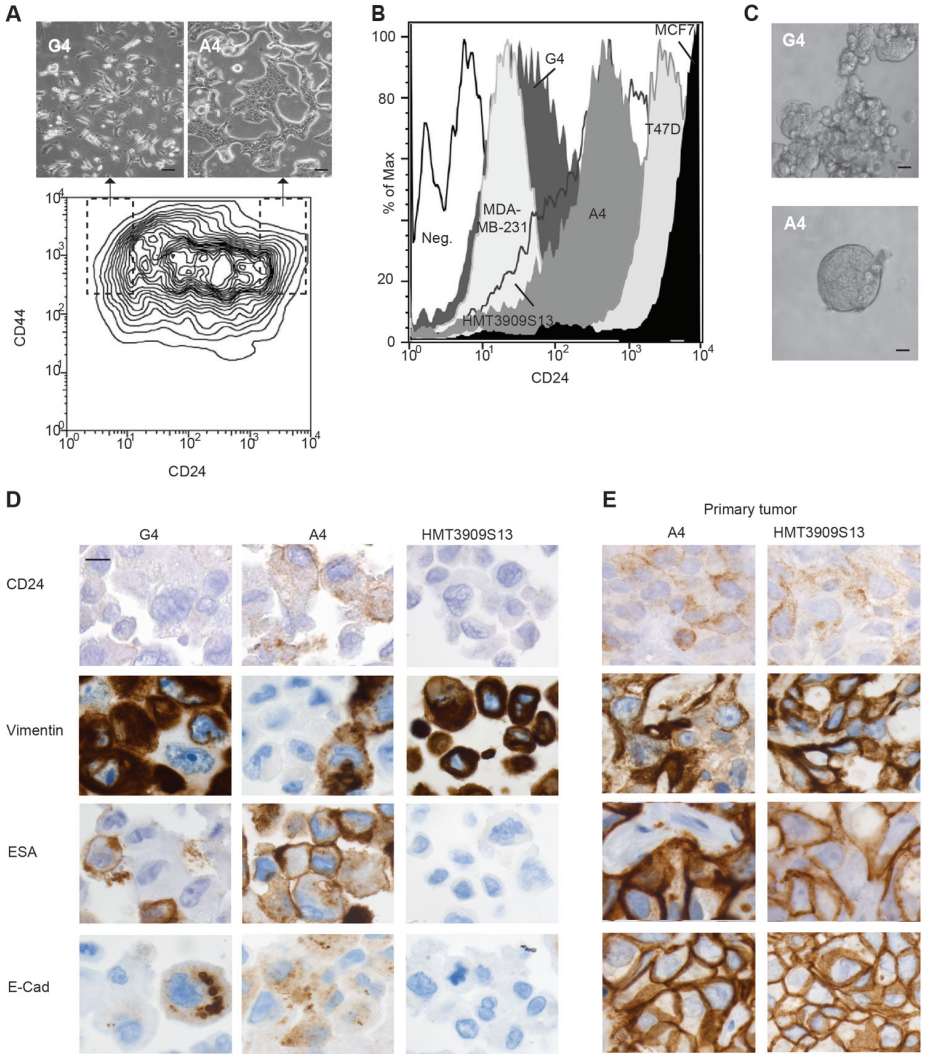


Figure 1. Phenotypic characterization of single-cell clones obtained from the parental cell line, HMT3909S13, demonstrates G4 as $CD44^{hi}/CD24^{-}$ with a mesenchymal-like morphology and A4 as $CD44^{hi}/CD24^{lo}$ with an epithelial-like morphology capable of forming mammospheres. A4 cells can recapitulate cells similar to the parental cell line *in vivo*. (A) CD44 and CD24 expression of HMT3909S13 as determined by flow cytometry. The $CD44^{hi}/CD24^{-}$ G4 and $CD44^{hi}/CD24^{lo}$ A4 single-cell clones were obtained from marked gates. G4 exhibits a mesenchymal-like morphology, whereas A4 exhibits an epithelial-like morphology, as determined by phase contrast imaging. (B) Relative CD24 expression of HMT3909S13, A4 and G4 compared with the MDA-MB-231, T47D and MCF7 cell lines, as determined by flow cytometry. HMT3909S13 without primary antibody is shown as a representative negative control. (C) Mammospheres were formed from A4, but not G4, after growth of cells in nonadherent culture flasks at d 7, as depicted by phase contrast imaging. Only loosely adhered clumps of cells were observed in the G4 cell culture flask. Scale bar: 30 μm. (D) ICC analysis of CD24, vimentin, ESA and E-cadherin demonstrating altered expression levels between the three cell lines. Scale bar: 20 μm. (E) IHC staining of primary tumors from immunodeficient mice generated by inoculation of A4 and HMT3909S13 cells, respectively, into the mammary fat pad showing the ability to express the proteins heterogeneously and more similar to the A4 than HMT3909S13 cells *in vitro*, emphasizing the CSC-like characteristics of A4 clone cells. See also Supplementary Figures S2 and S3 and Supplementary Table S1.

noise ratios. Only proteins that exhibited altered expression in at least two samples and were identified with at least two peptides were subjected to further analysis.

Real-Time Polymerase Chain Reaction

Relative quantification of gene expression was performed in triplicate using SYBR Green PCR Master Mix (Applied Biosystems; Life Technologies Corporation). Median relative expression levels were normalized by using the reference genes TATA box binding protein (*TPB*), *β-actin* and Na^{+}/K^{+} ATPase (*ATP1A1*). The primers for specific amplification were obtained from QuantiTect Primer assays collection (Qiagen, Hilden, Germany): *TBP*: QT00000721; *β-actin*: QT00095431; *ATP1A1*: QT00059962; *VIM*: QT00095795; *MUC1*: QT00015379; *KRT6A*: QT01666791; *KRT19*: QT00081137; *S100A14*: QT00217028. Primers for *CD24* were obtained from Invitrogen: F (CD24) GCACT GCTCC TACCC ACGCA GATTT, R (CD24) GCCTT GGTGG TGGCA TTAGT TGGAT; and primers for *CD44* were obtained from Hill *et al.* (38): F (CD44) TTTGC ATTGC AGTCA ACAGT C, R (CD44) GTTAC ACCCC AATCT TCATG TCCAC.

Quantitative Transcriptomics

mRNA was purified by TRIzol followed by ethanol precipitation. Each cell line in five biological replicates was analyzed on GeneChip® Human Gene 1.0 ST Arrays (Affymetrix, Santa Clara, CA, USA). Microarray data were normalized and processed using the Partek® Genomics Suite (false discovery rate [FDR] 0.01) (Gene Expression Omnibus [GEO] accession number GSE32455). Out of 599 genes, 406 were compared with a publicly available data set (GEO accession number GSE7390) (39) as training set for a prognostics signature. The training set was divided into ER^{+} (n = 134) and ER^{-} (n = 64) patients before classification. The remaining 193 genes were not used because of absence of probe IDs or transcripts in the patient data. Data from GSE2990 (40) (KJ125) was chosen as the

testing set (n = 34). Only data sets were chosen with available information on patient ER^+/ER^- status, time to distant metastasis, negative lymph node status and no systemic treatment in addition to data sets obtained from Affymetrix platforms. Kaplan-Meier plot and Cox proportional hazard regression model was performed on the two data sets (n = 98) using Partek Genomics Suite (Partek Inc., St. Louis, MO, USA).

Western Blotting

On the basis of protein concentration, equal amounts of cell lysates were resolved on 4–20% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Pierce; Thermo Fisher Scientific, Rockford, IL, USA) and electroblotted onto a polyvinylidene fluoride (PVDF) membrane. The membranes were blocked and incubated with primary antibody and HRP-conjugated secondary antibody, respectively. The immunoreactive bands were visualized using an ECL Western Blot Kit (GE Healthcare, Buckinghamshire, UK). Loading of equal amounts of lysate in each lane was further validated by Coomassie staining of the PVDF membrane.

All supplementary materials are available online at www.molmed.org.

RESULTS

Phenotypic Characterization of Single-Cell Clones with CSC-Like Features

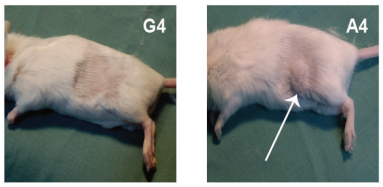
To study the $CD44^{hi}/CD24^{-/lo}$ subpopulation at a single-cell level, a panel of breast cancer cell lines was examined for CD44/CD24 expression to identify one that contained a high percentage of $CD44^{hi}/CD24^{-/lo}$ cells. The triple-negative cell line HMT3909S13, which is also known to exhibit a bilineage phenotype, was found to fulfill this requirement, and a large panel of isogenic cell clones with different phenotypes was isolated by flow cytometry sorting of HMT3909S13 by using the gates $CD44^{hi}/CD24^-$ and $CD44^{hi}/CD24^{lo}$, respectively, and transfer of the sorted cells directly into individual wells for single-cell cloning. Two single-cell clones were selected for further comparison and designated $CD44^{hi}/CD24^-$ (clone G4) and $CD44^{hi}/CD24^{lo}$ (clone A4), respectively (Figure 1A). The level of CD24 expression in G4 was equal to that of the basal B-cell line, MDA-MB-231, which is characterized as $CD24^-$. The level of CD24 expression in A4 was 10- and 20-fold lower, respectively, compared with the luminal cell lines T47D and MCF7, usually designated $CD24^+$ (Figure 1B) (8,41). The level of CD24 expression differed six-fold between the $CD44^{hi}/CD24^-$ G4 clone and the $CD44^{hi}/CD24^{lo}$ A4 clone. The $CD44^{hi}/CD24^-$ G4 cells exhibited a spindle shape consistent with a mesenchymal-like morphology resembling basal B cells, whereas the $CD44^{hi}/CD24^{lo}$ A4 cells were shaped like cobblestones and grew in clusters, consistent with an epithelial-like morphology resembling basal A cells (see Figure 1A). G-band karyotyping of the HMT3909S13, G4 and A4 showed identical genetic abnormalities in all three cell lines, confirming the isogenic background (Supplementary Figure S1). A total of 44 additional single-cell clones obtained from HMT3909S13 were analyzed according to CD44/CD24 phenotypes and morphology. The majority of the $CD44^{hi}/CD24^-$ phenotype clones exhibited mesenchymal-like morphology, whereas the majority of clones with a $CD44^{hi}/CD24^{lo}$ phenotype exhibited epithelial-like morphology. Some clones also exhibited a reverse relationship or mixed populations, indicating that CD24 levels in the $CD44^{hi}/CD24^{-/lo}$ subpopulation are not directly associated with individual cell morphology (42).

The ability to self-renew and form mammospheres *in vitro* was examined by seeding single-cell suspensions in nonadherent culture flasks, resulting in mammosphere formation of $CD44^{hi}/CD24^{lo}$ A4 clone within 2 wks. These could, by dissociation and upon transfer to new flasks, lead to new mammospheres. In contrast, the $CD44^{hi}/CD24^-$ G4 clone mostly produced loosely adherent clumps of cells (Figure 1C).

A

Number of cells inoculated	10^2	10^3	10^4	10^5	10^6	Total
CD44hi/CD24-, G4						
CB17	-	0/2	0/3	0/3	0/3	0/11
NOG	-	-	-	0/1	0/4	0/5
CD44hi/CD24lo, A4						
CB17	-	-	2/2	1/3	-	3/5
NOG	(0/3)	2/3	3/3	4/4	1/1	10/11
HMT3909S13						
CB17	-	-	-	1/1	-	1/1
NOG	-	-	-	1/1	1/1	2/2
Hs578t						
NOG	-	-	-	-	0/3	0/3

B



C

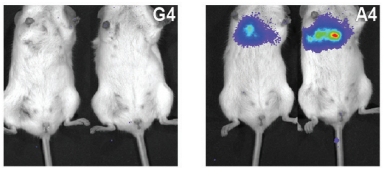


Figure 2. Only the $CD44^{hi}/CD24^{lo}$ A4 and the parental HMT3909S13 cell lines initiated tumors in immunodeficient mice. (A) Tumor formation upon orthotopic inoculation into the mammary fat pad of $CD44^{hi}/CD24^-$ G4, $CD44^{hi}/CD24^{lo}$ A4, HMT3909S13 or Hs578T cells into two strains of immunodeficient mice (CB17-SCID and NOG). Ratios refer to the number of mice in which primary tumors developed relative to total number of mice inoculated with cells within each group. (B) Tumors established after orthotopic inoculation of 10^4 A4 cells in a NOG mouse (right, arrow), whereas no tumor formation was observed in any of the NOG mice orthotopically inoculated by 10^6 G4 cells (left). (C) Whole-body bioluminescent imaging of lung metastases established after tail vein inoculation of 10^6 Luc2-transfected A4 or G4 cells.

The two clones and the parental HMT3909S13 cell line were next stained with a panel of antibodies against relevant markers (Figure 1D, Supplementary Table S1). The expression of ER, PR and HER2 were negative in all three cell

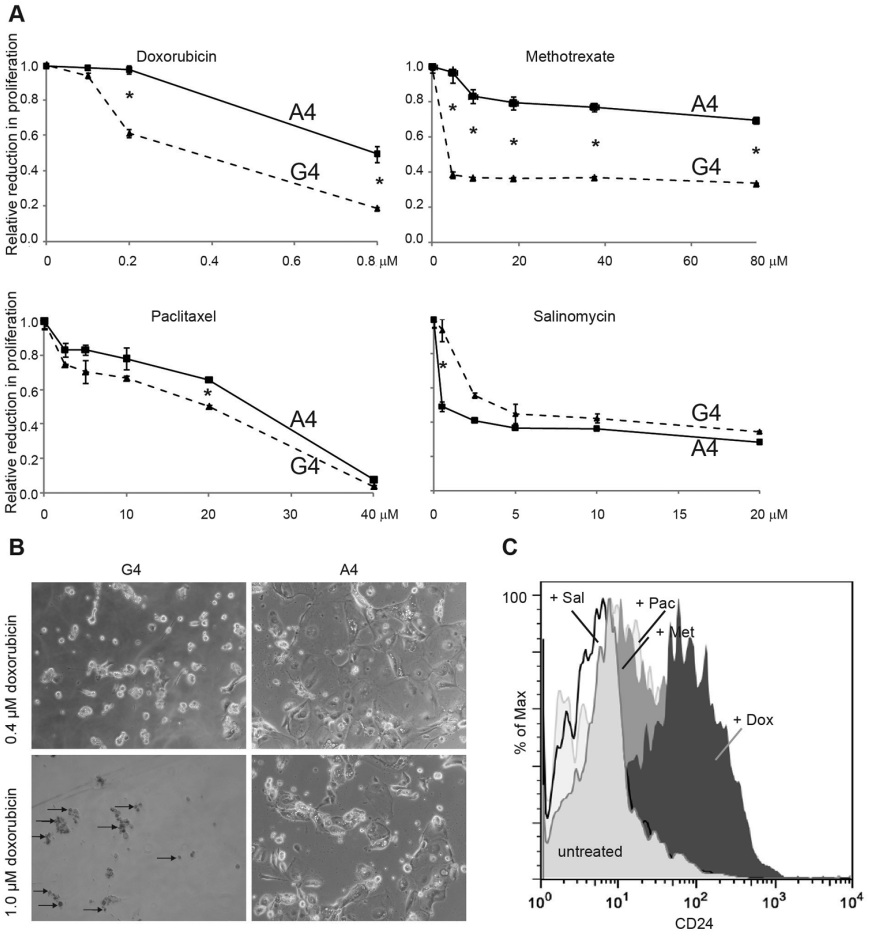


Figure 3. CD44^{hi}/CD24^{lo} A4 cells are significantly more resistant to conventional chemotherapeutic reagents, but are more sensitive to salinomycin treatment than CD44^{hi}/CD24^{-} G4 cells. (A) Relative reduction in proliferation of A4 and G4 cells after treatment with various chemotherapeutics in at least four different concentrations for 96 h, as determined by an MTT proliferation assay. Three independent experiments were performed, and a representative experiment is shown. *$p < 0.01$. (B) Phase contrast pictures of G4 versus A4 treated for 10 d with 0.4 or 1 μmol/L doxorubicin, respectively. The arrows show dead Tryphan blue–stained cells. (C) CD24 expression of HMT3909S13 as shown by flow cytometry after treatment with 0.4 μmol/L doxorubicin, 40 μmol/L methotrexate, 5 μmol/L paclitaxel or 2 μmol/L salinomycin for 10 d, showing a selection of doxorubicin-, methotrexate- and paclitaxel-treated cells that express more CD24 compared with untreated cells. No selection is observed with salinomycin treatment.

lines, whereas the relatively absent and low expression level of CD24 in G4 and A4, respectively, as observed by flow cytometry, were validated by ICC. Vimentin, cytokeratin 17 and E-cadherin exhibited higher expression and cytokeratins 5/6 lower expression in G4 versus A4 cells. However, E-cadherin was expressed in a larger percentage of A4 compared with G4 cells. More A4 than G4 cells expressed ESA, but with approximately equal intensities. Cytokeratins 7/8, 14, 18 and 19; HER1; and integrins β1 and α6 were equally expressed in the two cell lines. Interestingly, HMT3909S13 did not express E-cadherin or ESA, whereas both the G4 and A4 cells showed intense expression, although at varying levels, supporting the CSC features of the selected clones.

The cell lines are considered stable since the A4 and G4 clones have been grown discontinuously under stable growth conditions and in the presence of growth factors for >2 years without phenotypic changes (Supplementary Figure S2).

An Epithelial-Like CSC-Like Clone, but Not a Mesenchymal-Like CSC-Like Clone, Exhibited *In Vivo* Tumor Initiation Capacity

CD44^{hi}/CD24^{-} G4, CD44^{hi}/CD24^{lo} A4 and HMT3909S13 were inoculated into the mammary fat pad of two strains of immunodeficient mice, CB17 SCID and NOG. Primary tumors were formed within a period of 90 d upon orthotopic inoculation of as low as 10^3 A4 cells and 10^5 HMT3909S13 cells, respectively, whereas no tumor formation was observed from inoculation of up to 10^6 G4 cells after 180 d of observation (Figures 2A, B). The percentage of mice that formed tumors upon inoculation with A4 cells was higher in NOG mice than in CB17-SCID mice. Metastatic potential was evaluated by tail vein injection and whole-body bioluminescent imaging of Luc2 transfected A4 and G4 cell lines in NOG mice and revealed that lung metastases were established within a period of 3 wks by injection of 10^6 A4 cells. However, G4 did not give rise to lung metastases within a period of 8 wks (Figure 2C).

Whereas well-characterized basal A cell lines (for example, BT20 and MDA-MB-468) are known to establish tumors in immunodeficient mice (43), the lack of tumor potential of basal B cells, other than G4, was examined by inoculation of 10^6 cells of another well-characterized basal B-cell line, Hs578T (32), into the mammary fat pad of NOG mice. The cell line did not show any sign of tumor formation after 90 d of observation.

IHC staining was performed to evaluate whether the tumor cells in the primary tumors established by A4 and HMT3909S13 cells exhibited the same markers as the cells used for inoculation (Figure 1E). Comparing the expression of CD24, vimentin, ESA and E-cadherin in the primary tumors established by A4

with that of A4 cells *in vitro* showed some differences, for example, vimentin and E-cadherin were more highly expressed in the primary tumor established by A4 compared with A4 cells grown *in vitro*, and ESA became more heterogeneously expressed in the primary tumor. Interestingly, the staining pattern of the tumors formed by inoculation of A4 and HMT3909S13 cells, respectively, closely resembled each other. This result indicates that inoculation of A4 cells into mice leads to primary tumors containing cells with diverse expression patterns of epithelial/mesenchymal markers that recapitulate the original heterogeneity of HMT3909S13 cells.

An Epithelial-Like CSC-Like Clone Exhibited More Resistance to Conventional Chemotherapy than a Mesenchymal-Like CSC-Like Clone

To evaluate the sensitivity/resistance of G4 and A4 to different chemotherapeutic compounds, the cells were treated for 96 h *in vitro* with the conventional chemotherapeutics doxorubicin (antracycline, 0.05–0.8 μmol/L), paclitaxel (taxane, 0.2–40 μmol/L) and methotrexate (antifolate, 0.07–75 μmol/L), as well as salinomycin (antibacterial ionophore, 0.125–20 μmol/L), which recently were reported to predominantly target CSCs (44). Sensitivity/resistance toward chemotherapeutic compounds was measured using an MTT proliferation assay that showed the relative reduction in proliferation after chemotherapy treatment *in vitro* (Figure 3A), whereas the death of treated cells was shown by Tryphan blue staining and cell counting (Figure 3B). The A4 clone was significantly ($p < 0.01$) more resistant to doxorubicin, methotrexate and paclitaxel treatment than G4. In contrast, G4 was significantly ($p < 0.01$) more resistant to salinomycin treatment than A4.

HMT3909S13 cells that survived treatment with conventional chemotherapeutics were stained for CD44 and CD24 and compared with untreated cells by flow cytometry. Interestingly, treatment with 0.4 μmol/L doxorubicin, 40 μmol/L methotrexate or 5 μmol/L paclitaxel for 10 d led to an eight-, three- and two-fold higher mean fluorescence intensity for CD24, respectively, compared with untreated cells. On the basis of the expression of CD24, treatment with salinomycin (2 μmol/L) did not select for a specific population (Figure 3C). CD44 expression showed no changes after any of the above treatments compared with the untreated cell line. This result suggests that only cells expressing some CD24 (such as the A4 cells) survived treatment with doxorubicin, methotrexate and paclitaxel.

A

Protein name	UniProt accession number	*peptides/ number of samples	**Fold higher expression in CD44hi/CD24-
Ribosomal protein L7	P18124	3/3	2.1
Vimentin	P08670	20/21	3.8

Protein name	UniProt accession number	*peptides/ number of samples	***Fold higher expression in CD44hi/CD24lo
CK 5	P13647	6/8	1.5
CK6A	P02538	17/10	2.5
CK7	P08729	16/9	1.7
CK8	P05787	11/13	2
CK18	P05783	12/10	1.5
CK19	P08727	19/19	1.9
MUC1	P14941	2/5	1.6
S100A14	Q9HCY	2/4	2.5

*Maximum number of quantified peptides per sample/number of samples
**CD44hi/CD24- relative to CD44hi/CD24lo
***CD44hi/CD24lo relative to CD44hi/CD24-

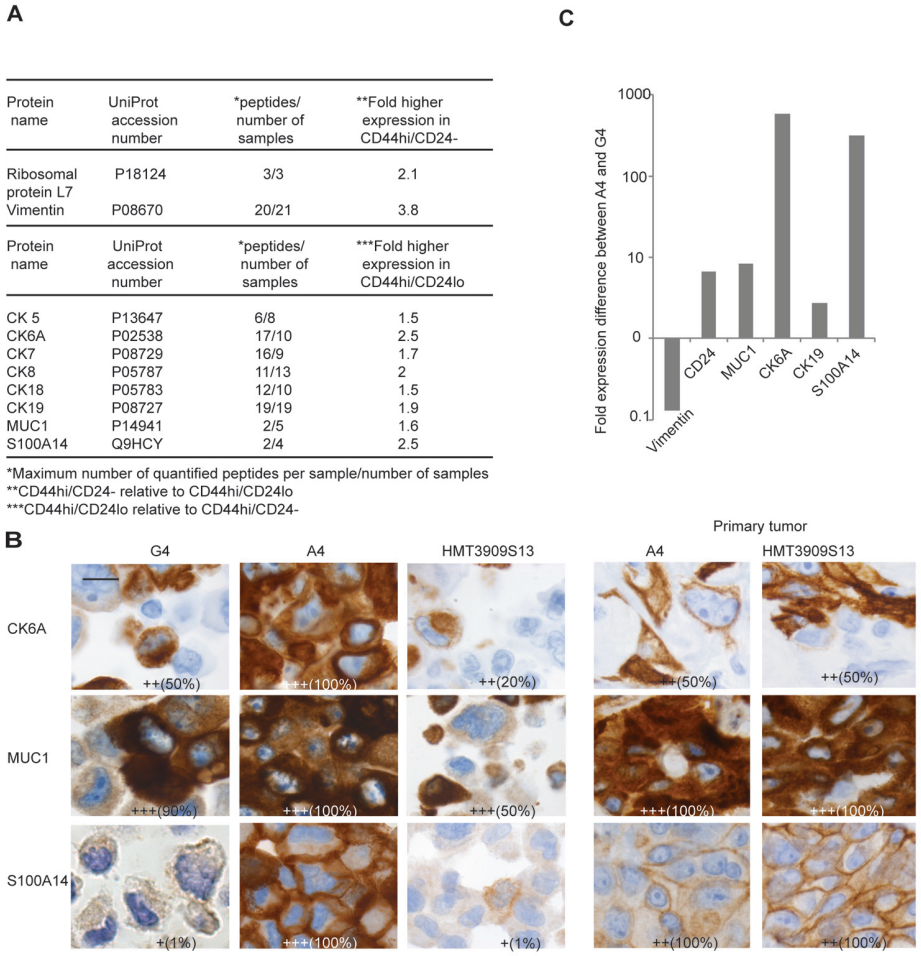


Figure 4. Identification of proteins exhibiting altered expression levels between $CD44^{hi}/CD24^{lo}$ and $CD44^{hi}/CD24^{-}$ cells using comparative quantitative mass spectrometry–based proteomics of SILAC cells and validation of selected proteins. (A) A list of proteins exhibiting altered expression between $CD44^{hi}/CD24^{lo}$ and $CD44^{hi}/CD24^{-}$ cells. (B) ICC with estimated staining intensity and number of cells expressing CK6A, MUC1 and S100A14 in G4, A4 and the parental cell line compared with IHC of primary tumors established with A4 and HMT3909S13 inoculated into NOG mice. Scale bar: 20 μm. (C) Fold-change expression levels of mRNA of selected candidates in A4 and G4 measured by quantitative real-time PCR identifying *CD24*, *MUC1*, *KRT6A* (*CK6A*), *KRT19* (*CK19*) and *S100A14* upregulation in A4 at the mRNA level, whereas *VIM* is downregulated compared with G4. The number 1 (*y* axis) corresponds to equal mRNA expression in the two cell lines.

Novel Markers Distinguishing the $CD44^{hi}/CD24^{-/lo}$ Subsets Identified by Quantitative Proteomics

Proteins distinguishing the $CD44^{hi}/CD24^{lo}$ and $CD44^{hi}/CD24^{-}$ phenotypes were identified by comparative quantita-

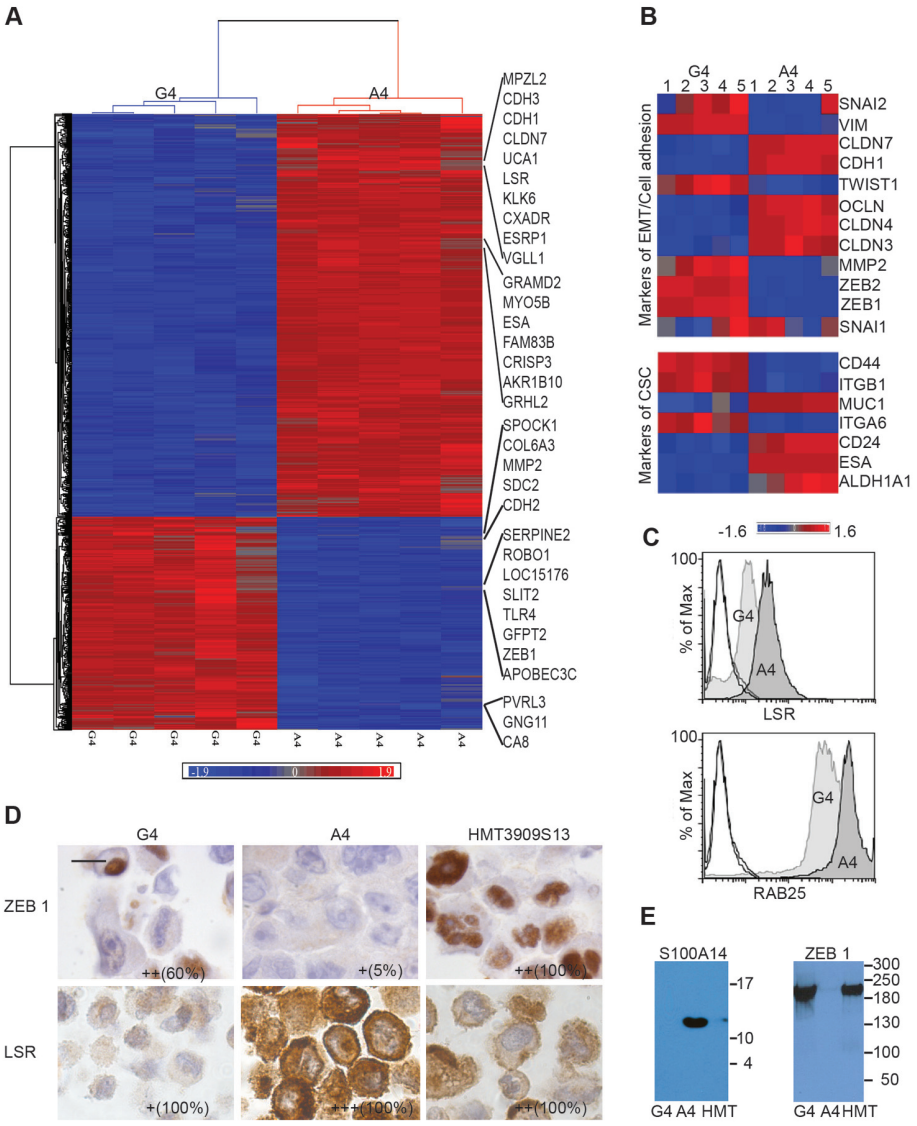


Figure 5. Transcriptomic analysis of CD44^{hi}/CD24^{lo} G4 and CD44^{hi}/CD24^{-} A4 clones identifying genes and gene signatures related to the CSC-like phenotype and potential new biomarkers. (A) Heat map of the genes differentially expressed more than two-fold in A4 relative to G4 (FDR 0.01), showing representative gene clusters. (B) Heat map focusing on gene expression levels of EMT markers: *VIM*, transcription factors (*ZEB1*, *ZEB2*, *TWIST1*), snail 1 and 2 (*SNAI1*/*SNAI2*), matrix metalloproteinase-2 (*MMP2*), E-cadherin (*CDH1*) and claudins (*CLDN3*, *CLDN4*, *CLDN7*) and CSC markers: *CD24*, *CD44*, aldehyde dehydrogenase (*ALDH1A1*), epithelial cell adhesion (*ESA*) molecule, *MUC1*, integrin β1 (*ITGB1*) and integrin α6 (*ITGA6*). Each colored square represents the relative transcript abundance (log 2 space) of each of five biological replicates of A4 and G4, respectively. Highest expression is red and lowest is blue. (C) Flow cytometry analysis of LSR and RAB25 showing higher expression in A4 versus G4. Controls without primary antibody were included. (D) Comparison of the expression levels of ZEB1 and LSR in G4, A4 and HMT3909S13 by ICC showing higher expression of ZEB1 and lower expression of LSR in G4 versus A4. Scale bar: 20 μm. (E) Western blotting of ZEB1 and S100A14 showing higher expression of ZEB1 and lower expression of S100A14 in G4 versus A4. The parental cell line (HMT3909S13) shows intermediate expression of both proteins. See also Supplementary Figure S4 and Supplementary Table S2.

tive mass spectrometry of SILAC-labeled cells sorted from the parental cell line HMT3909S13 (Figure 4A). A number of cytokeratins (CK5, CK7, CK8, CK18 and CK19) and vimentin, in the proteomic study, exhibited altered expression between CD44^{hi}/CD24^{lo} versus CD44^{hi}/CD24^{-}. In addition, four potentially interesting proteins (S100A14, MUC1, CK6A and ribosomal protein L7) exhibited altered expression. Three of these (S100A14, MUC1 and CK6A) were validated by ICC, showing more intense CK6A and S100A14 staining in A4 versus G4 cells, whereas more A4 than G4 cells stained for MUC1. Fewer cells in the HMT3909S13 cell line than in the A4 and G4 sublines expressed S100A14, MUC1 and CK6A. IHC of primary tumors generated from A4 and HMT3909S13 cells for MUC1 and S100A14 showed staining similar to ICC of A4, whereas CK6A showed a more heterogeneous staining, similar to G4 cells (Figure 4B). The observed altered protein expression levels of all three proteins (MUC1, S100A14 and CK6A) were shown to correspond with alterations in mRNA expression levels measured by quantitative real-time polymerase chain reaction (PCR) (Figure 4C).

Gene Array Profiling of the CD44^{hi}/CD24$^{-/lo}$ Subsets Revealed Novel Markers and Pathways Linked to Tumor-Initiating Cells as well as a Prognostic Gene Signature

To define a gene expression signature of CSC-like cells with tumor-initiating and chemoresistant potential, gene arrays of five biological replicates of CD44^{hi}/CD24^{lo} A4 and CD44^{hi}/CD24^{-} G4, respectively, were compared, revealing 1,777 genes that were upregulated and 1,883 that were downregulated ($p <$ 0.0001, FDR 0.01). Of these, 599 genes exhibited more than two-fold difference in A4 versus G4 (Figure 5A and Supplementary Table S2). The fold-changes were in the range of 2- to 71.7-fold. Ingenuity pathway analysis revealed several networks, including a network of 59 genes that were upregulated more than

three-fold in A4 versus G4, including *CDH1*, *MUC1*, *JUP*, *DSP*, *MUC16* and *FOXA1*, genes already known to be involved in cancer, reproductive systems and genetic disease (Supplementary Figure S4).

Focusing on important pathways described for EMT and CSC, the expression of central genes were evaluated. Expression of epithelial cell–cell adhesion genes, such as E-cadherin, claudin 3, claudin 4, claudin 7 and occludin, were significantly higher in A4 than G4, whereas EMT markers such as vimentin, twist1, matrix metalloproteinase 2, ZEB1 and ZEB2 were lower in A4 versus G4. Snail1 and snail2 were fairly equally expressed between the two clones. Protein markers of CSC and epithelial differentiation, such as MUC1, ESA, CD24 and aldehyde dehydrogenase, were more highly expressed in A4 versus G4, whereas CD44, integrin α6 and integrin β1 were more highly expressed in G4 versus A4 (Figure 5B). To identify potential novel biomarkers capable of distinguishing $CD44^{hi}/CD24^{-/lo}$ cancer cells with and without tumorigenic and chemoresistant potential in a tumor cell population of triple-negative breast cancer, candidate biomarkers were selected from genes that exhibited extensive alteration in expression between A4 versus G4. These genes were examined at the protein level, including *LSR*, *RAB25* and *S100A14*, which were upregulated at the mRNA levels 11-, 24- and 10-fold, respectively, in A4 versus G4 and *ZEB1*, which was upregulated 11-fold in G4. The higher expression of the cell surface proteins LSR and RAB25 in A4 versus G4 were confirmed by flow cytometry (Figure 5C). LSR and ZEB1 were examined by ICC, showing higher expression of LSR and lower expression of ZEB1 in A4 versus G4, respectively (Figure 5D). The altered expression of ZEB1 was also confirmed by Western blotting (Figure 5E). Additionally, Western blotting demonstrated higher expression of S100A14 in A4 versus G4 (see Figure 5E).

To examine whether A4 and G4 exhibited characteristics similar to the basal A and basal B subpopulations, the expression of selected proteins that exhibited altered expression in A4 versus G4 was analyzed in well-characterized cell lines of basal A (MDA-MB-468 and BT20) and basal B (Hs578T) by using ICC. Similar to A4, MDA-MB-468 and BT20 exhibited higher expression of S100A14, LSR, MUC1 and ESA than G4 and Hs578T. Similarly, G4 and Hs578T exhibited higher expression of vimentin and ZEB1 compared with A4, MDA-MB-468 and BT20 (Figures 1D, 4B, 5D and 6).

As it may be expected that breast cancers exhibiting a higher frequency of CSC may have a poor outcome and more frequent disease recurrence, we examined whether some of the 599 genes identified by comparison of A4 and G4 could predict distant metastasis of lymph node–negative primary breast cancers. A microarray data set of 198 tumor samples of lymph node–negative breast cancer patients who had received no adjuvant systemic treatment was divided into ER^+ and ER^- patients (39). Because our cell model system was generated from an ER^- tumor, it was expected that the gene signature only would relate to this breast cancer subtype. The study contained 64 ER^- patients, of which 25 developed dis-

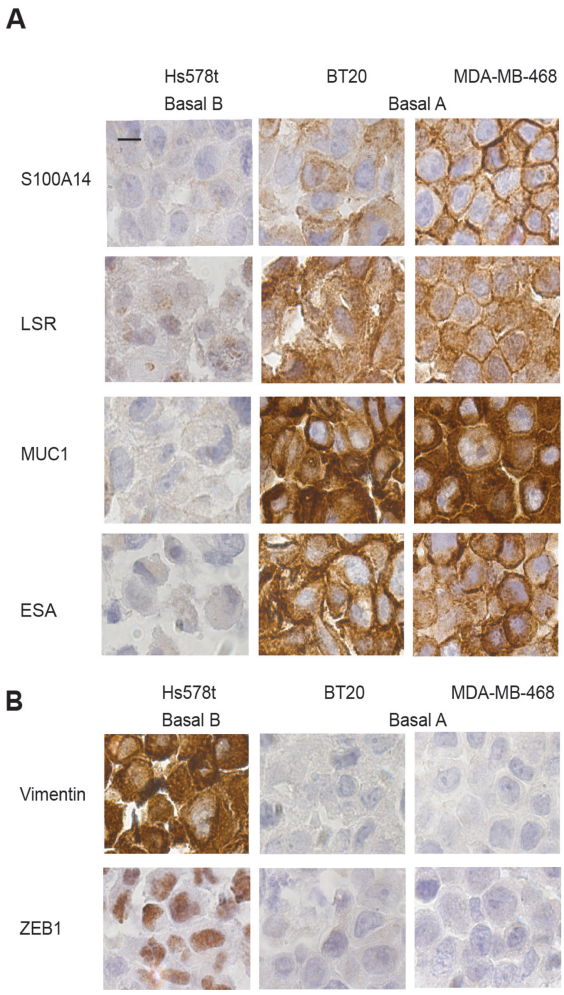
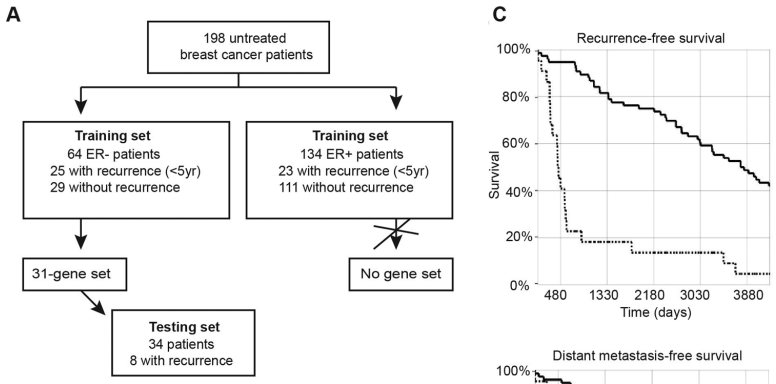


Figure 6. Proteins identified to exhibit altered expression between A4 and G4 were further analyzed in cell lines of basal A and basal B subtype. (A) ICC of S100A14, LSR, MUC1 and ESA showed higher expression in the basal A cell lines, BT20 and MDA-MB-468, relative to the basal B cell line, Hs578T. (B) Vimentin and ZEB1 showed higher expression in the basal B cell line relative to the basal A cell lines.



B

Symbol	Entrez Gene name
ADAP1	ArfGAP with dual PH domains 1
ANXA6	annexin A6
APOBEC3C	apolipoprotein B mRNA editing enzyme,
B3GNT3	BGNT3
CACNG4	Voltage-dependent calcium channel gamma-4
CDH2	cadherin 2, type 1, N-cadherin
DSP	desmoplakin
EPCAM/ESA	epithelial cell adhesion molecule
EYA4	eyes absent homolog 4
FGFR2	fibroblast growth factor receptor 2
GALNT3	Polypeptide N-acetylgalactosaminyltransferase 3
GPR56	G protein-coupled receptor 56
GRHL2	grainyhead-like 2
HIST1H3A	histone cluster 1, H3a
JUP	junction plakoglobin
KRT13	keratin 13
LCN2	lipocalin 2
LSR	lipolysis stimulated lipoprotein receptor
MAP1B	microtubule-associated protein 1B
MPZL2	myelin protein zero-like 2
OCLN	occludin
PADI4	peptidyl arginine deiminase, type IV
PDE3B	phosphodiesterase 3B, cGMP-inhibited
PRKAR2B	protein kinase, cAMP-dependent, regulatory, type II, beta
PVRL3	poliovirus receptor-related 3
RARRES1	retinoic acid receptor responder (tazarotene induced)
SLC7A7	solute carrier family 7 member 7
SLC7A8	solute carrier family 7 member 8
SLCO1B3	solute carrier organic anion transporter family, member 1B3
SPINT1	serine peptidase inhibitor, Kunitz type 1
ST14	suppression of tumorigenicity 14

Figure 7. A 31-gene breast cancer signature capable of predicting distant metastasis of lymph node–negative ER^- primary breast cancer. (A) A microarray data set of a total of 198 lymph node–negative breast cancer patients (64 ER^- and 134 ER^+) were used as a training set to identify a 31-gene signature from the 599 genes exhibiting altered expression between A4 and G4. A testing set of 34 patients in combination with the training set was used to validate the signature (39,40). (B) The 31 genes included in the prognostic signature of ER^- patients. (C) Kaplan-Meier plots illustrating the probability of relapse-free and distant metastasis-free survival predicted by the 31-gene signature. Upper curve (solid line) illustrates patients predicted as metastasis-free within a 5-year (1,825-d) period. Lower curve (dashed line) predicts patients with distant metastasis before a 5-year period.

tant metastasis within a 5-year period (Figure 7A). A 31-gene signature was identified in this training set by using a leave-one-out cross-validation procedure with the K-nearest neighbor algorithm as a classifier (Figure 7B). Subsequently, the accuracy of the 31-gene signature was tested in an independent set of 34 patients (40) combined with the training set, demonstrating an accuracy of 86.6% (sensitivity of 70.37%, specificity of 97%).

The very high specificity indicates that the signature is accurate in identifying patients with triple-negative breast cancer who will not experience recurrence. Interestingly, when similar analysis was performed on the ER^+ subgroup of the microarray training data set, no signature with high prognostic accuracy could be identified. A Kaplan-Meier plot illustrates survival analysis of untreated, ER^- and node-negative breast cancer cohorts classified as metastatic or nonmetastatic by the 31-gene signature. Metastatic patients had significantly shorter relapse-free survival time ($p < 0.00001$; metastatic versus nonmetastatic mean/median: 938 versus 3,493 d/440 versus 3,688 d) and distant metastasis-free survival ($p < 0.00001$; metastatic versus nonmetastatic mean/median: 1,630 versus 3,640/743 versus 4,124 d) compared with those classified by the signature as nonmetastatic (Figure 7C). Relapse-free survival was defined as the interval between the date of breast surgery and the date of diagnosis of any local, regional or distant relapse. Furthermore, multivariate Cox regression analysis using clinical parameters including age, tumor size, tumor grade and 31-gene signature class (metastatic/nonmetastatic) revealed that only the signature class significantly correlated with distant metastasis-free survival (hazard regression 0.19; $p < 0.00001$).

The 31-gene signature includes some of the genes/proteins we had already biochemically validated, such as LSR and ESA. These markers were originally selected from the 599-gene expression list on the basis of reports in the literature indicating their potential role in cancer progression.

DISCUSSION

The CSC theory of tumorigenesis may have major implications for clinical treatment practice. This potential may be particularly significant for triple-negative or basal-like breast cancers, the focus of this study, since chemotherapy is the only medical treatment for this subtype; the known targeted treatments, such as endocrine and anti-HER2 therapy, have no effect in this patient group. Considering that CSCs are generally more resistant to standard chemotherapy, it is not surprising that chemotherapy is of limited use, and this subgroup has a poor outcome.

Evidence of the presence of CSCs in primary tumors has mainly relied on primary and early passage xenograft models (25,45), whereas a limited number of cell lines, some generated by transform-

ing normal mammary cells with oncogenes, have been used to study stem cell characteristics relative to functional and phenotypic differences and to isolate subpopulations with CSC characteristics based on markers such as $CD44^{hi}$ and $CD24^{-/lo}$, ALDH1 and ESA (1,8,46). Examination at the single-cell level to identify novel markers and functionally characterize CSCs is advantageous, although technically difficult.

In this study, we isolated such single-cell clones from an established cell line and performed detailed characterization. We found that the $CD44^{hi}/CD24^{lo}$ A4 cells resembling cells of the previously described basal A cluster (32), which closely matches the Perou basal–like signature (31,32,47,48), could form mammospheres, which could be propagated. These cells could also initiate tumors in mice when as few as 1,000 cells were implanted into the mammary fat pad, and these tumors recapitulated the phenotype of the parental cell line. A4 cells showed metastatic potential in the lungs and also exhibited resistance to standard chemotherapeutic drugs, all hallmarks of breast CSCs. Surprisingly, whereas earlier models for breast CSCs have shown that these cells exhibit a mesenchymal-like phenotype (31), the A4 cells exhibited a clear epithelial-like phenotype both with regard to morphology and molecular markers. In contrast, the $CD44^{hi}/CD24^{-}$ G4 clone, which exhibited a mesenchymal-like phenotype and resembles the previously described basal B cells, could not form mammospheres, tumors or metastasis, even when 10^6 cells were implanted, and were less resistant to traditional chemotherapeutic drugs. However, it should be mentioned that at least one cell line with basal B characteristics, MDA-MB-231 (32,43), was shown to be tumorigenic in immunodeficient mice (49,50); however, MDA-MB-231 cells are derived from a pleural effusion, whereas HMT3909S13 is derived from a primary tumor, which may contribute to the different tumorigenic behavior. It is of interest that primary tumors established by A4, in contrast to A4 cells in culture, expressed the mesenchymal marker vimentin, indicating that EMT had occurred during the process of establishing the tumor. We presume that not all tumor-initiating cells in this model exhibit an epithelial-like phenotype, as suggested by the heterogeneity of the epithelial-like/mesenchymal-like morphology and CD44/CD24 expression of the $CD44^{hi}/CD24^{lo}$ single-cell clones isolated from HMT3909S13, but it demonstrates that it is not a prerequisite for breast CSC to exhibit a mesenchymal-like morphology. The transition of cells observed *in vivo* also supports the observations made by Gupta *et al.* (16), who showed that cells are in transition between states to obtain phenotypic equilibrium in populations of cancer cells. Further, a recent study by Kim *et al.* (51) demonstrated that, within a single tumor, there are multiple stemlike cells with tumorigenic potential.

ALDH1 and ESA have been suggested as additional CSC markers in breast cancer (1,9,46,52,53). Interestingly, our transcriptomic analysis revealed higher expression of ALDH1 and ESA in A4 versus G4. Further, our quantitative proteomic and transcriptomic analysis revealed a panel of additional markers that are more highly expressed in A4 versus G4, such as MUC1, S100A14 and LSR, proteins that have potential as novel markers and/or targets of CSC subpopulations. These markers were also more highly expressed in basal A cell lines (BT20 and MDA-MB-468) than basal B cell lines (Hs578T), as shown by ICC. Markers such as vimentin and ZEB1 showed higher expression in G4 and basal B cell lines compared with A4 and basal A cell lines, verifying that the data obtained by our bilineage model could be confirmed in other model systems. The basal A cell lines, such as A4, are tumorigenic in athymic nude mice (43), whereas the basal B cell lines and G4 are nontumorigenic in NOG mice.

Despite the lack of ESA, E-cadherin and CD24 expression in the HMT3909S13 cell line *in vitro*, high expression of all three proteins were obtained *in vivo* after inoculation, indicating a small CSC-like population within HMT3909S13 harboring features similar to A4.

On the basis of the genes that exhibit altered expression between A4 and G4, we identified a 31-gene prognostic signature able to predict distant metastases in triple-negative and lymph node–negative breast cancer patients. The accuracy of this profile was confirmed in a second cohort of triple-negative breast cancer patients, but was found not to be the case in ER^{+} breast cancer patients, indicating that CSCs differ between breast cancer subtypes. Thus, the prognostic profiles should be subtype specific. The specificity of 97% is substantially higher than generally reported (54,55) and demonstrates that this signature is a highly accurate indication of those patients with triple-negative breast cancer who will not experience recurrence. This assay may, when further developed, be clinically useful in avoiding the deleterious side effects of overtreatment. Some of the genes included in the signature, such as ESA and LSR, also show promise as protein biomarkers of cells with CSC features, either alone or in combination, and may be used as future targeting of this subpopulation of cells involved in tumorigenesis and chemoresistance. As expected for a CSC, A4 was more resistant to doxorubicin, paclitaxel and methotrexate relative to G4. In agreement with earlier studies, A4 was less resistant to salinomycin, a polyether antibiotic that acts in different biological membranes as an ionophore and that was recently shown to selectively deplete breast CSCs from mammospheres and inhibit breast tumor growth in mice by as yet unknown mechanisms (44,56).

CONCLUSION

We propose that the A4 and G4 single-cell clones represent subgroups within a $CD44^{hi}/CD24^{-/lo}$ population in which only epithelial-like A4 cells contain the traditional described CSC properties. The clones give rise to a 31-gene signature identifying ER^{-} cancer patients who are not likely to develop dis-

tant metastasis. The two cell types also seem to represent the previously described basal A and basal B subtypes. The study challenges the current view that breast CSCs only exhibit a mesenchymal-like phenotype, since we demonstrate that cells with different phenotypes have CSC-like properties. The clones were propagated discontinuously for more than 2 years without changing characteristics. Thus, these single-cell clones may represent a unique model for CSC marker discovery, some of which have been identified here. In addition, the single-cell clones may be useful for evaluation of the efficacy of novel combinatorial drugs targeting different cell types as well as for elucidating the biology of CSCs.

ACKNOWLEDGMENTS

We thank B Preiss, Department of Pathology, Odense University Hospital, for cytogenetic analysis of the cell lines; Eva C Nielsen for expert assistance in quantitative proteomics experiments; and MK Occhipinti-Bender for editorial assistance.

This work was supported by grants from the Danish Cancer Society, Danish Research Council, A Race Against Breast Cancer, The Novo Nordisk Foundation, The Simon Spies Foundation and the Sino-Danish Breast Cancer Research Centre.

R Leth-Larsen and HJ Ditzel planned the experimental design, analyzed data and wrote the manuscript. D Elias performed the analysis of microarray data. OW Petersen and ON Jensen were involved in the experimental design. MG Terp, T Kühlwein and AG Christensen conducted the experiments. The accession number of microarray data set was GSE32455.

DISCLOSURE

The authors declare that they have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

REFERENCES

1. Charafe-Jauffret E, *et al.* (2009) Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer Res.* 69:1302–13.
2. Trumpp A, Wiestler OD. (2008) Mechanisms of disease: cancer stem cells—targeting the evil twin. *Nat. Clin. Pract. Oncol.* 5:337–47.
3. Jordan CT, Guzman ML, Noble M. (2006) Cancer stem cells. *N. Engl. J. Med.* 355:1253–61.
4. Lobo NA, Shimono Y, Qian D, Clarke MF. (2007) The biology of cancer stem cells. *Annu. Rev. Cell Dev. Biol.* 23:675–99.
5. Gupta PB, Chaffer CL, Weinberg RA. (2009) Cancer stem cells: mirage or reality? *Nat. Med.* 15:1010–2.
6. Dave B, Chang J. (2009) Treatment resistance in stem cells and breast cancer. *J. Mammary Gland Biol. Neoplasia.* 14:79–82.
7. Liu S, Wicha MS. (2010) Targeting breast cancer stem cells. *J. Clin. Oncol.* 28:4006–12.
8. Fillmore CM, Kuperwasser C. (2008) Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res.* 10:R25.
9. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. (2003) Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 100:3983–8.
10. Dean M, Fojo T, Bates S. (2005) Tumour stem cells and drug resistance. *Nat. Rev. Cancer.* 5:275–84.
11. Diehn M, *et al.* (2009) Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature.* 458:780–3.
12. Diehn M, Clarke MF. (2006) Cancer stem cells and radiotherapy: new insights into tumor radioresistance. *J. Natl. Cancer Inst.* 98:1755–7.
13. Baumann M, Krause M, Thames H, Trott K, Zips D. (2009) Cancer stem cells and radiotherapy. *Int. J. Radiat. Biol.* 85:391–402.
14. Li X, *et al.* (2008) Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J. Natl. Cancer Inst.* 100:672–9.
15. Takebe N, Ivy SP. (2010) Controversies in cancer stem cells: targeting embryonic signaling pathways. *Clin. Cancer Res.* 16:3106–12.
16. Gupta PB, *et al.* (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell.* 146:633–44.
17. Chaffer CL, *et al.* (2011) Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc. Natl. Acad. Sci. U. S. A.* 108:7950–5.
18. Scheel C, Weinberg RA. (2011) Phenotypic plasticity and epithelial-mesenchymal transitions in cancer and normal stem cells? *Int. J. Cancer.* 129:2310–4.
19. Yang YM, Chang JW. (2008) Current status and issues in cancer stem cell study. *Cancer Invest.* 26:741–55.
20. Ricci-Vitiani L, *et al.* (2007) Identification and expansion of human colon-cancer-initiating cells. *Nature.* 445:111–5.
21. Singh SK, *et al.* (2004) Identification of human brain tumour initiating cells. *Nature.* 432:396–401.
22. Creighton CJ, *et al.* (2009) Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. U. S. A.* 106:13820–5.
23. Tanei T, *et al.* (2009) Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin. Cancer Res.* 15:4234–41.
24. Ginestier C, *et al.* (2010) CXCR1 blockade selectively targets human breast cancer stem cells in vitro and in xenografts. *J. Clin. Invest.* 120:485–97.
25. Ginestier C, *et al.* (2007) ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell.* 1:555–67.
26. Honeth G, *et al.* (2008) The CD44+/CD24– phenotype is enriched in basal-like breast tumors. *Breast Cancer Res.* 10:R53.
27. Mani SA, *et al.* (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell.* 133:704–15.
28. Morel AP, *et al.* (2008) Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One.* 3:e2888.
29. Hugo H, *et al.* (2007) Epithelial—mesenchymal and mesenchymal—epithelial transitions in carcinoma progression. *J. Cell Physiol.* 213:374–83.
30. Santisteban M, *et al.* (2009) Immune-induced epithelial to mesenchymal transition in vivo generates breast cancer stem cells. *Cancer Res.* 69:2887–95.
31. Blick T, *et al.* (2010) Epithelial mesenchymal transition traits in human breast cancer cell lines parallel the CD44(hi/)CD24 (lo/-) stem cell phenotype in human breast cancer. *J. Mammary Gland Biol. Neoplasia.* 15:235–52.
32. Neve RM, *et al.* (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell.* 10:515–27.
33. Petersen OW, *et al.* (1990) Differential tumorigenicity of two autologous human breast carcinoma cell lines, HMT-3909S1 and HMT-3909S8, established in serum-free medium. *Cancer Res.* 50:1257–70.
34. Brothman AR, Persons DL, Shaffer LG. (2009) Nomenclature evolution: changes in the ISCN from the 2005 to the 2009 edition. *Cytogenet. Genome Res.* 127:1–4.
35. Ong SE, Foster LJ, Mann M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods.* 29:124–30.
36. Ong SE, *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics.* 1:376–386.
37. Lund R, Leth-Larsen R, Jensen ON, Ditzel HJ. (2009) Efficient isolation and quantitative proteomic analysis of cancer cell plasma membrane proteins for identification of metastasis-associated cell surface markers. *J. Proteome Res.* 8:3078–90.
38. Hill A, *et al.* (2006) Cortactin underpins CD44-

promoted invasion and adhesion of breast cancer cells to bone marrow endothelial cells. *Oncogene.* 25:6079–91.
39. Desmedt C, *et al.* (2007) 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.* 13:3207–14.
40. Sotiriou C, *et al.* (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* 98:262–72.
41. Sheridan C, *et al.* (2006) CD44+/CD24– breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res.* 8:R59.
42. Meyer MJ, *et al.* (2009) Dynamic regulation of CD24 and the invasive, CD44posCD24neg phenotype in breast cancer cell lines. *Breast Cancer Res.* 11:R82.
43. Lacroix M, Leclercq G. (2004) Relevance of breast cancer cell lines as models for breast tumours: an update. *Breast Cancer Res. Treat.* 83:249–89.
44. Gupta PB, *et al.* (2009) Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell.* 138:645–59.
45. Al-Hajj M. (2007) Cancer stem cells and oncology therapeutics. *Curr. Opin. Oncol.* 19:61–4.
46. Wright MH, *et al.* (2008) Brca1 breast tumors contain distinct CD44+/CD24– and CD133+ cells with cancer stem cell characteristics. *Breast Cancer Res.* 10:R10.
47. Perou CM, *et al.* (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. U. S. A.* 96:9212–7.
48. Chung CH, Bernard PS, Perou CM. (2002) Molecular portraits and the family tree of cancer. *Nat. Genet.* 32 (Suppl.):533–40.
49. Minn AJ, *et al.* (2005) Genes that mediate breast cancer metastasis to lung. *Nature.* 436:518–24.
50. Wang Y, *et al.* (2006) Adiponectin modulates the glycogen synthase kinase-3beta/beta-catenin signaling pathway and attenuates mammary tumorigenesis of MDA-MB-231 cells in nude mice. *Cancer Res.* 66:11462–70.
51. Kim J, *et al.* (2012) Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity. *Proc. Natl. Acad. Sci. U. S. A.* 109:6124–9.
52. Croker AK, *et al.* (2009) High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. *J. Cell Mol. Med.* 13:2236–52.
53. Liu Q, Li JG, Zheng XY, Jin F, Dong HT. (2009) Expression of CD133, PAX2, ESA, and GPR30 in invasive ductal breast carcinomas. *Chin. Med. J. (Engl.).* 122:2763–9.
54. Naoi Y, *et al.* (2011) Prediction of pathologic complete response to sequential paclitaxel and 5-fluorouracil/epirubicin/cyclophosphamide therapy using a 70-gene classifier for breast cancers. *Cancer.* 117:3682–90.
55. Kuderer NM, Lyman GH. (2009) Gene expression profile assays as predictors of distant recurrence-free survival in early-stage breast cancer. *Cancer Invest.* 27:885–90.
56. Fuchs D, Heinold A, Opelz G, Daniel V, Naujokat C. (2009) Salinomycin induces apoptosis and overcomes apoptosis resistance in human cancer cells. *Biochem. Biophys. Res. Commun.* 390:743–9.