ALDH1-positive cancer stem cells predict engraftment of primary breast
tumors and are governed by a common stem cell program

Running title: Clinical relevance of breast CSCs from PDXs

Emmanuelle Charafe-Jauffret1,2,3, Christophe Ginestier1, François Bertucci1,3,4, Olivier Cabaud1, Julien Wicinski1, Pascal Finetti1, Emmanuelle Josselin1, José Adelaide1, Tien-Tuan Nguyen1, Florence Monville1, Jocelyne Jacquemier2, Jeanne Thomassin-Plana2, Guillaume Pinna7, Aurélie Jalaguier6, Eric Lambaudie5, Gilles Houvenaeghel3,5, Luc Xerri2,3, Annick Harel-Bellan7, Max Chaffanet1, Patrice Viens1,3,4, Daniel Birnbaum1

Affiliations:
1 Inserm, U1068, CRCM, Molecular Oncology, Marseille, F-13009, France;
2 Institut Paoli-Calmettes, Biopathology, Marseille, F-13009, France;
3 Aix-Marseille Univ, UM 105, F-13284, Marseille, France
4 Département d’Oncologie Médicale, Institut Paoli-Calmettes, Marseille, France
5 Département de Chirurgie oncologique Institut Paoli-Calmettes, Marseille, France
6 Département de Radiologie, Institut Paoli-Calmettes, Marseille, France
7 Plateforme ARN interference PArI, CEA SACLAY, Gif-sur-Yvette, France

Keywords: breast cancer-cancer stem cells- primary tumors derived-xenograft-stemness-clinical validation
This work is supported by INSERM, Institut Paoli-Calmettes, Institut National du Cancer-DGOS grant TRANSLA11-103 (ECJ), Ligue Nationale Contre le Cancer Label DB (DB).

Corresponding Author: Emmanuelle Charafe-Jauffret, Centre de Recherche en Cancérologie de Marseille (CRCM), départements d’Oncologie Moléculaire et de Pathologie, 232, Bd de Ste Marguerite, 13009 Marseille, France
Tel:+33 4 91 22 34 57/Fax :+33 4 91 22 35 44

There is no Conflict of Interest to disclose for any author of the manuscript
Abstract

Cancer stem-like cells (CSC) have been widely studied but their clinical relevance has yet to be established in breast cancer (BC). Here we report the establishment of primary breast tumor-derived xenografts (PDXs) that encompass the main diversity of human BC and retain the major clinicopathological features of primary tumors. Successful engraftment was correlated with the presence of ALDH1-positive CSCs which predicted prognosis in patients. The xenografts we developed demonstrated a hierarchical cell organization of BC with the ALDH1-positive CSCs constituting the tumorigenic cell population. Analysis of gene expression from functionally-validated CSCs yielded a breast CSC signature and identified a core transcriptional program of 19 genes shared with murine embryonic, hematopoietic and neural stem cells. This generalized stem cell program allowed the identification of potential CSC regulators, which were related mainly to metabolic processes. Using an siRNA genetic screen designed to target the 19 genes, we validated the functional role of this stem cell program in the regulation of breast CSC biology. Our work offers a proof of the functional importance of CSCs in breast cancer, and it establishes the reliability of PDXs for use in developing personalized CSC therapies for breast cancer patients.
Introduction

Heterogeneity is a limitation to cure breast cancer (BC), overcome recurrences and metastases (1). A major cause of heterogeneity is the hierarchical cellular organization: that endows a small, phenotypically distinct population of cancer stem cells (CSC) with the capacity to support tumor growth, metastasis and therapeutic resistance (2, 3, 4, 5, 6, 7, 8, 9, 10, 11). Multiple pools of stem cells have been identified within a tumor by different means including (12, 13, 14) mouse lineage tracing experiments (15, 16, 17). To measure the capacity of a subset of tumor cells to transfer a cancer, tumorigenesis assays with iso-or xenogenic transplantation have been used for many years. They evaluate both tumor-initiating and self-renewal capacities and are widely used to assess CSC activity. They have demonstrated that leukemia and many solid tumors are organized along a hierarchical model (18, 19, 20). Orthotopic BC xenografts have been shown to keep the features of their corresponding human tumors and to predict prognosis in patients (21). Yet, the validity of xenotransplant assays for CSC studies has been questioned. The use of highly immunodeficient mice increases the frequency of tumorigenic cells in certain types of cancers such as melanoma (22, 23). Furthermore, this model tests the potential of cells to form tumors, not their actual fate in the tumor in which they are born, regardless of microenvironmental variables. Notwithstanding these issues, the CSC model suggests that focusing attention and therapies on CSCs rather than on non-CSCs may improve prognosis and help cure cancer. However, both the clinical relevance and universality of the CSC model remain to be firmly established in solid tumors. An approach would be to demonstrate that features of CSC but not of non-
CSCs, influence clinical outcome. A recent study has shown emerging evidence supports the idea that properties of leukemic stem cells (LSCs) may be associated with prognosis (24). Here, we report the establishment of a bank of patient-derived xenografts (PDX) from primary breast cancer (xenobank). We found that a xenograft retains the main features of its cognate primary tumor and that successful engraftment is correlated with the presence of CSCs in the primary tumor, and predicts prognosis in patients. We established the gene expression signatures (GES) of functionally validated CSC populations (breast CSC-GES) and tested their clinical relevance in patients with breast cancers. This functionally-validated CSC population is correlated with survival and expresses genes governing stem cell functions, supporting a major prediction of the CSC model and opening further promises for new anti-CSC therapies using valid preclinical models.
Materials and Methods

Ethics statement. Samples of human origin and associated data were obtained from the IPC / CRCM Tumour Bank, that operates under authorization # AC-2007-33 granted by the French Ministry of Research. Prior to scientific use of samples and data, patients were appropriately informed and asked to consent in writing, in compliance with French and European regulations. The project was approved by the IPC Institutional Review Board. Animal studies were approved by the Inserm office for Laboratory Animal Medicine.

Primary tumor samples. All tissue samples were collected prospectively at the Institut Paoli-Calmettes from January 2008 to June 2009. Samples were obtained from fresh entire core biopsies or surgical specimen. Tumoral cells content was evaluated under light microscope: when it reached at least 70% of cells, the sample was selected for study and for one part frozen in liquid nitrogen and for another part directly processed for mice implantation. All the histoclinical informations of 74 different samples collected are listed in Table S1. Patient's treatments followed the standard guidelines of our institute: 56 patients underwent anthracyclin-based chemotherapy, 59 irradiation and 41 received hormone therapy.

Primary tumor processing and implantation. Human breast primary tumors obtained after surgery were dissociated mechanically and enzymatically using collagenase/hyaluronidase (StemCell technologies, Vancouver, Canada) digestion to generate single cell suspension for the in vivo implantation. For each primary tumor
dissociated, 1x10^6 cells were implanted orthotopically in humanized cleared fat pads of NSG (NOD/Shi-scid/IL-2Rγnull) mice for establishing xenotransplants as previously described (19) (Supporting information).

**Flow cytometry analysis.** The analysis was processed on single cell suspension from our patient-derived xenografts (PDXs) obtained as described above. ALDEFLUOR kit (StemCell technologies) was used to isolate the population with high ALDH enzymatic activity as previously described (19). In order to eliminate cells of mouse origin from the PDXs, we used staining with an anti-H2Kd antibody (BD biosciences, 1/200, 20 min on ice) followed by staining with a secondary antibody labeled with phycoerythrin (PE) (Jackson labs, 1/250, 20 min on ice).

**Tumorigenicity assay.** We evaluate the outgrowth potential of each population (ALDEFLUOR-positive, ALDEFLUOR-negative, Unselected) sorted from three different PDXs (CRCM226 x, CRCM174 x, CRCM168 x) and injected in cleared humanized fat pads of NSG mice. We performed serial passages in vivo, using limiting dilutions of ALDEFLUOR-positive, -negative and unselected cells (50,000 cells; 5,000 cells; 3,000 cells; 500 cells; 300 cells). For each PDX model and for each limiting dilutions 10 fat pads were injected. Each tumors generated with 500 sorted cells were serially re-implanted three times and in 10 different fat pads. Each mouse that present a tumor reaching a size of 10mm was considered as a tumor-bearing mouse.
**Immunohistochemistry.** Expression of ALDH1, CD44/CD24, ER, ERBB2, Ki67, PR and P53 was studied by IHC. The characteristics of the antibodies and experimental procedures are described in the Supporting information.

**DNA and RNA extraction.** DNA and RNA were extracted from frozen primary tumor samples, PDXs, and ALDEFLUOR-positive/-negative cells by using the All prep DNA/RNA/Protein Mini Kit (Qiagen, Hilden, Germany) for primary tumors and PDXs, and the RNeasy Micro Kit (Qiagen) for ALDEFLUOR-sorted cells.

**Gene expression analysis.** Gene expression profiling was done with Affymetrix U133 Plus 2.0 human oligonucleotide microarrays. Preparation of cRNA, hybridizations, washes and detection were done as previously described (25). See Supporting information for further detailed description of gene expression array and bioinformatics analysis.

**Array comparative genomic hybridization analysis.** Array-based comparative genomic hybridization (array-CGH) was applied to 75 samples (53 primary tumors and 22 PDXs including 19 early passages and 3 late passages) using high-resolution 244K CGH microarrays (Hu-244A, Agilent Technologies, Massy, France). See Supporting information for further detailed description of gene expression array and bioinformatics analysis.

**siRNA screen and ALDEFLUOR analysis.** We use the SUM159 mesenchymal BCL and S68 luminal BCL to perform the siRNA screen. The cell lines were grown using the recommended culture conditions (26). We have adapted a methodology to
optimize the reverse transfection of SUM159 and S68 cells using lipofectamine RNAiMAX (Invitrogen). Five thousand cells were plated in triplicate in 96-well plates with 0.25 pmol of siRNA in 25ul of Opti-MEM I Medium. The ALDEFLUOR-positive population was evaluated in triplicate at 72-h after lipofection. We have validated that CSCs and nontumorigenic cells were equally transfected (data not shown). We miniaturized the ALDEFLUOR assay for CSC-enriched population detection per cell in 96-wells plate. For each well we used 0.25 ul of substrate (BAAA) and 100 ul of ALDEFLUOR buffer. We used an automated FACS (LSR2) to detect the ALDEFLUOR-positive population in each well. Each experiments were controlled in 6-plicate using cells incubated with 0.5ul of DEAB. The mini siRNA library of 19 genes was designed by Qiagen based on a published library siRNA that have been shown in peer reviewed publication to provide effective gene silencing. We used three siRNAs per selected gene, most likely in triplicates and considered as a “hit” each siRNA construct that induces a variation of the number of ALDEFLUOR-positive cells over a threshold of two-fold the CSC proportion detected in the control (BCLs infected with a MOCK siRNA). Cell viability was evaluated using DAPI staining.

**siRNA screen and SFE analysis.** We have performed the siRNA screen with both BCLs following the reverse transfection protocol described before. Then, 72-h after lipofection, BCLs were dissociated and plated as single cells in ultra-low attachment plates (Corning, Acton, MA, www.sigmaaldrich.com) at low density (1,000 viable cells per milliliter). Cells were grown as previously described (27). After 5 days of suspension culture, the capacity of cells to form tumorspheres (SFE) was quantified.
**Statistical analysis.** Statistical analyses used the SPSS software (version 10.0.5) and are described in Supporting information.
Results

*Generation of patient-derived tumor xenografts (PDXs).* We transplanted 74 fresh primary breast tumors from 71 different individuals into cleared and humanized mammary fat pads of female NSG mice. All implanted tumors were from primary tumor and not from metastasis. The 74 primary tumors implanted cover the main histoclinical features met in an unselected set of breast tumors (*Table S1*). Twenty PDXs were established out of the 74 primary samples transplanted with an engraftment rate of 27%. We successfully maintained 13 PDXs through serial passages (transplantable rate: 17.5%); seven were lost between the first and the fourth passage. The average delay of growth of PDXs was 212 days (range 77-466). PDX growth kinetic (time to reach 10 mm) was stable or accelerated with serial passages (*Figure S1*).

*PDXs maintain the histological and phenotypic features of primary tumors.* We compared histology, SBR grade and pathological phenotype (based on ERBB2, ER, PR, Ki67, P53 protein expression) of PDXs with those of the cognate primary tumors. We observed a striking preservation of the main histological (*Figure 1A-B*) and immunohistochemical features (*Table I*). Among the 20 PDXs, only the expression of hormonal receptors (ER and/or PR) was lost in 4 PDXs out of 11 compared to their corresponding primary tumors. ERBB2 overexpression or “triple-negative” phenotype was maintained in all the PDXs. All the primary tumors that successfully engraft presented a high grade (SBR grade ≥ 2), which was retained in PDXs. Concerning
P53 status, only one discordant case (CRCM274) was observed with a gain of P53 overexpression in the PDX.

**PDXs maintain the genome profile and retain the molecular subtype of cognate primary tumors.** In 18 paired primary tumors/PDXs, we determined the genomic profiles according to DNA copy number alterations (CNAs) (Figure S2). We observed an almost perfect preservation of the genomic profiles in paired PDXs with a correlation score of 0.81 (Figure 1D). Furthermore, late passages in mice (CRCM168 X P11, CRCM226 X P10, CRCM237 X P8) preserved a genomic profile identical to early passages (P0 to P2) (Figure 1D). We profiled 19 paired primary tumors/PDXs using Affymetrix whole-genome oligonucleotide microarrays and determined the molecular subtypes by using the PAM50 predictor. Primary tumors from the five major molecular subtypes successfully engraft with 1 luminal A, 4 luminal B, 4 ERBB2-like, 8 basal, and 2 normal-like. In 14 pairs out of 19, the molecular subtype of the PDXs was identical to that of the corresponding primary tumor (Figure 1C). CRCM214 PT and CRCM274 PT were respectively of luminal B and luminal A subtype, but their paired PDXs lost expression of ER or PR and switched to an ERBB2 subtype, without modification of ERBB2 expression. CRCM184 X and CRCM272 X were respectively of ERBB2 and luminal B subtype, whereas the parental primary tumors were of normal-like subtype; this is likely to be due to the dilution of ERBB2- or ER-overexpressing cancer cells by abundant stromal and normal breast tissue in the human samples (Figure S3). Thus, our model of xenografts generally phenocopied the major molecular features of the corresponding primary tumors, and maintained these features among passages.
**PDXs present a hierarchical organization driven by a CSC population.** To evaluate the presence of a CSC population in our PDXs, we used the ALDEFLUOR assay. We detected an ALDEFLUOR-positive population in all tested xenografts, ranging from 0.2 to 12.3% (average 3.88%) of the total cancer cell population (Figure S4). ALDEFLUOR-positive and -negative cells isolated from three different PDXs were transplanted in limited dilutions in recipient mice. The percentage of tumor-bearing mice decreased with the number of injected ALDEFLUOR-negative cells from 77% with 50,000 ALDEFLUOR-negative cells to 0% with 300 ALDEFLUOR-negative cells, whereas 100% of mice injected with ALDEFLUOR-positive cells developed tumors, even with 300 cells only (Figure 2A). The breast CSC frequency as determined by limiting dilution analysis (LDA) *in vivo* was higher in the ALDEFLUOR-positive cell population (1/1; CI [1-697]) than in the ALDEFLUOR-negative cell population (1/14038; CI [7007-28126]; p=2.13e-18) or the unselected cell population (1/1572; CI [880-2806]; p=3.08e-06) (Table SII). Moreover, tumor growth rates were higher in tumors generated from ALDEFLUOR-positive population than in tumors generated from ALDEFLUOR-negative or unsorted populations (Figure 2B, Figure S5). Furthermore, only ALDEFLUOR-positive cells were able to regenerate both ALDEFLUOR-positive and -negative cells along passages, demonstrating that they are able to self-renew and differentiate (Figure 2C). The tumorigenic potential of the ALDEFLUOR-negative population was lost with passages whereas it was maintained in the ALDEFLUOR-positive population suggesting that the ALDEFLUOR-negative population contains cells that have proliferation potential but lack self-renewal ability (Figure 2D). Thus, our PDXs retained a hierarchical organization with an ALDEFLUOR-positive CSC population that initiates and maintains tumor growth.
Histological grade and ALDH1-positive CSC content of primary tumors predict engraftment. The clinical relevance of PDX models is suggested by the good correlation between success in engraftment and survival of patients (21). In our series we thus first questioned whether engraftment was associated with clinical outcome. Only 4% of primary tumors that failed to engraft versus 34% of primary tumors that did developed metastasis within the first three years of follow-up (log Rank test, p=0.003), (Figure 2E). Moreover, among the different histoclinical parameters associated with MFS, only positive axillary lymph node (pN) and primary tumor engraftment were independent factors associated with MFS (Figure S6). This observation confirms that primary tumor engraftment is strongly correlated with disease evolution. However, the factors that influence tumor engraftment are not known. In our series, engraftment was not correlated with the type of sample (core biopsy or surgical sample) used for primary injection (Figure S7). Among the different histoclinical and molecular factors tested (Figure S8), engraftment was associated with high SBR grade (p=0.0005), absence of ER expression (p=0.005), high proliferation rate (p=0.003), and molecular subtype (P=0.006) (Figure 2F-G and S6). Interestingly, within a molecular subtype, the engraftment rate perfectly matched the clinical outcome associated (Figure 2H). In contrast, no recurrent genomic alteration was associated with engraftment (Figure S9). We next tested whether the expression of CSC markers (ALDH1, CD44+/CD24-) in primary tumors was correlated with engraftment. Out of the 69 primary tumors injected and analyzed for ALDH1 expression, 24 had ALDH1-positive cells. We observed only 5 ALDH1-positive tumors in the 45 (11%) non-engrafters whereas 15 out of the 20 (75%) engrafters were ALDH1-positive (p=7.2e-5) (Figure 2I). In contrast, the presence of
CD44+/CD24- cells in primary tumors was not associated with engraftment (Figure S6). Using multivariate analysis with Cox proportional model, only SBR grade and ALDH1 expression were independent factors associated with successful engraftment (Figure 2J). These results confirm the correlation between patient's outcome, the capacity for a primary tumor to be xenografted, and the presence of ALDH1-positive CSC. They further suggest that the molecular machinery governing CSC properties is likely to influence clinical outcome.

**Transcriptional profiles of CSCs isolated from PDXs.** To identify molecular networks regulating CSC biology we used Affymetrix microarrays to establish the gene expression profiles of sorted ALDEFLUOR-positive and -negative populations from 8 different PDXs. SAM analysis identified 837 genes as differentially expressed (FDR=0.05) - hereafter designated as the breast CSC gene expression signature (BCSC-GES) - between the ALDEFLUOR-positive and -negative populations, (Figure 3A, Table SIII). To determine if genes overexpressed in ALDEFLUOR-positive cells were commonly enriched in known corresponding signaling pathways, we ran a gene ontology analysis (Figure 3B and Table SIV). Among the different pathways identified several play a role in stem cell biology (BRCA1, P53, SHH, and retinoic pathways), supporting the potential stem cell nature of our BCSC-GES. The remaining pathways were associated with two main cell functions (Figure S10 and S11): cell cycle control and especially the G1/S checkpoint controlling a crucial step for cell fate decision (28) and DNA damage repair.

**BCSC-GES is associated with poor prognosis in BC.** We studied the correlation between the BCSC-GES and histoclinical data using 13 clinically annotated gene
expression data sets corresponding to 2,609 patients (Table SV). Based on the BCSC-GES, the samples were classified as “BCSC-like” (1,443 samples) or “non-BCSC-like” (1,196 samples). Correlations were found between these two classes and all histoclinical features except patient’s age and histological type: as compared to “non-BCSC-like” samples, “BCSC-like” samples were more frequently SBR grade 3, with a TNBC (triple negative breast cancer) phenotype and axillary lymph node metastasis (Table SVI). The prognostic value regarding metastasis-free survival (MFS) was tested within the 1,642 patients with available follow-up (8 out of 13 datasets). The 5-year MFS was 70% (95%CI, 67-74) in the “BCSC-like” group and 80% (95%CI 77-83) in the “non-BCSC-like” group (p=5.5e-04, log-rank test) (Figure 3C). In multivariate analysis, all these variables remained significant, including the BCSC-GES-based classification (p = 0.027; hazard ratio (HR= 1.39; 95%CI [1.04-1.87]) (Figure 3D), suggesting that our BCSC-GES determined in our PDXs was independently associated with prognosis in primary BC.

Identification of a “common” stem cell core transcriptional program. To gain insight into the molecular mechanisms that govern stem cell intrinsic functions, we searched for common expression patterns between our 837-gene BCSC-GES and three stemness GES defined from embryonic, neural, and hematopoietic stem cells (29). Enrichment test was significant with each of the tested stemness GES and the core stemness GES (3SC) defined by Ramalho-Santos et al. (Table SVII). By contrast, no enrichment was found between our BCSC-GES and the 65-gene signature identified to be upregulated in differentiated cells. A total of 19 genes were common to the four stem cell signatures (BCSC, ESC, HSC,NSC) (Figure 3E). These leading genes were called the core-enriched 4SC genes (CE-4SC). We then...
derived a metagene classifier based on the combined expression of these 19 genes and determined its prognostic relevance in our public data set of 1,642 BC patients. Based on the metagene, two classes of patients were defined and associated with different MFS (Figure S12): the CE-4SC-positive class (733 patients) displayed a 67% 5-year MFS (95%CI [64-71]) and the CE-4SC-negative class (909 patients) displayed an 81% 5-year MFS (95%CI [78-83]; p=4.8E-06, log-rank test). In multivariate analysis, the CE-4SC signature kept its independent prognostic value (Figure S12). Collectively, these data indicate that our “common” stem cell transcriptional program influences patient outcome and represent a short list of genes to identify key CSC regulators.

The “common” stem cell core transcriptional program functionally regulates CSC self-renewal and differentiation. Overall, the 19 CE-4SC genes are poorly characterized but might be preferentially active in CSCs as compared with the bulk of tumor cells and represent candidates for CSC targeting. Using the SUM159 mesenchymal breast cancer cell line (BCL), we performed a screening of an RNA-interference library specially designed to block the expression of these 19 genes. We used three independent siRNA sequences per gene (57 siRNA constructs) and evaluated the effect of single gene knock-down (KD) on the breast CSC population using a miniaturized ALDEFLUOR assay in 96-well plates. Among the 19 genes, the knock-down of 14 genes significantly modified the ratio CSC/non-CSC (Figure 4A). Within these 14 genes, two groups were identified. In the first group of 5 genes (ACAT2, CBR3, DTYMK, NDUFAF1, PHTF2), the CSC population decreased when gene expression was knocked-down. Conversely, in the second group of 10 genes (ACAT2, FCF1, GFER, GCLM, KLHL7, MDFIC, MRPS10, MSH2, RAD23B, UMPS),
the CSC population significantly increased. To validate this observation we tested the
effect of the 57 siRNAs on tumorsphere-formation efficiency (SFE). SFE is an *in vitro*
assay used as a surrogate method to evaluate the proportion of CSCs. We observed
a strong correlation between modification of the ALDEFLUOR-positive/ALDEFLUOR-
negative ratio and the SFE after gene KD ($\rho = 0.826; \text{CI} [0.72-0.89]; \text{p-value}= 4.4\text{e}-15$) *(Figure 4B)*. To confirm these results, we screened our siRNA library using the
luminal BCL S68. We observed similar results in the two BCLs analyzed (p=0.003),
with 13 out 19 genes with comparable effect after the gene KD on the ALDEFLUOR-
positive population *(Figure S13)*. We never observed an opposite effect of a specific
gene KD in the two BCLs and only five genes (*EIF4EBP1, FCF1, GCLM, KLHL7, SEC23IP*) presented a restricted effect in one BCL when gene expression was KD.
The results on the ALDEFLUOR-phenotype obtained in S68 were similarly validated
using SFE assay ($\rho = 0.79; \text{CI} [0.61-0.87]; \text{p-value}= 6.3\text{e}-10$) *(Figure S14)*. These
results represent a first validation of the functional role of our CE-4SC signature in
the regulation of breast CSC biology with genes that may be implicated either in self-
renewal program or in differentiation process.
Discussion

We established a bank of serially transplantable, orthotopic xenografts of primary BCs in humanized mouse fat pad. The PDXs maintained the main histological, phenotypical and molecular features of the cognate primary tumors as demonstrated before in other PDX models (21, 30). We show that the xenografts maintain the same hierarchical organization as the primary tumors, with CSCs able to self-renew and differentiate. Our findings have three important implications for BC understanding and treatment.

First, our study contributes to the definition of CSCs. As pointed out in the report of “The Year 2011 Working Conference on CSCs”, despite the deep clinical implications of the CSC model, CSC investigation has been hampered both by a lack of consistency in the terms used for these cells and by how they are defined (31). Whereas the phenotypic definition (CD44+/CD24-/low or ALDEFLUOR-positive cells for breast CSCs) can restrain the population to a subset of CSCs, a functional definition based on tumor-initiating or tumor-maintaining properties seems more appropriate. Xenograft assays outlined a population of cells in an operational rather than in a conceptual context, based on tumor-initiating or tumormaintaining properties, recapitulation of the entire tumoral heterogeneity. The paradigm of xenograft assay as a gold standard to evaluate CSC activity has also been challenged since the ability of a subpopulation of tumor cells to generate a tumor in the mouse might reflect the permissivity of these cells for a specific environment (32, 22). In that context, we questioned the clinical relevance of the xenograft assay. We demonstrated that primary breast tumors that generate PDXs have more ALDH1-positive cells (CSCs) and are more prone to form metastasis in patients than tumors
that failed to graft. We demonstrated that tumor engraftment is an independent prognostic factor, as strong as lymph node invasion to predict metastasis. Our study establishes a direct link between primary tumor engraftment, functionally-defined breast CSCs and patient’s clinical outcome suggesting that CSCs are not artefacts of the experimental PDX model but reflect intrinsic tumor biology.

Second, we show that the molecular machinery that governs CSCs is clinically relevant. This crucial issue has been demonstrated in acute myeloid leukemia, where a functionally-defined LSC or HSC signature is correlated with disease outcome (24). In breast, an “invasiveness” gene signature generated by comparing CD44+/CD24-low BC cells with that of normal breast epithelium is associated with disease outcome (33) and a CD44+ cells-signature that contains stem cell markers is correlated with decreased survival in patients (34). Furthermore, a normal mammary stem cell signature predicts BC biological and molecular features, and high grade tumors contain more CSCs than low grade tumors (35). Yet, the potential link between functionally-defined CSCs and engraftment has never been described so far. This approach provides a paradigm for assessing both the identity and clinical relevance of CSCs from BC and other solid tumors. Importantly, because our finding support CSC clinical relevance, it suggests that therapies targeting CSCs would improve survival outcome and that xenograft models based on primary breast CSC engraftment can be crucial in the preclinical evaluation of new cancer drugs.

Third, the identification of shared transcriptional profiles in CSCs and ESC/HSC/NSC suggests that common genes have a role in establishing and maintaining the stem cell state, and that these shared determinants of stemness influence clinical outcome. We identified a “common” stem cell core transcriptional program. This core of genes common to four stem cell gene expression signatures contains genes
implicated in oxidative phosphorylation, detoxification, lipid metabolism, and genomic stability. Some of these genes encode drug resistance enzymes (UMPS) (36), or cell cycle regulators through P27 function (GFER) (37, 38). Metabolic alterations in tumor cells are thought to be important for the tumor phenotype and evolution, and to affect response to therapy (39). In embryonic, neural, normal and tumoral breast tissues, stem/progenitor cells contain lower level of Reactive Oxygen Species (ROS) than their more mature progeny, sustaining their relative resistance to radiation (40, 41). In leukemia, oxidative phosphorylation triggers genomic instability and is a putative mechanism explaining relapse after tyrosine kinase inhibitor treatment (42, 43). Noteworthy, aldehyde dehydrogenase, a hallmark of embryonic, normal adult tissue and cancer stem cells oxidizes aldehydes to the corresponding carboxylic acids (44). Furthermore, we recently identified the importance of mevalonate metabolism in the P27-dependent regulation of breast CSCs (45). Our findings reinforce recent literature data and suggest that metabolic therapies anti-CSC might be a new therapeutic approach to cure cancer. These novel results however need further validation. I

In the opening era of personalized medicine, targeting the cell subpopulation that sustains tumor growth and development could help design new therapeutic strategies to cure cancer. In BC, utilization of PDXs that mimic human tumor is increasing to capture tumor heterogeneity and elaborate the best therapeutic strategy for patients (46, 47). The clinical relevance of PDXs studies for CSCs in BC suggests that this model is reliable to test anti-CSC therapies. If confirmed, our data indicate the importance of developing CSC biomarkers and will move forward a new era of accelerated transfer for personalized cancer therapy into clinics.
Acknowledgments. We thank CRCM Centre de Ressources Biologiques CRB (C Chabannon) for tissue conservation and providing, CRCM animal core facility for animal housing, CRCM flow cytometry core, Xentech company, Evry France (JG Judde) for helpful technical discussions. We are grateful to J-P Borg for discussions and encouragements and to CIML flow cytometry core (M. Malissen and P. Grenot) for the technical assistance. We acknowledge the support of FEDER for the purchase of the ARIA III cell sorter. This work is dedicated to the memory of Rémy Galindo.
Reference List


Figure Legends

Figure 1. PDXs resemble the primary tumors from which they are derived. A-B. Histology of two paired primary tumors/PDXs: CRCM181 a triple negative metaplastic breast cancer (A) and CRCM224 a luminal B breast cancer (B). Histology evaluated after H&E staining and immunohistochemical staining for ER, PR and ERBB2 protein expression are strictly conserved in PDXs. C. Correlation of gene expression of the 19 paired primay tumors (top)/PDXs (bottom) to five molecular subtypes defined by the PAM50 predictor. For each subtype a median profile is defined and compared to the sample's median profile. Correlation coefficients are plotted by colors indicating the subtype: dark blue for luminal A, light blue for luminal B, red for basal, pink for ERBB2-like, and green for normal-like. The dashed line represents the threshold for correlation coefficient significance (0.15). D. Correlation matrix evaluating the genomic profile similarity between primary tumors and PDXs. Each sample (19 pairs and 3 late passage PDXs) are classified according to their genomic profile similarities (Spearman test). Correlation coefficient are depicted according to the color scale, with blue for a correlation coefficient null meaning an absence of correlation, and red for a correlation coefficient equal to 1 meaning a perfect correlation.

Figure 2. ALDH1-CSCs drive tumorigenicity in PDXs and predict primary tumor engraftment. A-D The outgrowth potential of sorted populations (ALDEFLUOR-positive, A+; ALDEFLUOR-negative, A-; Unselected, U) from three independent PDXs (CRCM226 x, CRCM168 x, CRCM174 x) was evaluated, revealing a limited potential of the ALDEFLUOR-negative population whereas ALDEFLUOR-positive
cells generated tumors in all fat pads injected even with only 300 cells (A). An example of outgrowth kinetic is represented in B, for CRCM226 x, and an injection of 300 sorted cells. Outgrowth kinetics for all PDX models and all limited dilutions are represented in Figure S5. For each PDX model tested the ALDEFLUOR-positive population recreated the native cellular heterogeneity and gave rise to ALDEFLUOR-positive and -negative cells, as shown in C for CRCM168 x (similar results were observed for CRCM226 x and CRCM174 x). The outgrowth potential of the ALDEFLUOR-positive population was maintained within three serial passages whereas it was lost at passage II for the ALDEFLUOR-negative population (D). Data represent mean ± SD. E. Successful engraftment of primary tumor specimens predicts metastasis formation in patients (LogRank test, P=0.00352). A Kaplan-Meier metastasis-free survival analysis shows patient’s outcome in engrafter and non-engrafter groups. F-I. Factors predicting successful engraftment in mice. Kaplan-Meier analysis shows proportion of tumor outgrowth for different groups of primary tumor injected and stratified according to SBR grade, ER protein expression, molecular subtypes, or the presence of ALDH1-positive cells. Each of these factors identifies different groups of primary tumor with opposite engraftment rate. J. Multivariate analysis with Cox proportional model identifies only SBR grade and ALDH1 expression as independent factors associated with successful engraftment.

Figure 3. Identification of a “common” stem cell core transcriptional program (CE-4SC) associated with clinical outcome. A. Supervised analysis identified a breast CSC gene expression signature (BCSC-GES). Gene expression profiles of ALDEFLUOR-positive and -negative populations, isolated from five PDXs, were compared: 837 genes were identified as differentially expressed between CSCs and
non-CSCs and are represented on hierarchical clustering. **B.** To determine the pathways associated with the BCSC-GES we ran Ingenuity Pathway Analysis (IPA) using Ingenuity® software. Bar plot represents enrichment for each of the network components identified, where the strength of the association is represented by the –log (p-value). **C.** Kaplan-Meier metastasis-free survival curves according to BCSC-GES status. Tumors that express the BCSC-GES (BCSC-like) are associated with a reduced metastasis-free survival (p=5.10-4). **D.** Multivariate analysis with Cox proportional model identified BCSC-GES as independent factor associated with metastasis-free survival. **E.** Identification of a "universal" stem cell core transcriptional program corresponding to the common 19 genes of four stemness GES. The overlap of the four GES is represented with a Venn Diagramm. The 19 common genes (named CE-4SC) are listed.

**Figure 4. Functional validation of the CE-4SC using siRNAs screen in SUM159 BCL. A.** Screening of a library of 57 siRNAs targeting the 19 identified genes using the variation of the ALDEFLUOR-positive population as read out. Each siRNA construct that induced a variation in the number of ALDEFLUOR-positive cells over a threshold of two-fold the CSC proportion detected in the control was considered as a hit. Eighteen hits were identified, corresponding to 14 unique genes, with *in green* siRNA constructs that reduced significantly the CSC population and *in red* siRNA constructs that increased significantly the CSC population. **B.** Comparison of CSC variation after gene knock-down by ALDEFLUOR phenotyping (right) and tumorsphere-forming efficiency (SFE) (left) using the 57 siRNAs. Results for each siRNA (from top to bottom) are represented as opposite bars. Correlations are measured using Spearman’s rank correlation (p). Data represent mean ± SD.
<table>
<thead>
<tr>
<th>ID</th>
<th>Histological type</th>
<th>Grade</th>
<th>ER status</th>
<th>PR status</th>
<th>ERBB2 status</th>
<th>Ki 67 status</th>
<th>P53 status</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRCM162 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>negative</td>
<td>positive</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM162 X</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>negative</td>
<td>positive</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM168 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>negative</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM168 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM172 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM172 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM174 PT</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>CRCM174 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>CRCM177 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>CRCM177 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
<td>herogeneous</td>
</tr>
<tr>
<td>CRCM180 PT</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM180 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>CRCM181 PT</td>
<td>MED</td>
<td>2</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM181 X</td>
<td>MED</td>
<td>2</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>CRCM183 PT</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>herogeneous</td>
<td></td>
</tr>
<tr>
<td>CRCM183 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>herogeneous</td>
<td></td>
</tr>
<tr>
<td>CRCM184 PT</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>CRCM184 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Histoclinical data for primary tumors and corresponding PDXs
<table>
<thead>
<tr>
<th>Sample</th>
<th>Type</th>
<th>Grade</th>
<th>ER</th>
<th>PR</th>
<th>Heterogeneity</th>
<th>Sample</th>
<th>Type</th>
<th>Grade</th>
<th>ER</th>
<th>PR</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRCM214 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>negative</td>
<td>CRCM214 X</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>CRCM216 PT</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>CRCM216 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>CRCM224 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
<td>CRCM224 X</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>CRCM226 PT</td>
<td>IDC</td>
<td>2</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>CRCM226 X</td>
<td>IDC</td>
<td>2</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>CRCM236 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
<td>CRCM236 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td>CRCM237 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
<td>CRCM237 X</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>CRCM256 PT</td>
<td>CS</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td>CRCM256 X</td>
<td>CS</td>
<td>3</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>CRCM258 PT</td>
<td>IDC</td>
<td>2</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
<td>CRCM258 X</td>
<td>IDC</td>
<td>2</td>
<td>negative</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>CRCM268 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
<td>CRCM268 X</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>CRCM272 PT</td>
<td>IDC</td>
<td>2</td>
<td>positive</td>
<td>negative</td>
<td>negative</td>
<td>CRCM272 X</td>
<td>IDC</td>
<td>2</td>
<td>positive</td>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>CRCM274 PT</td>
<td>IDC</td>
<td>3</td>
<td>positive</td>
<td>positive</td>
<td>negative</td>
<td>CRCM274 X</td>
<td>IDC</td>
<td>3</td>
<td>negative</td>
<td>positive</td>
<td>negative</td>
</tr>
</tbody>
</table>

IDC: Invasive Ductal Carcinoma; MED: Medullary breast carcinoma; CS: Carcinosarcoma; ER: Estrogen Receptor; PR: Progesterone Receptor; Heterogeneous: <70% of positive cells
Figure 1

CD
Table 3: Prognostic Analyses of miRNA Expression Profile in Breast Cancer Cell Lines

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>0.72 [0.54-0.97]</td>
<td>2.9E-02</td>
</tr>
<tr>
<td>pN</td>
<td>1.67 [1.24-2.24]</td>
<td>6.8E-04</td>
</tr>
<tr>
<td>pT</td>
<td>1.38 [1.02-1.86]</td>
<td>3.5E-02</td>
</tr>
<tr>
<td>SBR Grade</td>
<td>1.99 [1.24-3.18]</td>
<td>4.1E-03</td>
</tr>
<tr>
<td>BCSC GES</td>
<td>1.39 [1.04-1.87]</td>
<td>2.7E-02</td>
</tr>
</tbody>
</table>

Figure 3
Figure 4
ALDH1-positive cancer stem cells predict engraftment of primary breast tumors and are governed by a common stem cell program

Emmanuelle Charafe-Jauffret, Christophe Ginestier, Francois Bertucci, et al.

Cancer Res Published OnlineFirst October 18, 2013.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-4704

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2014/01/09/0008-5472.CAN-12-4704.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.