Cell

Biological and Molecular Heterogeneity of Breast Cancers Correlates with Their Cancer Stem Cell Content

Salvatore Pece,[1,2,3,4,*] Daniela Tosoni,[1,3,4] Stefano Confalonieri,[1] Giovanni Mazzarol,[3] Manuela Vecchi,[1] Simona Ronzoni,[3] Loris Bernard,[3] Giuseppe Viale,[2,3] Pier Giuseppe Pelicci,[2,3,*] and Pier Paolo Di Fiore[1,2,3,*]
[1]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Via Adamello 16, 20139 Milan, Italy
[2]Dipartimento di Medicina, Chirurgia ed Odontoiatria, Universita' degli Studi di Milano, Via di Rudini' 8, 20122 Milan, Italy
[3]Istituto Europeo di Oncologia, Via Ripamonti 435, 20141 Milan, Italy
[4]These authors contributed equally to this work
*Correspondence: salvatore.pece@ifom-ieo-campus.it (S.P.), piergiuseppe.pelicci@ifom-ieo-campus.it (P.G.P.), pierpaolo.difiore@ifom-ieo-campus.it (P.P.D.F.)
DOI 10.1016/j.cell.2009.12.007

SUMMARY

Pathways that govern stem cell (SC) function are often subverted in cancer. Here, we report the isolation to near purity of human normal mammary SCs (hNMSCs), from cultured mammospheres, on the basis of their ability to retain the lipophilic dye PKH26 as a consequence of their quiescent nature. PKH26-positive cells possess all the characteristics of hNMSCs. The transcriptional profile of PKH26-positive cells (hNMSC signature) was able to predict biological and molecular features of breast cancers. By using markers of the hNMSC signature, we prospectively isolated SCs from the normal gland and from breast tumors. Poorly differentiated (G3) cancers displayed higher content of prospectively isolated cancer SCs (CSCs) than did well-differentiated (G1) cancers. By comparing G3 and G1 tumors in xenotransplantation experiments, we directly demonstrated that G3s are enriched in CSCs. Our data support the notion that the heterogeneous phenotypical and molecular traits of human breast cancers are a function of their CSC content.

INTRODUCTION

Cancer is frequently characterized by the alteration of pathways that control the homeostasis of normal stem cells (SCs) (Visvader and Lindeman, 2008). The elucidation of the molecular mechanisms that govern normal SC function might, therefore, advance our understanding of tumorigenesis.

In the mammary gland, resident multipotent mammary SCs (MSCs) orchestrate the development of the gland during embryogenesis, and its modifications in postnatal life (Williams and Daniel, 1983). Since MSCs are rare, their purification constitutes a major hurdle to their characterization. A number of approaches, based on the exploitation of MSC surface markers, have allowed the prospective isolation of mouse and human MSCs (Eirew et al., 2008; Liao et al., 2007; Lim et al., 2009; Raouf et al., 2008; Shackleton et al., 2006; Stingl et al., 2006). However, the relative promiscuity of these markers (Carter et al., 1990; Jones et al., 2004; Stingl et al., 1998; Stingl et al., 2006) limits their usefulness when highly purified MSCs are needed.

At the onset of the present study, we devised a strategy to obtain highly pure populations of MSCs, based on their functional, rather than immunophenotypical, characteristics, and relying on the "mammosphere" technology (Dontu et al., 2003). Since MSCs can withstand *anoikis*, they proliferate/differentiate in anchorage-independent conditions, giving rise to clonal spheroids, which can in part recapitulate the mammary morphogenetic program. MSCs, however, constitute less than 1% of all cells in a mammosphere (Dontu et al., 2003). To identify this fraction of cells, we used a lipophilic fluorescent dye, PKH26, which labels relatively quiescent cells within a proliferating population (Huang et al., 1999; Lanzkron et al., 1999). During the growth of a mammosphere, the rare quiescent/slowly dividing MSCs retain PKH26 epifluorescence, while the bulk population, derived from the proliferation of progenitors of the transit-amplifying compartment, progressively lose it by dilution. We were able to purify, by fluorescence-activated cell sorting (FACS), a minority of PKH-positive (PKH^{POS}) cells from human mammospheres to near homogeneity and to show that they represent human normal MSCs (hNMSCs). We then obtained the transcriptional profile of PKH^{POS}/hNMSCs and compared it to that of their immediate progeny, thus identifying a hNMSC signature.

The apparently simple cytoarchitecture of the mammary gland, composed of an internal layer of luminal epithelial cells and an external layer of myoepithelial cells, is difficult to reconcile with the diversity of breast cancer phenotypes (Stingl and Caldas, 2007). This has led to the hypothesis that, despite its morphological simplicity, the mammary gland is functionally complex and molecularly heterogeneous (Stingl and Caldas, 2007). This hypothesis has stimulated much study and debate regarding the cellular origin of breast cancer subtypes, as it affects our ability to predict tumor behavior and responsiveness to therapy

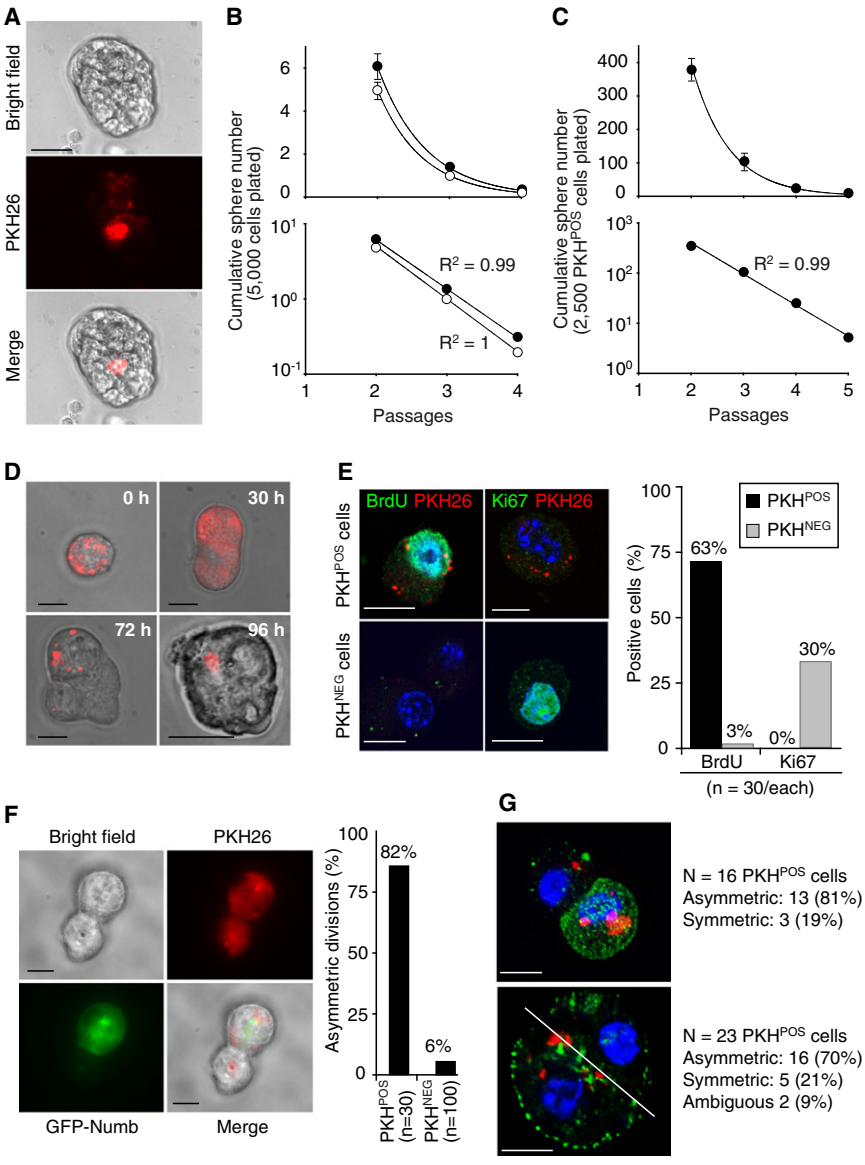

Figure 1. Identification and Purification of hNMSCs

(A) A mammosphere derived from PKH26-labeled epithelial cells. The scale bar represents 100 μm.

(B and C) Propagation of mammospheres for multiple generations starting from bulk epithelial cells (B, open and closed circles indicate two different patients) or FACS-sorted PKH^{POS} cells (C). Top, linear scale; bottom, semilog scale. Regression analysis trend lines and coefficients (R2) are shown. Typical experiments, in triplicate, are shown (additional statistical analyses are in Figures S1B and S1C).

(D) A single PKH^{POS} cell, embedded in methylcellulose, was monitored at the indicated times. Scale bars represent 10 μm, and 100 μm in the bottom-right panel.

(E) BrdU (left) and Ki67 (right) staining of PKH^{POS} and PKH^{NEG} cells. Quantification is shown on the right.

(F and G) Asymmetric segregation of lentivirally transduced GFP-Numb (F) or endogenous Numb (G, confocal sections are shown) to one daughter cell during cell division of a single PKH^{POS} cell either embedded in methylcellulose (F) or grown in suspension (G). Top, mock-treated; bottom, treated with blebbistatin). Scale bars represent 10 μm. The solid white bar in (G) indicates the cleavage plane of the two daughter cells. Quantifications are also shown.

See also Figure S1.

(Stingl and Caldas, 2007). With this background in mind, we used the hNMSC signature to interrogate expression data sets of breast cancers. We discovered that the signature can stratify breast cancers according to their biological and molecular features. We further showed that this is due to the different number of cancer SCs (CSCs) (or cancer-initiating cells) present within different types of breast tumors. Based on our data, we propose a model of breast tumorigenesis that might resolve the incongruence between the complexity of breast cancer phenotypes and the simplicity of the normal breast parenchyma.

RESULTS

Isolation of hNMSCs

Cells from normal human mammary glands were labeled with PKH26 and plated in suspension to allow mammosphere growth (Figure S1A available online). As expected, very few cells within mammospheres retained strong epifluorescence (Figure 1A). The sphere-forming efficiency (SFE) of cells from the mammary gland was 0.003%–0.01% (depending on whether bulk mammary cells or pre-enriched mammary epithelial cells were employed, Figure S1B). The SFE of cells obtained from dissociated mammospheres was ~0.1% (F2 in Figure S1B). Normal mammospheres could be propagated for at least four generations (Figure 1B), and their clonogenic ability decreased exponentially (Figure 1B, Figure S1C): at every generation, the SFE was ~23% of that measured in the preceding generation (Figure S1C). Mammospheres reproducibly contained ~300 cells per sphere (Figure S1D).

We estimated that in the normal mammary gland there is one mammosphere-initiating cell per ~7500 total cells or ~2000 epithelial cells (Figure S1D), in good agreement with previous reports in mouse and human (Shackleton et al., 2006; Stingl et al., 2006). In addition, we calculated that approximately one mammosphere-initiating cell is present per mammosphere, and that this cell is likely found within the PKH26-positive (PKH^{POS}) cell fraction (Figure S1D). This was confirmed by monitoring of the ability of single cells, obtained from dissociated mammospheres, to form second-generation mammospheres, as a function of their PKH positivity. Only PKH^{POS} cells formed

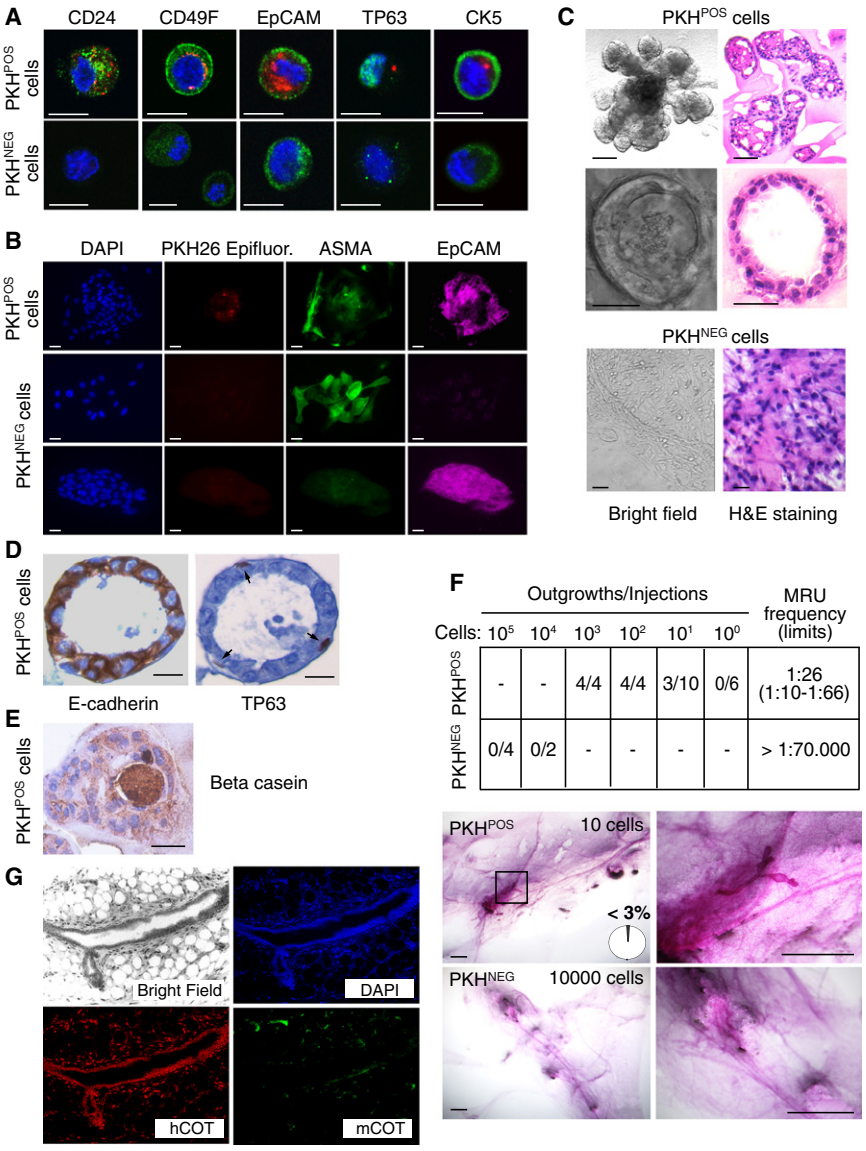

Figure 2. Immunophenotypical and Functional Characterization of PKH^{POS} and PKH^{NEG} Cells

(A) PKH^{POS} or PKH^{NEG} cells were analyzed with the indicated antibodies by IF. Green, antibody staining; red, PKH26 epifluorescence; blue, DAPI. Scale bars represent 10 μm. Quantification is shown in Figure S2A. Results are typical and representative of three independent experiments.

(B) PKH^{POS} or PKH^{NEG} cells were grown on Matrigel and analyzed in IF with the indicated antibodies. Scale bars represent 10 μm. Quantification is shown in Figure S2B. Results are typical and representative of three independent experiments.

(C) PKH^{POS} or PKH^{NEG} cells were grown in 3D in Matrigel (bright field) and then stained with hematoxylin/eosin (H&E). Scale bars represent 100 μm (PKH^{POS}) and 10 μm (PKH^{NEG}).

(D) Serial sections (3 μm thick) of organotypic outgrowths from PKH^{POS} cells were analyzed in IHC with the indicated antibodies. Scale bars represent 50 μm. Arrowheads point to basally located, spindle-like shaped cells that are E-cadherin negative (left) and TP63 positive (right).

(E) IHC analysis of β-casein accumulation in the lumen of 3D organotypic structures stimulated with prolactin. The scale bar represents 50 μm.

(F) Top: The indicated amounts of PKH^{POS} and PKH^{NEG} cells were transplanted into humanized epithelium-cleared mammary fat pads of 21-day-old immunocompromised NOD/SCID mice. MRU, mammary-repopulating unit. Bottom: Whole-mount analysis of 10-week-old mammary outgrowths. Scale bars represent 500 μm. The boxed area is magnified on the right to better visualize the outgrowth. A graphic indication (pie) of the percentage of fat pad filled by the outgrowth is also shown.

(G) Genomic FISH analysis to assess the human origin of the fat pad mammary outgrowths from PKH^{POS} cells, with specific human (hCOT, red) or mouse (mCOT, green) Cot-1 probes (Extended Experimental Procedures). Original magnification ×20.

See also Figure S2.

mammospheres, with a SFE of ~23% (PKH^{POS} pre-FACS in Figure S1B; note that this number is in agreement with the maximum expected SFE, Figures S1B and S1C).

We used FACS to isolate PKH^{POS} cells. Based on the typical PKH26 distribution in mammospheres and the SFE of cells from mammospheres, we isolated the most epifluorescent 0.2%–0.4% of the total cell population (PKH^{POS} cells, gated at 10^3–10^4 fluorescence units), as well as a "dull" population (gated at 10^1–10^2 fluorescence units, PKH^{NEG} cells) (Figure S1A, inset). Only the PKH^{POS} population gave rise to mammospheres (Figure S1A, inset), with a SFE of ~11% (Figure S1B; note that on the basis of this value and a series of normalizations, we estimate that our PKH^{POS} population is ~90% pure, Figure S1B). Of note, PKH^{POS} cells could support at least four additional generations of mammosphere growth (Figure 1C, Figure S1C).

By monitoring microscopically the growth of PKH^{POS} cells embedded in methylcellulose, we established that mammospheres are truly the result of the clonal expansion of single PKH^{POS} cells (Figure 1D). PKH^{POS}, but not PKH^{NEG}, cells showed features of the quiescent state, as evidenced by retention of BrdU and absence of expression of the proliferation marker Ki-67 (Figure 1E). Moreover, PKH^{POS} cells divided asymmetrically, as demonstrated by the unequal partitioning at mitosis of the cell fate determinant Numb (Figures 1F and 1G) (Gönczy, 2008).

Characterization of PKH^{POS} Cells

We analyzed PKH^{POS} cells, by immunofluorescence (IF), with a panel of markers (Experimental Procedures). PKH^{POS} cells shared features with both epithelial (CD24$^+$/EpCAM$^+$) and myoepithelial (CD49F$^+$/CK5$^+$/TP63$^+$) cells (Figure 2A, Figure S2A); however, they did not express detectable terminal differentiation markers, such as Muc1, E-cadherin, and ASMA (data not shown), suggesting that lineage specification has not yet

occurred in these cells. PKH^{NEG} cells, conversely, expressed little or no CD24, CD49F, and TP63; compared to PKH^{POS} cells, they displayed similar levels of EpCAM and lower levels of CK5 (Figure 2A, Figure S2A). Finally, PKH^{NEG} cells did not display detectable levels of Muc1 or ASMA (data not shown).

These findings suggest that, as expected of SCs, PKH^{POS} cells are bipotent. This was confirmed in 2D-differentiation assays on Matrigel. The majority (~87%) of colonies generated by PKH^{POS} cells expressed both epithelial (EpCAM) and myoepithelial (ASMA) markers (Figure 2B, Figure S2B). Within these colonies, individual cells expressed either EpCAM or ASMA, but never both, indicating terminal differentiation of cells into either the epithelial or the myoepithelial lineage (Figure 2B, Figure S2B). By contrast, PKH^{NEG} cells originated either pure epithelial colonies (97%) or, in rarer cases, pure myoepithelial colonies (Figure 2B, Figure S2B). When individual colonies were picked and replated in suspension, bipotent colonies, originating from PKH^{POS} cells, formed mammospheres at high efficiency, whereas lineage-restricted colonies (either epithelial—originating from PKH^{POS} or PKH^{NEG} cells—or myoepithelial—originating from PKH^{NEG} cells) were unable to do so (data not shown).

We also used 3D Matrigel cultures, in which hNMSCs can generate mammary gland organotypic outgrowths (Dontu et al., 2003). PKH^{POS} cells generated two types of outgrowths, which recapitulated several aspects of the mammary gland in vivo: hollow, branched lobuloalveolar, and cavitated, acinar-like structures (Figure 2C). Both epithelial ($E\text{-cadherin}^+$) and myoepithelial ($TP63^+$) cells could be detected in the acinar-like structures (Figure 2D) by immunohistochemistry (IHC). In addition, these structures accumulated β-casein in the lumen, upon prolactin stimulation (Figure 2E). PKH^{NEG} cells did not form 3D structures, but either grew as monolayers of terminally differentiated cells (Figure 2C) or formed small structures that never reached the dimension or acquired the differentiation-specific features detected in the outgrowths generated by PKH^{POS} cells (data not shown).

Finally, we analyzed the ability of PKH^{POS} cells to reconstitute a normal mammary epithelium when transplanted into humanized epithelium-cleared fat pads of immunocompromised NOD/SCID mice. PKH^{POS} cells were able to do so with an efficiency of around 4% (one in 26 cells, range 1:10–66), while PKH^{NEG} cells could not reconstitute the mammary gland, even when injected at concentrations as high as 10^5 cells/transplant (Figure 2F). Of note, this value is compatible with the maximum expected value of reconstitution, estimated on the basis of the replicative kinetics of PKH^{POS} cells (Figure S1E). The outgrowths generated by PKH^{POS} cells displayed the normal mammary gland cytoarchitecture (Figure 2G, Figure S2C) and were derived unequivocally from transplanted human cells (Figure 2G, Figure S2D).

Transcriptomic Analysis of hNMSCs

We performed expression profile analysis of PKH^{POS} and PKH^{NEG} cells. We extracted RNAs from FACS-sorted cells derived from three independent pools of mammospheres, each from five to six individuals. We compared profiles of PKH^{POS} and PKH^{NEG} cells for each pool and obtained three separate gene lists of differentially expressed genes (see the Experimental Procedures). Considering the heterogeneity of the samples (see above), the three lists overlapped significantly (Figure 3A). In particular, 2306 probesets (psets) (the "hNMSC signature") showed a concordant trend of regulation (increased/decreased in two pools, and unchanged in the third one), and significant enrichment ($p < 1 \times 10^{-8}$, Figure 3A). Within this group, 377 psets (the "3/3 signature") were consistently up- or downregulated in all the three pools (expected 8.35, actual 377, enrichment factor 45.11, $p < 1 \times 10^{-60}$, Figure 3A). The genes of the hNMSC signature and of the 3/3 signature are listed in Table S1, together with their functional annotation and other characteristics that will be discussed later (see the Discussion).

The hNMSC signature, which readily distinguished PKH^{POS} from PKH^{NEG} cells (Figure 3B), was validated by testing the expression of 69 transcripts on three independent RNA preparations of PKH^{POS} and PKH^{NEG} cells, with 64% concordance with the GeneChip data (83% if one considers only those genes, 53, for which a reliable trend of regulation could be obtained, Table S1). Similar results were obtained if only transcripts present in the 3/3 signature were considered (Table S1).

We also used IF to validate the expression of selected genes (based on the antibody availability) upregulated in PKH^{POS} (JAG1, SOX4, DNER, and DLL1/DELTA) or PKH^{NEG} (HEY1) cells. JAG1, SOX4, DNER, and DLL1 were significantly enriched in PKH^{POS} cells, whereas HEY1 was enriched in PKH^{NEG} cells (Figures 3C and 3D). Finally, we selected markers (CK5, TP63, and SERPINB5/Maspin, for which antibodies suitable for IHC are available), from those upregulated in PKH^{POS} cells, to test whether they could identify hNMSCs within mammospheres. All three markers were expressed in rare (frequently single) cells within the mammosphere (Figure S3A).

Identification of hNMSCs In Vivo Using Markers from the hNMSC Signature

hNMSCs are thought to reside in the basal layer of ducts and/or at the level of the duct-branch points (Chepko and Smith, 1997; Shackleton et al., 2006; Stingl et al., 2006). We analyzed the expression of markers from the hNMSC signature (DLL1, CK5, CD49F, JAG1, and TOPO2A) by IF (Figure 3E, Figure S3B) or IHC (Figure S3B) on sections of human mammary glands. In IF experiments, we used double labeling with EpCAM, which allows the visualization of the epithelial layer. Since we have shown that PKH^{POS} cells are $EpCAM^+$, double-positive staining for EpCAM and the selected hNMSC marker should identify those rare cells displaying an $EpCAM^+/\text{hNMSC-marker}^+$ phenotype as putative hNMSCs. Indeed, such rare double-positive cells were found to reside in close proximity to the basal/myoepithelial layer in ducts, or at the duct-branch point (Figure 3E, Figure S3B).

Next, we used surface markers, derived from the hNMSC signature, to isolate cells endowed with mammosphere-forming ability, directly from the mammary gland. We used markers (CD24, CD49F, DNER, and DLL1) for which we had confirmed the differential protein expression between PKH^{POS} and PKH^{NEG} cells (Figures 2A and 3C). In monoparametric FACS, all markers were able to select populations enriched in mammosphere-forming cells, as witnessed by SFEs 15- to 50-fold higher than that of the bulk mammary population (Table 1, Figure S4A). Of note, DNER and DLL1, which were not previously known as hNMSC markers, were the most efficient ones (Table 1). We then tested

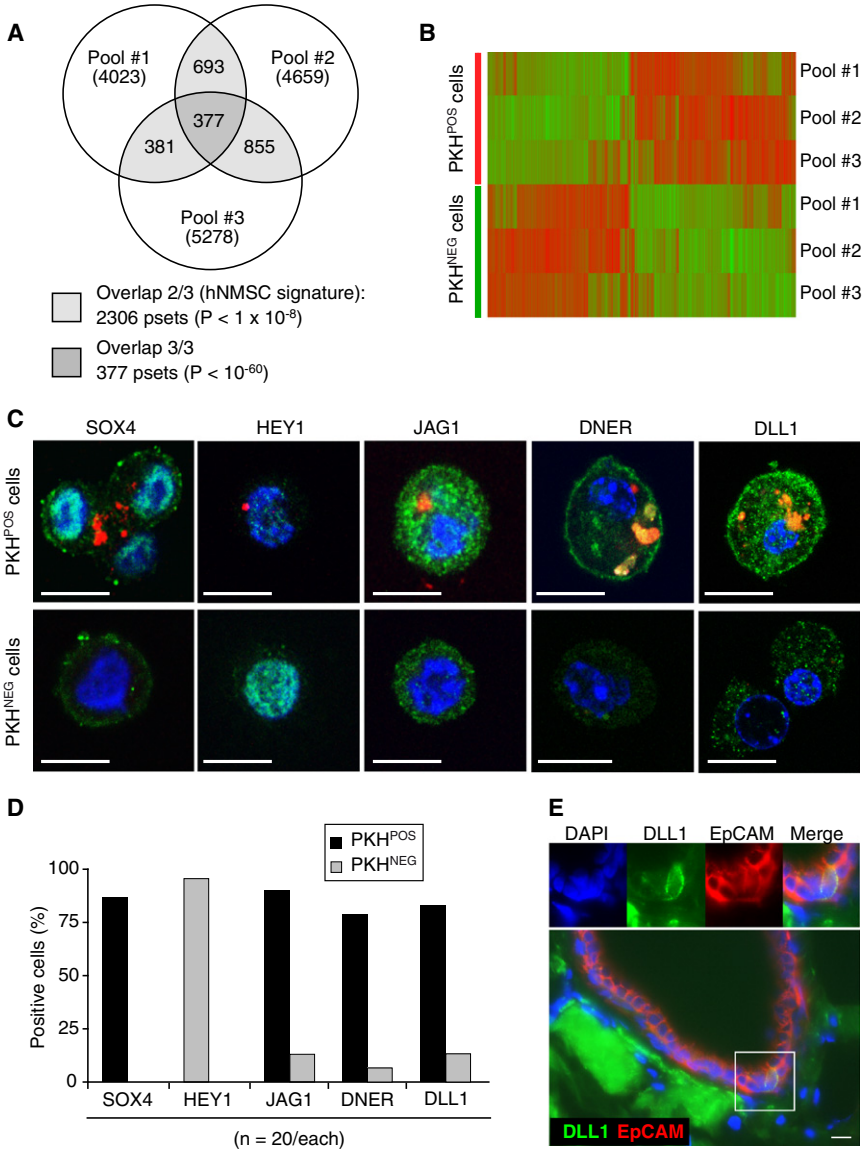

Figure 3. Gene Expression Profiling of hNMSCs and Immunophenotypical Validation

(A) Venn diagram of the psets differentially expressed in the PKH^{POS}/PKH^{NEG} comparison in the three RNA pools. The number of psets differentially expressed, in each pool, is shown in parentheses. At the intersections of the Venn diagram, the psets that fulfilled criteria for inclusion in the hNMSC signature or in the 3/3 signature are shown. p values were calculated as described in the Extended Experimental Procedures.

(B) Two-way hierarchical clustering of PKH^{POS} and PKH^{NEG} samples by the hNMSC signature.

(C and D) FACS-sorted PKH^{POS} or PKH^{NEG} cells were analyzed with the indicated antibodies by IF (C). Green, antibody staining; red, PKH26 epifluorescence; blue, DAPI. Scale bars represent 10 μm. Quantification is shown in (D). Results are typical and representative of three experiments that yielded comparable results.

(E) Fresh sections of normal mammary glands were analyzed in IF with the indicated antibodies. Magnifications of the boxed regions are shown on the top. The scale bar represents 10 μm.

See also Figure S3 and Table S1.

our markers in three combinations (CD24/CD49F/DNER, CD24/CD49F/DLL1, and CD49F/DNER/DLL1) in triparametric sorting. In all cases, only cells positive for the expression of all three markers were able to support mammosphere growth (Table 1, Figures S4B, S4C, and S4D), with efficiencies significantly increased with respect to the monoparametric sorting (additional data on biparametric FACS experiments are in Figure S4E). In particular, the combination CD49F/DNER/DLL1 was the most effective one and yielded cells displaying a SFE >500-fold higher than that of the bulk mammary population (Table 1). Finally, we showed that $CD24^{H}$/$CD49F^{H}$/$DNER^{H}$ cells, but not $CD24^{H}$/$CD49F^{H}$/$DNER^{L}$ cells, could reconstitute a normal mammary epithelium upon transplant (Figures S4F, S4G, and S4H).

We note that some of results reported herein on markers of hNMSC are at variance with recently published ones (Lim et al., 2009; Shipitsin et al., 2007). While these differences, and possible explanations for them, are discussed in details in the legend to Figure S4 (and in the corresponding section of Extended Experimental Procedures), the sum of our results shows that markers of our hNMSC signature reliably allow the identification and the prospective isolation in vivo of hNMSCs.

The hNMSC Signature Predicts Biological and Molecular Features in Breast Cancer

We used the hNMSC signature (and the 3/3 signature) to perform a meta-analysis of published breast cancer expression data sets. We initially investigated whether the hNMSC signature could stratify breast cancers according to their biological characteristics, exemplified by tumor grade. We employed two data sets published by Ivshina et al. (2006) and Pawitan et al. (2005), from which we extracted the information relative to poorly differentiated (G3) and well-differentiated (G1) tumors (Extended Experimental Procedures). We tested, by gene set enrichment analysis (GSEA), whether the genes of the hNMSC signature were coordinately expressed in G3 versus G1 tumors. There was a significant enrichment of genes upregulated in PKH^{POS} cells among the genes upregulated in G3 tumors, and a mirroring enrichment of concordantly downregulated genes (Figure 4A, and Table S2, "Specificity controls" sheet). From the GSEA of the Ivshina and Pawitan data sets, we identified 792 and 842 "core enriched genes," respectively, with 595 genes in common (Figure 4C, Table S2). Comparable results were obtained when we used, as a starting point for GSEA, the more stringent 3/3 signature of 377 psets (Figures 4B and 4D).

Table 1. Sphere-Forming Potential of Purified Mammary Cell Fractions

Type of Sorting	Cell Fraction	Gate	SFE (%)	Enrichment (Fold)
Nonsorted	Bulk mammary		0.003 ± 0.001	
Monoparametric Sorting	CD24 HIGH	Fr. 1	0.100 ± 0.007	33
	CD24 MEDIUM	Fr. 2	<0.01	
	CD24 LOW	Fr. 3	<0.01	
	CD49F HIGH	Fr. 4	0.049 ± 0.015	16
	CD49F MEDIUM	Fr. 5	<0.01	
	CD49F LOW	Fr. 6	<0.01	
	DNER HIGH	Fr. 7	0.146 ± 0.005	49
	DNER LOW	Fr. 8	<0.01	
	DLL1 HIGH	Fr. 9	0.121 ± 0.004	40
	DLL1 LOW	Fr. 10	<0.01	
CD24/CD49F/DNER Multiparametric Sorting	$CD24^{HIGH}$, $CD49F^{HIGH}$, $DNER^{HIGH}$	Fr. 17	0.659 ± 0.040	220
	$CD24^{HIGH}$, $CD49F^{HIGH}$, $DNER^{LOW}$	Fr. 18	<0.01	
CD24/CD49F/DLL1 Multiparametric Sorting	$CD24^{HIGH}$, $CD49F^{HIGH}$, $DLL1^{HIGH}$	Fr. 19	0.512 ± 0.058	171
	$CD24^{HIGH}$, $CD49F^{HIGH}$, $DLL1^{LOW}$	Fr. 20	<0.01	
CD49F/DLL1/DNER Multiparametric Sorting[a]	$CD49F^{POS\ (MED+HIGH)}$, $DLL1^{HIGH}$, $DNER^{HIGH}$	Fr. 21	1.591 ± 0.330	530
	$CD49F^{POS\ (MED+HIGH)}$, $DLL1^{HIGH}$, $DNER^{LOW}$	Fr. 22	<0.01	
	$CD49F^{POS\ (MED+HIGH)}$, $DLL1^{LOW}$, $DNER^{LOW}$[b]	Fr. 23	<0.01	
	$CD49F^{NEG\ (LOW)}$, $DLL1^{HIGH}$, $DNER^{LOW}$[b]	Fr. 24	<0.01	
	$CD49F^{NEG\ (LOW)}$, $DLL1^{LOW}$, $DNER^{LOW}$[b]	Fr. 25	<0.01	

Cells from the indicated FACS experiments (see also Figure S4) were tested for their mammosphere-forming ability. The column "gate" shows the cellular fractions as from Figure S4. The fold enrichment was calculated with respect to the bulk mammary population.

[a] In these experiments CD49F-sorted cells were divided into a "positive" fraction (corresponding to the medium + high fractions of the monoparametric sorting experiments) and in a "negative" fraction (corresponding to the low fraction of the monoparametric sorting experiments).

[b] $CD49F^{POS}/DLL1^{LOW}$, $CD49F^{NEG}/DLL1^{HIGH}$, and $CD49F^{NEG}/DLL1^{LOW}$ cells were also tested for DNER and found to be $DNER^{LOW}$ (data not shown).

We then investigated whether the hNMSC signature could distinguish among the molecular subtypes of breast cancer identified by Perou et al. (2000) and Sørlie et al. (2001). The signature could separate basal-type tumors from other molecular types of breast cancers (ErbB2-type, luminal-A or -B), in a manner that was apparently independent of their histological grade (Figure S5). Furthermore, GSEA showed enrichment of genes concordantly upregulated in PKH^{POS} cells and in basal-type tumors, despite the fact that 63% of the nonbasal tumors were G3s, as compared to 16% G1s in the same group (Figure S5).

Cancer-Initiating Cells in G3 and G1 Tumors

The above results show that the hNMSC signature can stratify breast cancers on the basis of their biological and—at least in part—molecular characteristics. One possible interpretation of these results is that the heterogeneity of breast cancers might reflect their content in SC-like cells (i.e., cancer stem cells [CSCs]). To verify this possibility, we initially performed IHC/IF analysis of G1 and G3 tumors with markers from the hNMSC signature (SERPINB5, TOP2A, CK5, TP63, SOX4, CD24, ADRM1, DNER, DLL1, and JAG1). We found that the proportion of cells expressing hNMSC markers was significantly higher (~3- to 4-fold, on average) in G3 versus G1 tumors (Figure 5A, Figure S6A). In consecutive 3 μm thick sections from G3 tumors, we detected simultaneous expression of hNMSC markers in clusters of tumor cells (Figure S6B), suggesting that tumor cells might express en bloc features associated with a SC program (as also confirmed by IF results with other markers, Figure 5C). Finally, we observed a similar trend in the SC-like content of G3 versus G1 subtypes of preinvasive breast lesions (ductal carcinomas in situ [DCIS]) (Figure 5B), supporting the idea that the cells expressing hNMSC markers might constitute true CSCs (or cancer-initiating cells) already present at a very early stage of tumorigenesis.

We also used markers from the hNMSC signature (CD49F, DNER, DLL1, Figure 5C; see also the legend to Figure S6C—and the corresponding section of Extended Experimental Procedures—for the rationale for the use of these markers) to isolate prospectively CSCs from breast tumors. Consistent with the data obtained in IHC/IF (Figures 5A and 5C), in triparametric FACS, we recovered 2- to 4-fold more triple-positive cells from G3 than from G1 tumors (Figure 5D). Triple-positive ($CD49F^{+}/DLL1^{H}/DNER^{H}$) cells, but not cells negative for DNER ($CD49F^{+}/DLL1^{H}/DNER^{L}$) (see Figure S6D for FACS profiles) were able to form mammosphere-like structures (Figure 5E). Of note, cells sorted from G3 tumors were ~3-fold enriched in mammosphere-initiating cells compared to cells sorted from G1 tumors (Figure 5E). Finally, we proved that $CD49F^{+}/DLL1^{H}/DNER^{H}$ cells, both from G3 and G1 tumors, are enriched in cancer-initiating cells, as they were able to form tumors upon xenotransplantation, more efficiently than unsorted cells (Figure 5F, compare also to Figure S7B for more experiments with unsorted cells).

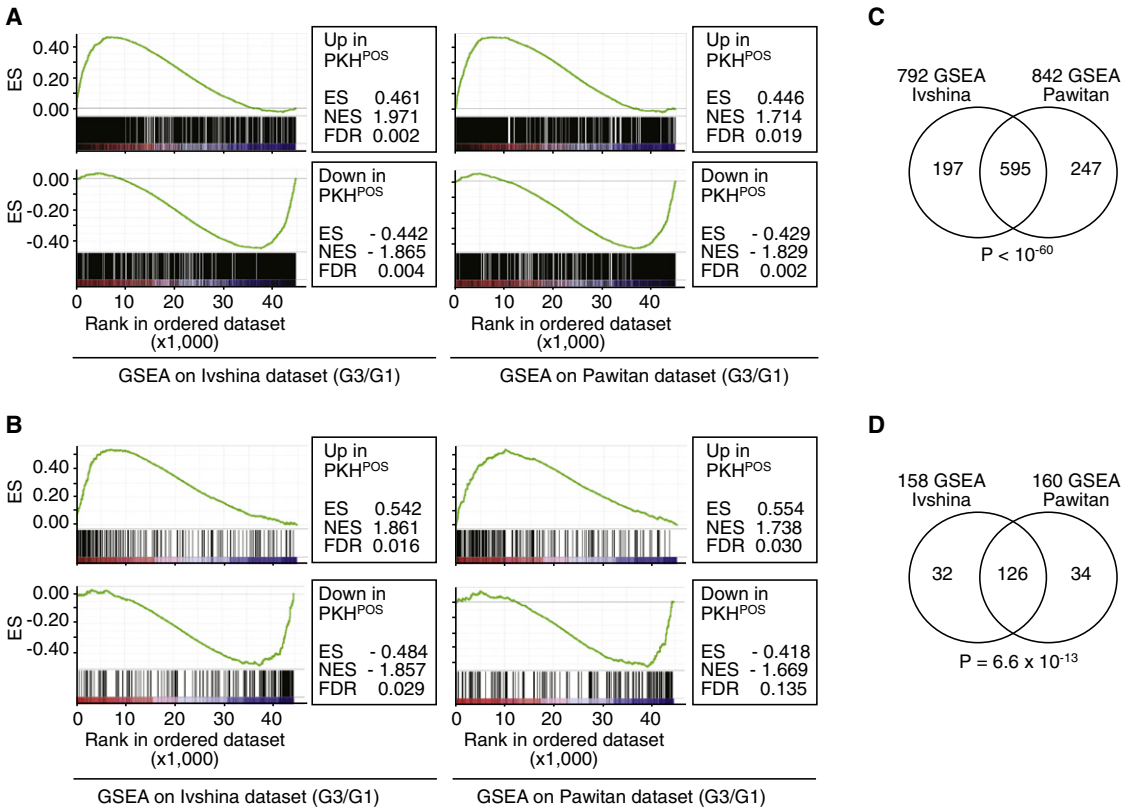

Figure 4. Prediction of Tumor Grade by the hNMSC Signature

(A and B) GSEA results for genes upregulated (top) and downregulated (bottom) in PKH^{POS} cells, in the Ivshina (left) and Pawitan (right) cohorts of samples.

(A) GSEA analysis with the hNMSC signature.

(B) GSEA analysis with the 3/3 signature. ES, enrichment score; NES, normalized enrichment score; FDR, false discovery rate-adjusted q value. Genes were first ranked according to the expression ratio (G3 over G1).

(C and D) Venn diagrams of the core enriched genes identified in (A) and (B) (C and D correspond to the analyses in A and B, respectively); the significance of the overlap (P) was calculated by binomial distribution.

See also Figure S5 and Table S2.

The precise correspondence between the data obtained in IHC/IF (Figures 5A and 5B), and those obtained by prospective isolation of CSCs (Figures 5D–5F) strongly argues in favor of the possibility that G3 tumors indeed display a higher CSC content than G1 tumors. If so, this should be reflected in different abilities of unsorted cells from the two types of tumors to form mammosphere-like structures in vitro and to give rise to tumors, when transplanted in vivo. To test this, we initially employed a cohort of 28 patients (eight normal, five G1, and 15 G3). Cells from all patients grew as spheroids, when plated in suspension culture (Figure 6A). G1 tumors formed mammosphere-like structures with an efficiency slightly, but significantly, higher than hNMSCs, both at the first and second generations (Figure 6A). Conversely, G3 tumors displayed ~3-fold higher SFE than normal or G1-derived cells (Figure 6A), in good agreement with the 3- to 4-fold increase in putative CSCs evidenced in the IHC/IF analysis (Figure 5A). In addition, G3 tumors gave rise to spheroids that were almost twice as big as those generated by normal or G1 cells (Figures 6A and 6B). Of note, in PKH26-labeling experiments, spheroids from G3 tumors displayed ~4-fold more PKH^{POS} cells than spheroids from G1 tumors (Figure 6B, Figure S7A).

Finally, in xenotransplantation experiments, cells from both G1 and G3 tumors gave rise to tumors. In G1s (seven tumors tested), the frequency of CSCs was comparable to that of hNMSCs present in the normal gland (Figure 6C, quantitations in Figure S7B), in agreement with the mammosphere data of Figure 6A. In G3 tumors (eight tumors tested), we detected an ~4-fold increase in CSCs (Figure 6C, Figure S7B). In addition, G3 tumors grew to a larger size with respect to G1 tumors (Figure 6D, top), again in agreement with data on mammosphere size, shown in Figure 6A. Tumors obtained in mice retained the G1 and G3 characteristics of the original human tumors (Figure 6D, bottom). Finally, comparable results were obtained when cells derived from G1 or G3 mammospheres were used to induce tumors in mice (Figure S7C), thus formally linking the data obtained in the mammosphere assay and in the tumor transplantation experiments.

DISCUSSION

We report here a method, based on the functional labeling of hNMSCs, that enables the purification of hNMSCs to near homogeneity. We demonstrated that cells prospectively isolated from

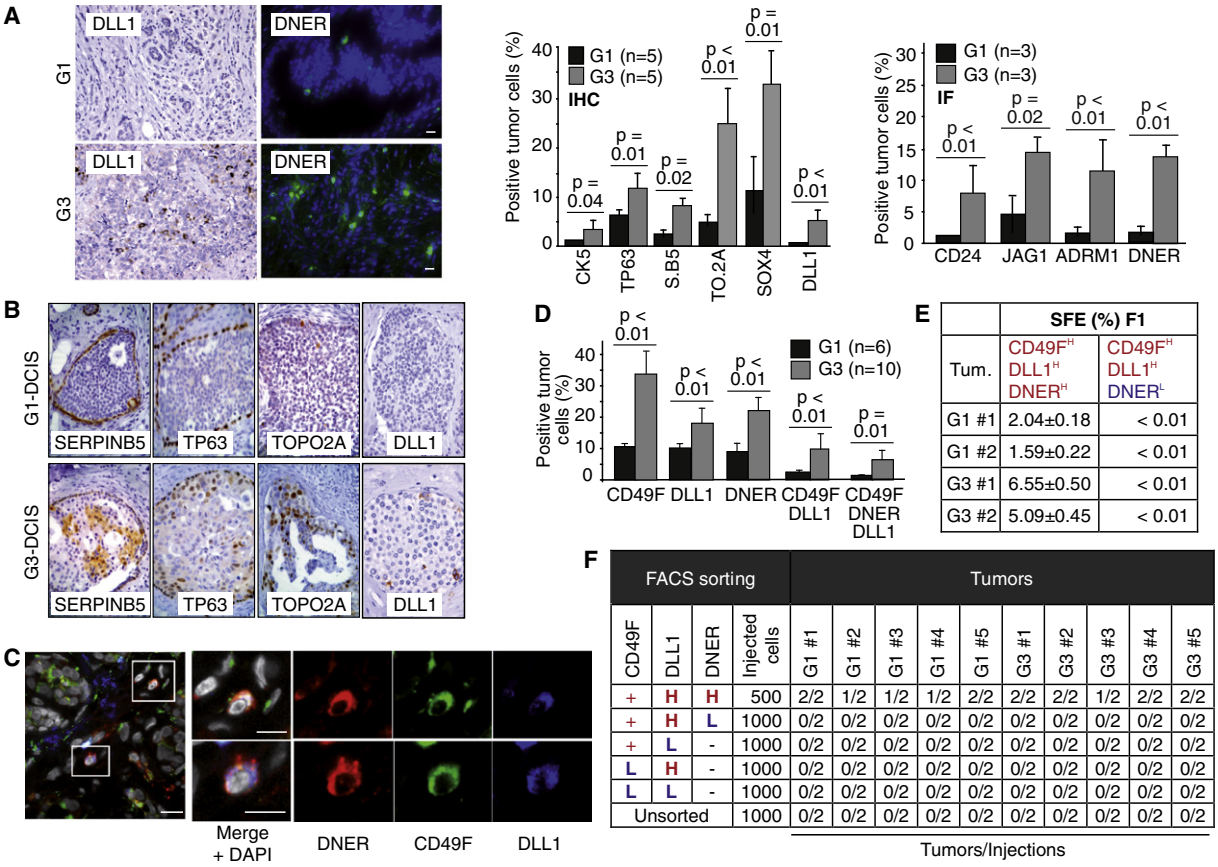

Figure 5. Cells Expressing Markers of the hNMSC Signature Are Enriched in G3 versus G1 Tumors

(A) Paraffin-embedded or frozen OCT-embedded sections were analyzed with markers derived from the hNMSC signature in IHC (CK5, TP63, SERPINB5, TOPO2A, SOX4, and DLL1) or IF (CD24, JAG1, ADRM1, and DNER). Examples of staining are on the left (original magnification ×40). Scale bars represent 50 μm (right); additional examples of stainings are in Figure S6A. Quantification is on the right, as obtained on multiple sections from three to five tumors in each group (statistical significance G3 versus G1 is shown). Data are expressed as the percentages of positive tumor cells (mean ± SD) in the total epithelial population.

(B) Paraffin-embedded sections (3 μm thick) from the indicated type of preinvasive DCIS lesions were analyzed in IHC with the indicated antibodies. Results were confirmed on multiple sections from three different DCIS for each subtype. Original magnification ×40.

(C) Representative image of a fresh section of a G3 tumor analyzed in IF with the indicated antibodies (DNER, CD49F, and DLL1). Magnifications of the boxed regions are shown on the right. Scale bars represent 10 μm.

(D) Bar graph depicting the frequency of different cell populations identified by FACS analysis with antibodies against the indicated markers, in G1 (n = 6) and G3 (n = 10) tumors (FACS profiles are in Figure S6D). Data are expressed as the percentage of the total number of epithelial cells. Data are expressed as the percentages of positive tumor cells (mean ± SD) in the total epithelial population.

(E) FACS-sorted $CD49F^{+}/DLL1^{H}/DNER^{H}$ and $CD49F^{+}/DLL1^{H}/DNER^{L}$ cells (fractions 5 and 6 in Figure S6D) from G1 (n = 2) and G3 (n = 2) tumors were tested for their mammosphere-forming ability. Data are from two independent experiments, each performed in duplicate.

(F) FACS-sorted (shown on the left, see Figure S6D) or unsorted cells, from three G1 and 3 G3 tumors, were transplanted at the indicated numbers into mammary fat pads of 21-day-old immunocompromised NOD/SCID mice. The number of tumors/injections is shown.

See also Figure S6.

cultured mammospheres through this methodology (PKH^{POS} cells) possess all the features expected for authentic hNMSCs. This allowed functional and molecular studies of hNMSCs, leading to a number of conclusions relevant to the homeostasis of the hMNSC compartment and its subversion in cancer.

Molecular Features of hNMSCs and of the Progenitors of the TA Compartment

By exploiting the high degree of purity of PKH^{POS} cells, we obtained comparative molecular profiles of hNMSCs and of their immediate progeny, which were validated by several molecular and cellular criteria. In principle, molecular determinants and pathways of the hNMSC signature should reflect circuitries that are relevant to the maintenance of the hNMSC compartment, and to the molecular strategies enacted by progenitors to exit this compartment. Many such circuitries could be readily identified (Table S1, "Functional classification" sheet), including: (1) regulators of cell survival, cell cycle, and telomerase activity, which seemingly underlie the hNMSC quiescent state and their refractoriness to apoptosis; (2) growth factor and chemokine

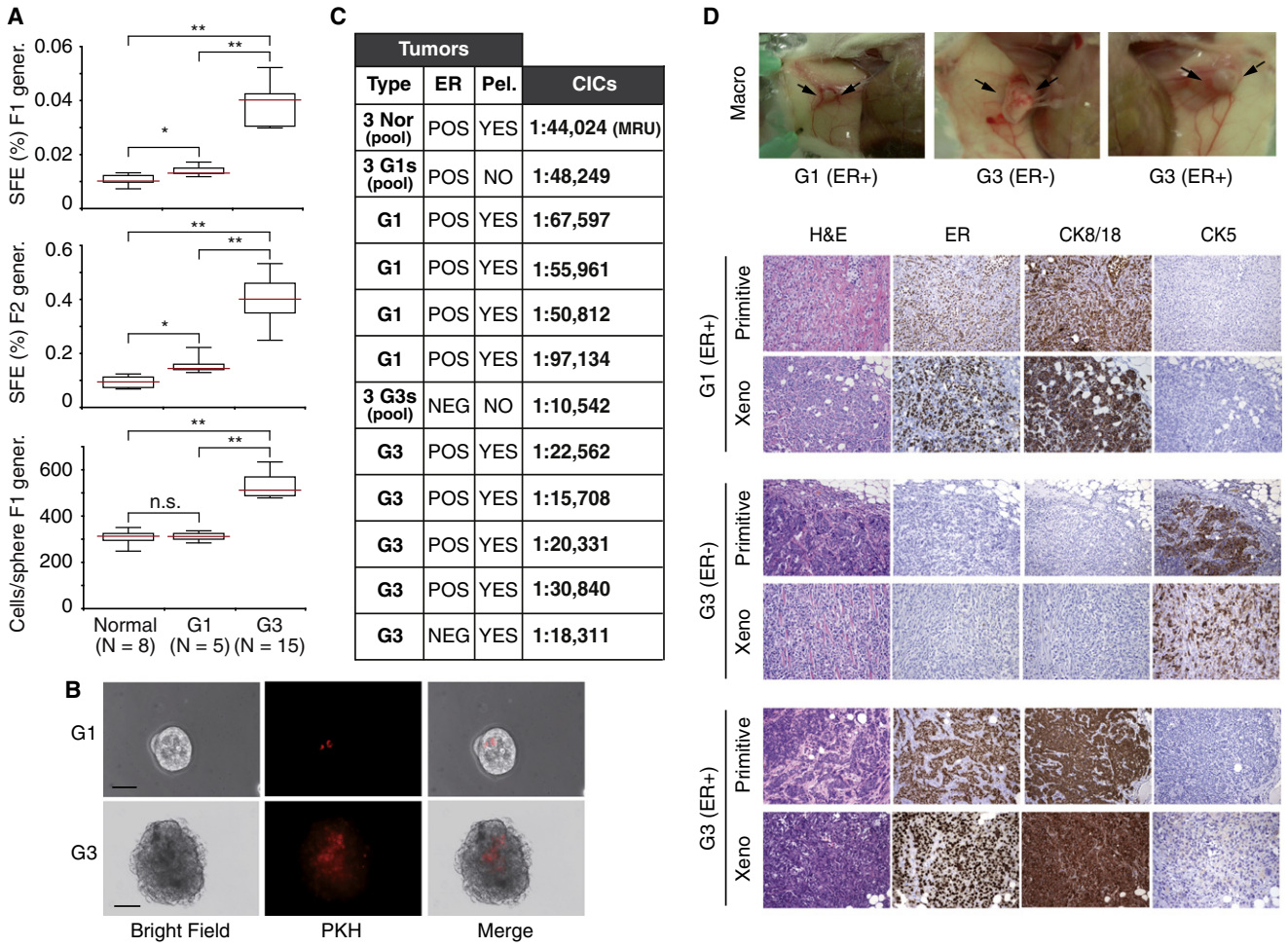

Figure 6. Poorly Differentiated Breast Cancers Are Enriched in Cancer-Initiating Cells

(A) SFEs (F1, top; F2, middle) and mammosphere size (bottom). Results are shown as box plots, extending from the 25th to the 75th percentiles (red line, median). The whiskers below and above each box plot extend to the 1.5× of the interquartile range (75th–25th percentile). *$p \leq 0.01$, **$p < 0.0001$.

(B) Representative images of mammospheres from PKH26-labeled G1-derived and G3-derived tumors (additional characterizations are in Figure S7A). Scale bars represent 100 μm.

(C) Frequency of mammary repopulating units (MRU) in bulk mammary normal cells (Nor), and of cancer-initiating cells (CICs) in tumor cells from individual G1 and G3 tumors (G1, G3). The complete set of data is in Figure S7B. A total of seven G1 and eight G3 tumors were analyzed. In two experiments (pool), tumor cells from three pooled G1 and G3 tumors were xenografted. The Estrogen receptor (ER) status of the primitive tumors is reported. When indicated, Estrogen pellets (Pel.) were used.

(D) Top: Representative tumor growths at the sites of injection of 100,000 tumor cells from G1 or G3 tumors (Macro). Arrows point to the tumors. Bottom: Representative images of IHC analysis of the different tumor types used in (C), and of their corresponding xenotransplants, with the indicated antibodies. Original magnification ×20.

See also Figure S7.

receptors, and molecules involved in cell-to-cell and cell-to-extracellular matrix contacts, which suggest the ability of SCs to organize their "niche" by interacting with neighboring cells, and to respond to their special environment; (3) transcription and chromatin remodeling factors, possibly required to regulate transcriptional programs; and (4) molecules involved in oxidative stress/drug response and in DNA damage checkpoint/repair, in line with the notion that SCs are uniquely programmed to preserve their homeostasis, and in particular their genome integrity (see Table S1, where a more detailed discussion is also provided).

While a comprehensive analysis of the molecular characteristics of the hNMSC signature will be impossible here, one feature is worth mentioning. We have previously shown that p53 critically controls the binary fate decision of NMSCs in the mouse mammary gland by influencing the rate of symmetric versus asymmetric self-renewing cell divisions (Cicalese et al., 2009). While some caution is due when extrapolating results from mice to humans, it is nevertheless tempting to speculate that similar regulatory mechanisms might also exist in hNMSCs. Indeed, we found that genes annotated as putative p53 targets on the basis of chromatin immunoprecipitation experiments

(Wei et al., 2006) were significantly enriched ($p < 0.002$) in the hNMSC signature. In particular, we found that 42 out of 426 "highly reliable" putative p53 targets (Wei et al., 2006) were present in the hNMSC signature (Table S1, "p53 putative targets sheet").

SCs and the Heterogeneity of Human Breast Cancers

When the hNMSC signature was applied to the meta-analysis of breast cancer expression data sets, we found out that it was predictive of biological and molecular features of human breast cancers. Thus, breast cancers can be distinguished based on their degree of resemblance to the hNMSC molecular phenotype.

Recently, Ben-Porath et al. described an embryonic stem-like signature, which could predict, in breast cancers, tumor grade and several additional features, including clinical outcome (Ben-Porath et al., 2008). A direct comparison of our results with those of Ben-Porath et al. is not straightforward, since they meta-analyzed several gene sets associated with human embryonic stem cell identity and used different tools for the analysis of gene set enrichment patterns. In general, however there is no significant overlap between the hNMSC signature and signatures of the ES state. For instance, the "ES exp1" signature of Ben-Porath et al., which includes 380 genes and separates clearly G3 from G1 tumors, shows an overlap of only 16 concordantly regulated genes with the hNMSC signature.

There are plausible reasons to explain the lack of overlap. First, the starting points of the signatures are not the same. In our case, adult mammary SCs were analyzed, while the study of Ben Porath et al. meta-analyzed signatures representative of various embryonic SC states. Second, signatures do not portray the "absolute" molecular picture of a given condition, but only the comparative picture with respect to another condition. Thus, a "SC signature" can be different according to whether it was derived by comparison of the SC to a pluripotent progenitor, a committed progenitor, or a differentiated cell—a consideration to bear in mind also when comparing our data to other signatures reported for mammary SCs (see for instance Raouf et al. [2008] and discussion in the legend to Table S1 and in the corresponding section of Extended Experimental Procedures). Still, individual signatures might explore the same molecular "universe" from different (and limited) viewpoints, possibly leading to similar biological conclusions. From this perspective, the conclusions of our studies and of those of Ben-Porath et al. are remarkably similar, in that in both cases genes associated with SC identity (adult or embryonic, respectively) could stratify breast tumors according to biological and molecular features.

There are two major ways of interpreting these findings. On the one hand, different types of breast cancers might represent a continuum in which each subclass corresponds to an incremental degree of divergence from the hNMSC. On the other, it is possible that the degree of similarity of a tumor to the hNMSC signature might simply reflect the CSC content of the tumor. Our experiments fully supported this latter possibility, as we could show that G3/poorly differentiated tumors contain more CSCs (or cancer-initiating cells) than G1/well-differentiated tumors, although we do not know yet whether this is a general property of poorly differentiated tumors or of a subset of them. In turn, these findings argue that the heterogeneity of breast cancers could be explained, at least in part, by the number/proportion of cells displaying stem-like features contained within the tumor.

The Cellular Origin of Human Breast Cancers

Our findings do not directly address the question of the origin of mammary cancer-initiating cells, i.e., whether they derive from malignant transformation of a hNMSC or from progenitors that have reacquired self-renewal ability. We have previously shown that skewing self-renewal division from an asymmetric (one stem → one stem + one progenitor) to a symmetric (one stem → two stems) mode is a critical step in a model of murine mammary cancerogenesis (Cicalese et al., 2009). Kinetic analysis also revealed that when CSCs undergo self-renewal, they tend to skip rounds of asymmetric cell division in favor of symmetric division (Cicalese et al., 2009). Indeed, we report in this manuscript evidence (see Figure 6B and discussion in the legend to Figure S7A) that a different rate of skipping of asymmetric division might determine the different number of cancer stem cells in G3 versus G1 tumors, and might therefore sit at the heart of the biological and clinical heterogeneity of breast cancers.

Together, data in our present and previous (Cicalese et al., 2009) studies suggest a scenario for mammary tumorigenesis. In this model, normal SCs are the targets of different oncogenic events. The nature of the transforming event(s) determines the frequency with which the transformed SCs will skip asymmetric self-renewing division. This, in turn, will determine the final number of CSCs within the tumor tissue, and a number of biological and clinical features of the tumor. This model does not preclude additional differential effects of the transforming events, which might, for instance, only allow differentiation toward a certain lineage or up to a certain point in the differentiation program, thus further contributing to the heterogeneity of breast cancers (Lim et al., 2009; Shipitsin et al., 2007).

Finally, we have shown, in a model of murine cancerogenesis, that reduced tumor growth can be achieved by pharmacological interference with the self-renewing properties of CSCs (Cicalese et al., 2009). Here, we show that the number of CSCs in human breast cancers can vary greatly, with discernible impact on several clinical and pathological features. Together, these data provide strong support for the concept of "cancer stem cell-targeted therapy" to eradicate cancer.

EXPERIMENTAL PROCEDURES

Clinical Samples

Fresh, frozen, or archival formalin-fixed paraffin-embedded (FFPE) mammary tissue specimens were collected at the European Institute of Oncology (Milan, Italy). All tissues were collected via standardized operative procedures approved by the Institutional Ethical Board, and informed consent was obtained for all tissue specimens linked with clinical data.

Cultivation of Mammospheres and FACS Procedures

Epithelial cells, from reductive mammoplasties (Pece et al., 2004), were allowed to adhere for 24 hr in complete SC medium (Dontu et al., 2003). Cells were trypsinized, filtered through a 100 μm and a 40 μm cell strainer, resuspendend in PBS (~500,000 cells/ml), and labeled with PKH26 (Sigma, 10^{-7} M, 5 min). Labeled cells were plated (30,000 cells/ml) in suspension (Dontu et al., 2003). After 7–10 days, mammospheres were harvested, dissociated

enzymatically (0.05% trypsin/0.5 mM EDTA for 10 min, plus filtering through a 40 μm cell strainer), and subjected to FACS analysis with a FACS Vantage SE flow cytometer (Becton & Dickinson) to yield PKH^{POS} and PKH^{NEG} cells.

IF of PKH^{POS} and PKH^{NEG} cells was performed on cytospin preparations fixed with 4% formaldehyde (5 min)/cold methanol (5 min). IF on mammary glands (from OCT-embedded, snap-frozen samples) was performed on 3 μm thick sections, fixed for 10 min with cold methanol/acetone (1:1, v/v). IHC was performed as previously described (Pece et al., 2004).

The preparation of FACS-sorted samples, directly from bulk mammary gland (Figure S4), is described in the Extended Experimental Procedures.

Differentiation Assays

For 2D differentiation assays, PKH^{POS} and PKH^{NEG} cells were plated at clonogenic density (500 viable cells/well) onto glass slides coated with Matrigel (BD Biosciences) and grown for 10 days in primary mammary epithelial cell medium (Pece et al., 2004), followed by IF with the appropriate antibodies. 3D Matrigel cultures were performed as described (Dontu et al., 2003); where indicated, treatment with prolactin (Sigma) was for 10 days at 5 μg/ml.

Immunofluorescence, Immunohistochemistry, and Image Quantitative Analysis

For the immunophenotypical characterization of PKH^{POS} cells and PKH^{NEG} cells shown in Figure 2 and in Figure 3C, we employed the following primary antibodies: for IF analysis, anti-CD49F and anti-TP63 (BD Bioscience PharMingen), anti-SOX4 and fluorescein isothiocyanate (FITC) anti-ASMA (Sigma), anti-CD24 and biotinylated anti-EpCAM (Neomarkers), anti-CK5 (Covance), anti-HEY1 (ABR-Affinity BioReagents), anti-JAG1 and anti-Delta/DLL1 (Santa Cruz Biotechnology), anti-ADRM1 (Abnova), and biotinylated anti-DNER (R&D Systems). Samples were analyzed under an AX-70 Provis (Olympus) fluorescence microscope equipped with a black and white cooled CCD camera (Hamamatsu c5985), or with a Leica TCS SP2 AOBS confocal microscope equipped with 405, 488, 543, and 633 nm laser lines. Digital images were computer processed with Adobe Photoshop CS2. For the quantitative analysis of IF experiments performed on PKH^{POS} and PKH^{NEG} cells, the ImageJ image-analysis software (W. Rasband, National Institutes of Health) was used to measure the specific mean intensity on an average of ten cells for each antibody staining. Background intensity was initially subtracted by placing "regions of interest" over areas devoid of specific signal.

For immunohistochemical analysis of FFPE sections obtained from normal and tumor breast biopsies or from mammospheres, the primary antibodies used were as follows: anti-CK5, anti-Estrogen receptor, anti-EpCAM (Dako, clone Ber-EP4), and anti-TP63 (Dako), anti-SERPINB5 (Oncogene), anti-TOPO2A (Chemicon), anti-E-cadherin (BD Biosciences PharMingen), anti-β-casein (a kind gift from C. Streuli), anti-SOX4 (Abcam), anti-CD24 (NeoMarkers), anti-DLL1 (SantaCruz), and anti-CK8/18 (Novocastra). Immunocomplexes were visualized by the ABC peroxidase method (Vector Laboratories).

RNA Extraction and GeneChip Hybridization

Total RNA was extracted with the Arcturus PicoPure RNA Isolation Kit (Arcturus Engineering, CA) according to the manufacturer's instructions and analyzed with the Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA (50–100 ng) was amplified with the Affymetrix T7 Polymerase-based double linear amplification protocol. Complementary RNA probes (10 μg) were hybridized onto the Affymetrix HG-U133_Plus_2 ChipSet (Santa Clara, CA), according to Affymetrix technical protocols. The average intensity of each array was scaled to a predefined value (target intensity) of 500, in order to make arrays comparable.

Expression Profiles and Meta-Analysis

Hybridization onto the Affymetrix (Santa Clara, CA) HG-U133_Plus_2 ChipSet was as previously described (Vecchi et al., 2007). The statistical analysis was performed with the Affymetrix MAS5 "Comparison Analysis" algorithm (Affymetrix Statistical Algorithm Reference Guide, version 5 edition) by direct comparison of each PKH^{POS} fraction (baseline array) against its corresponding PKH^{NEG} fraction (experimental array), within individual matched pairs. Among all the psets differentially expressed in at least one comparison (n = 10575), we selected only those showing a concordant trend of regulation (either increased or decreased) in at least two comparisons and assigned a "not changed" call in the third comparison. Given the differences in RNA yield and hybridization efficiency among the three pools of samples, for data presented in Figures 3A and 3B, expression data were independently mean normalized within each comparison, and normalized values were then imported in GeneSpring GX software version 7.3.1 (Agilent Technologies, Santa Clara, CA) without any further adjustment. Additional details (also with reference to validation experiments performed by quantitative RT-PCR) are in the Extended Experimental Procedures. Data sets have been deposited in NCBI's Gene Expression Omnibus, and are accessible through GEO Series accession number GSE18931.

Procedures for meta-analysis of published breast cancer data sets (Ivshina et al., 2006; Pawitan et al., 2005) are in the Extended Experimental Procedures.

Transplantation Assays into the Cleared Fat Pads of NOD-SCID Mice

Three-week-old female NOD/SCID mice were anesthetized by an i.p. injection of Tribromoethanol (Avertin) (150 mg/kg). For the xenotransplant of human normal mammary cells (either PKH^{POS} or PKH^{NEG} cells, or bulk mammary cells, or normal mammospheres), the number 4 inguinal mammary glands were cleared and humanized with 2.5×10^5 nonirradiated telomerase-immortalized human mammary fibroblasts (a generous gift from G. Dontu, University of Michigan, and R. B. Clarke, University of Manchester) and 2.5×10^5 irradiated (4 Gy) fibroblasts, as previously described (Kuperwasser et al., 2004; Proia and Kuperwasser, 2006). Two weeks after humanization, donor normal mammary cells, combined with a new batch of irradiated and nonirradiated fibroblasts to enhance the engrafment, were resuspended in an equal volume of a Phenol-red free Matrigel (Becton Dickinson) mixture, supplemented with 0.04% trypan blue (Sigma), to a final volume of 35 μl, and injected into the humanized site.

For the xenotransplant of human breast cancer cells, tumor biopsies were dissociated mechanically and enzymatically and cells, resuspended as above, were injected directly into cleared mammary fat pads of NOD/SCID mice. At the same time, a 60 day release Estrogen (17-β estradiol) pellet (0.72 mg/pellet; Innovative Research of America) was placed subcutaneously on the back of the mouse's neck with a trocar, when indicated.

Animals injected with normal breast cells were euthanized after 10 weeks. Animals injected with cancer cells were euthanized when the tumors were approximately 0.5–1.2 cm in the largest diameter, to avoid tumor necrosis and in compliance with regulations for use of vertebrate animal in research. For whole-mount analysis, the fat pads were removed, fixed in Carnoy's solution (ethanol, glacial acetic acid, chloroform), and stained with 0.2% carmine alum (carmine, AlKSO4). For IHC analysis, samples of normal mammary whole mounts and of tumor xenotransplants were FFPE and assayed with the appropriate antibodies. Measurements of total ductal outgrowth area on whole glands were taken essentially as previously described (McCaffrey and Macara, 2009). All measurements were made with ImageJ imaging software (W. Rasband, National Institutes of Health). In brief, fat pad filling was measured by drawing a shape around the fat pad, with the Freehand tool. Ductal areas were measured by drawing a shape that connected the tips of ducts. The percentage of fat pad filled was measured by the formula (ductal area/fat pad area) × 100, and is shown in the corresponding whole-mount images with a graphic indication (pie) of the extent to which each transplant filled the fat pad.

The transplantation frequency was calculated by Poisson statistics, using the "statmod" software package for the R computing environment (http://www.R-project.org/) as previously described (Shackleton et al., 2006) and a complementary log-log generalized linear model (two-sided 95% Wald confidence intervals or, in case of zero outgrowths, one-sided 95% Clopper-Pearson intervals), respectively. The single-hit assumption was tested as recommended and was not rejected for any dilution series ($p > 0.05$).

ACCESSION NUMBERS

Data sets reported in this paper have been deposited in NCBI's Gene Expression Omnibus (GEO Series) with the accession number GSE18931.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at doi:10.1016/j.cell.2009.12.007.

ACKNOWLEDGMENTS

We thank the Division of Senology at IEO, in particular V. Galimberti, A. Luini, P. Naninato, and P. Arnone, for the collection of human samples; G. Dontu, R.B. Clarke, C. Streuli, and J.M. Verdi for generously providing reagents; A. Cocito for statistical consultation and assistance; G. Matera, D. Galvagno, A. Fiocchi, G. Taliento, O. Oldazzi, A. Cavallon, V. Stufano, D. Nicosia, and M. Garré; the IEO Biobank and the centralized facilities of the IFOM-IEO Campus for technical assistance; P. Nuciforo for histopathological analysis; and G. Bonizzi, A. Cicalese, E. Belloni, and P.R. Romano for insightful discussions. This work was supported by grants from the Associazione Italiana per la Ricerca sul Cancro and MIUR to S.P., P.G.P., and P.P.D.F.; the European Community (FP6 and FP7) and the CARIPLO foundation to P.P.D.F. and P.G.P.; the Ferrari Foundation and the Monzino Foundation to P.P.D.F.; the G. Vollaro Foundation to S.P.; and the European Research Commission to P.P.D.F.

Received: April 9, 2008
Revised: February 18, 2009
Accepted: November 24, 2009
Published: January 7, 2010

REFERENCES

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat. Genet. *40*, 499–507.

Carter, W.G., Kaur, P., Gil, S.G., Gahr, P.J., and Wayner, E.A. (1990). Distinct functions for integrins alpha 3 beta 1 in focal adhesions and alpha 6 beta 4/bullous pemphigoid antigen in a new stable anchoring contact (SAC) of keratinocytes: relation to hemidesmosomes. J. Cell Biol. *111*, 3141–3154.

Chepko, G., and Smith, G.H. (1997). Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. Tissue Cell *29*, 239–253.

Cicalese, A., Bonizzi, G., Pasi, C.E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Di Fiore, P.P., and Pelicci, P.G. (2009). The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. Cell *138*, 1083–1095.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Eirew, P., Stingl, J., Raouf, A., Turashvili, G., Aparicio, S., Emerman, J.T., and Eaves, C.J. (2008). A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. Nat. Med. *14*, 1384–1389.

Gönczy, P. (2008). Mechanisms of asymmetric cell division: flies and worms pave the way. Nat. Rev. Mol. Cell Biol. *9*, 355–366.

Huang, S., Law, P., Francis, K., Palsson, B.O., and Ho, A.D. (1999). Symmetry of initial cell divisions among primitive hematopoietic progenitors is independent of ontogenic age and regulatory molecules. Blood *94*, 2595–2604.

Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H., et al. (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res. *66*, 10292–10301.

Jones, C., Mackay, A., Grigoriadis, A., Cossu, A., Reis-Filho, J.S., Fulford, L., Dexter, T., Davies, S., Bulmer, K., Ford, E., et al. (2004). Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. Cancer Res. *64*, 3037–3045.

Kuperwasser, C., Chavarria, T., Wu, M., Magrane, G., Gray, J.W., Carey, L., Richardson, A., and Weinberg, R.A. (2004). Reconstruction of functionally normal and malignant human breast tissues in mice. Proc. Natl. Acad. Sci. USA *101*, 4966–4971.

Lanzkron, S.M., Collector, M.I., and Sharkis, S.J. (1999). Hematopoietic stem cell tracking in vivo: a comparison of short-term and long-term repopulating cells. Blood *93*, 1916–1921.

Liao, M.J., Zhang, C.C., Zhou, B., Zimonjic, D.B., Mani, S.A., Kaba, M., Gifford, A., Reinhardt, F., Popescu, N.C., Guo, W., et al. (2007). Enrichment of a population of mammary gland cells that form mammospheres and have in vivo repopulating activity. Cancer Res. *67*, 8131–8138.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., et al; kConFab. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913.

McCaffrey, L.M., and Macara, I.G. (2009). The Par3/aPKC interaction is essential for end bud remodeling and progenitor differentiation during mammary gland morphogenesis. Genes Dev. *23*, 1450–1460.

Pawitan, Y., Bjöhle, J., Amler, L., Borg, A.L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S., et al. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res. *7*, R953–R964.

Pece, S., Serresi, M., Santolini, E., Capra, M., Hulleman, E., Galimberti, V., Zurrida, S., Maisonneuve, P., Viale, G., and Di Fiore, P.P. (2004). Loss of negative regulation by Numb over Notch is relevant to human breast carcinogenesis. J. Cell Biol. *167*, 215–221.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Proia, D.A., and Kuperwasser, C. (2006). Reconstruction of human mammary tissues in a mouse model. Nat. Protoc. *1*, 206–214.

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J., et al. (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. Cell Stem Cell *3*, 109–118.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Shipitsin, M., Campbell, L.L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., Nikolskaya, T., Serebryiskaya, T., Beroukhim, R., Hu, M., et al. (2007). Molecular definition of breast tumor heterogeneity. Cancer Cell *11*, 259–273.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA *98*, 10869–10874.

Stingl, J., and Caldas, C. (2007). Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. Nat. Rev. Cancer *7*, 791–799.

Stingl, J., Eaves, C.J., Kuusk, U., and Emerman, J.T. (1998). Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation *63*, 201–213.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.

Vecchi, M., Nuciforo, P., Romagnoli, S., Confalonieri, S., Pellegrini, C., Serio, G., Quarto, M., Capra, M., Roviaro, G.C., Contessini Avesani, E., et al. (2007). Gene expression analysis of early and advanced gastric cancers. Oncogene *26*, 4284–4294.

Visvader, J.E., and Lindeman, G.J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat. Rev. Cancer *8*, 755–768.

Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., et al. (2006). A global map of p53 transcription-factor binding sites in the human genome. Cell *124*, 207–219.

Williams, J.M., and Daniel, C.W. (1983). Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. Dev. Biol. *97*, 274–290.