Cell Stem Cell
Article

A Mammary Stem Cell Population Identified and Characterized in Late Embryogenesis Reveals Similarities to Human Breast Cancer

Benjamin T. Spike,[1,2,4] Dannielle D. Engle,[1,3,4,5] Jennifer C. Lin,[1,4] Samantha K. Cheung,[1,6] Justin La,[1] and Geoffrey M. Wahl[1,*]
[1]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92037, USA
[3]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92037, USA
[4]These authors contributed equally to this work
[5]Present address: Cambridge Research Institute, Cancer Research United Kingdom, Cambridge CB20RE, UK
[6]Present address: Division of Molecular and Cellular Biology, University of California, Berkeley, Berkeley, CA 94720, USA
*Correspondence: wahl@salk.edu
DOI 10.1016/j.stem.2011.12.018

SUMMARY

Gene expression signatures relating mammary stem cell populations to breast cancers have focused on adult tissue. Here, we identify, isolate, and characterize the fetal mammary stem cell (fMaSC) state since the invasive and proliferative processes of mammogenesis resemble phases of cancer progression. fMaSC frequency peaks late in embryogenesis, enabling more extensive stem cell purification than achieved with adult tissue. fMaSCs are self-renewing, multipotent, and coexpress multiple mammary lineage markers. Gene expression, transplantation, and in vitro analyses reveal putative autocrine and paracrine regulatory mechanisms, including ErbB and FGF signaling pathways impinging on fMaSC growth. Expression profiles from fMaSCs and associated stroma exhibit significant similarities to basal-like and Her2$^+$ intrinsic breast cancer subtypes. Our results reveal links between development and cancer and provide resources to identify new candidates for diagnosis, prognosis, and therapy.

INTRODUCTION

Breast cancers are a heterogeneous group of diseases distinguishable by histopathology and molecular profiling. Expression profiling of patient samples enabled their categorization into molecular subtypes referred to as luminal A, luminal B, Her2 positive, basal-like, and claudin-low (Herschkowitz et al., 2007; Perou et al., 2000). These divisions identify critical differences in cellular composition and molecular pathways, suggesting treatment options and correlating with patient survival (Prat and Perou, 2011). Prognostic expression signatures, refined by related approaches, are being tested or used clinically (Fan et al., 2011; Paik et al., 2006; van 't Veer et al., 2002; van de Vijver et al., 2002). Previously reported prognostic signatures and subtype designations identify a limited set of biologic programs correlating with hormone receptor status (estrogen receptors [ER] and progesterone receptors [PR]) and Her2 expression and proliferation (Desmedt et al., 2008; Fan et al., 2006; Haibe-Kains et al., 2008; Prat and Perou, 2011; Sotiriou and Piccart, 2007). While hormone receptors and Her2 are therapeutic targets, many breast cancers, including most of the basal-like subtypes, lack ER, PR, and Her2 expression and associated targeted treatment options (Pal et al., 2011).

Stem cell biology offers promise for understanding the origins and progression of breast and other cancers and may also reveal the next generation of molecular targets for breast cancers not susceptible to current agents. For example, basal-like breast cancers are poorly differentiated and exhibit gene expression similarities to embryonic and induced pluripotent stem cells (Ben-Porath et al., 2008; Mizuno et al., 2010). Expression profiles derived from adult mammary cells of different differentiation stages have also been used to designate cancers as stem-like or nonstem-like (Lim et al., 2009; Lim et al., 2010; Perou et al., 2010). Breast cancer cells that generate xenografted tumors with high efficiency, regenerate the cellular complexity of the originating tumor, and self-renew (as defined by secondary transplantation) exhibit properties attributed to stem cells and have consequently been called breast cancer stem cells (Al-Hajj et al., 2003). However, defining potential relationships between stem-like cells in breast cancer and normal mammary stem cells (MaSCs) requires MaSC isolation and characterization.

Adult MaSCs (aMaSCs) have been enriched via stem cell isolation methods, and their gene expression signatures have been reported (Lim et al., 2009; Lim et al., 2010; Pece et al., 2010; Raouf et al., 2008; Shackleton et al., 2006; Stingl et al., 2006). However, aMaSC rarity combined with the cellular complexity of the adult gland make purification challenging (Shackleton et al., 2006; Stingl et al., 2006), and copurifying stroma and differentiated mammary cells complicate elucidation of their core self-renewal and differentiation programs.

The developing mammary gland is less complex than the adult gland, suggesting that it may facilitate stem cell identification and purification. Furthermore, while the extensive proliferation, migration, and invasion required for mammogenesis do not occur in the resting adult mammary gland, they do resemble

processes mediating breast cancer progression (Veltmaat et al., 2003). These observations suggest that stem cells present in fetal mammary rudiments (i.e., fetal mammary stem cells [fMaSCs]) might express genes comprising pathways overlooked by analyses focused only on the adult mammary gland and that fMaSCs may reveal new targets to aide detection, prognosis, and treatment of breast cancers. Consistent with this idea, gene expression profiling of bulk epithelium from early mammogenesis revealed significant differences with the adult epithelium (Wansbury et al., 2011). Importantly, this study did not assess whether the profiled cells exhibited stem cell activity, so the relevance of these signatures to fMaSCs remains to be determined.

Mouse mammary gland development begins at approximately embryonic day (E)11.5 with a thickening of the ventral ectoderm to generate five pairs of mammary placodes. The placodes become spherical buds by E12.5, which elongate into ductal sprouts in females by E16 (Veltmaat et al., 2003). Invasion of the extending rudiment into the fat pad precursor begins by E16.5, and by E18.5 the mammary rudiment constitutes a primitive branched network within the mammary fat pad (Veltmaat et al., 2003).

Classic rudiment transplantation studies suggest that mammary stem cells may arise coincident with the morphologic specification of the mammary gland (Sakakura et al., 1979). For example, transplanting multiple intact E13 to E17 epithelial rudiments generated full mammary outgrowths (Sakakura et al., 1979). Heterotypic recombination experiments involving salivary mesenchyme demonstrated that the mammary epithelium is committed to develop into a mammary phenotype by E12.5 (Cunha and Hom, 1996).

Here, we quantify and characterize fMaSC activity during fetal mammogenesis, and analyze the relationships between their gene expression programs and those found in human breast cancer. Our data reveal the unexpected finding that fMaSCs are extremely rare in early embryogenesis but increase rapidly as the mammary rudiment invades into the fat pad precursor. We also show that fMaSCs and their associated stroma exhibit gene expression programs related to those found in specific forms of human breast cancer.

RESULTS

Mammary Stem Cells Are Rare Early but Increase Dramatically during Late Fetal Mammogenesis

We used limiting dilution transplantation analyses (LDTA) to measure mammary repopulating unit (MRU) frequency by using single-hit Poisson statistical analyses (Bonnefoix et al., 1996; Hu and Smyth, 2009; Stingl et al., 2006). Accordingly, fetal mammary repopulating units (fMRU) must contain an fMaSC but may additionally require other cells or components for mammary outgrowth. However, as a single stem cell can generate a mammary gland (Shackleton et al., 2006), we will refer to fMaSC frequency when quantifying fMRUs.

We transplanted limiting dilutions of rudiment-derived cell suspensions obtained at different developmental stages of CD1 embryos into immune compromised CB17-SCID recipients (Figure 1A, Table 1, and Table S1, available online) and in parallel transplanted intact mammary rudiments including surrounding mesenchyme (Figures S1A and S1D). Single intact mammary rudiments from as early as E12.5 transplanted at frequencies >70%. In contrast, measurable fMaSC activity was not evident in dissociated bulk mammary cell populations prior to E15.5 (Figure 1A, Table S1D, Table 1, column 2, Table S1, column 3, and Table S2). This was surprising as we transplanted 10,000 viable cells (epithelial and mesenchymal) per fat pad, which is more than we estimate to be present in a single rudiment (Figure 1A, Table 1, and Table S1, column 3). While rudiment dissociation could have reduced repopulation efficiency, especially if contextual cues analogous to niche interactions (Spradling et al., 2001) are required for stem cell function, identically dispersed E18.5 rudiment cells routinely generated mammary outgrowths from 100 cells (Table S1, column 3). Thus, fMaSC concentration increases dramatically during fetal development (Figure 1A; Table 1, column 2, and Table S1, column 3).

The importance of extracellular cues in stem cell function (Spradling et al., 2001) prompted us to determine whether Matrigel would increase fMaSC transplantation efficiency as it does aMaSCs (Lim et al., 2009). While Matrigel significantly increased the sensitivity of fMaSC detection, it still equated to one or fewer stem cells per E13.5 rudiment (1/12,000 cells; Figure 1A, Table 1, column 3, and Table S1, column 4). fMaSC activity peaked at E18.5, increasing to 1 in 60 cells with Matrigel, a 14-fold increase compared to transplantation in its absence (Figure 1A, Table 1, and Table S1). Thus, fMaSC concentration increases at least 200-fold between E13.5 to E18.5 when measured in the presence of Matrigel (Figure 1A and Table 1, column 3). This reflects a striking 9-fold increase in fMaSC abundance during the narrow developmental window between E15.5–E16.5 (1 in 1,800 to 1 in 200, $p < 0.001$) (Figure 1A, Table 1, and Table S1).

These quantitative measurements of stem cell frequency during embryogenesis reveal for the first time the surprising finding that transplantable fMaSCs are rare early in mammogenesis but are abundant late. Furthermore, the fMaSC frequency in unfractionated E18.5 embryonic mammary rudiments is ~5-fold higher than that obtained with bulk adult mammary cell populations measured under identical conditions (1 in 60 versus 1 in 300, $p < 0.001$) (Table 1 and Table S1). These differences may, in part, reflect different contaminating epithelial and stromal components within the cell populations obtained at different stages. Nevertheless, the data indicate that mammary stem cell frequency is higher in the E18.5 mammary rudiment than in the adult gland. This facilitates their purification for subsequent molecular analyses.

Stem Activity Is Restricted to a Unique Fetal Population Expressing High Levels of CD24 and CD49f

aMaSCs can be enriched with fluorescence activated cell sorting for surface markers such as CD24 (heat stable antigen [HSA]) and CD49f (α6 integrin) (Stingl et al., 2006). Both proteins were expressed in the stromal and epithelial compartments of E13.5 and E15.5 mammary rudiments (Figure 1B). In contrast, high CD24 and CD49f expression at E18.5 identifies a basal-epithelial compartment with negligible staining in the fat pad and surrounding mesenchyme, suggesting the utility of these markers for fMaSC enrichment (Figure 1B). Consistent with

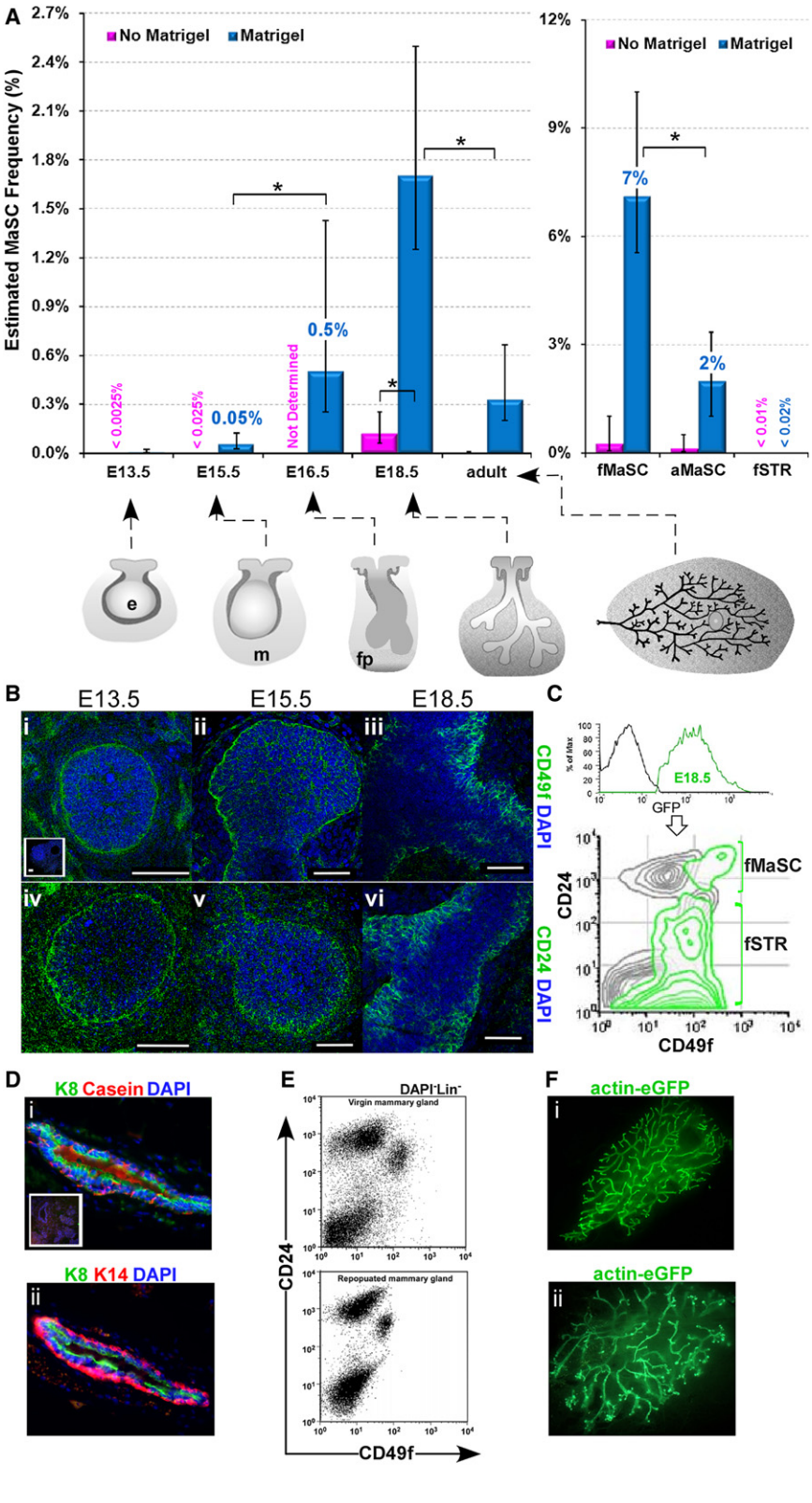

Figure 1. fMaSCs Identified in Late Embryogenesis Express High Levels of CD24 and CD49f

(A) Mammary stem cell frequency estimates at various stages of fetal development and in the adult in the presence or absence of Matrigel. Gross morphological appearance of the gland at various stages is illustrated. The following abbreviations are used: e = epithelium, m = mesenchyme, and fp = fat pad. *p < 0.001, pairwise group difference. The error bars indicate 95% confidence intervals (CI).

(B) Confocal images showing CD49f (i–iii) and CD24 (iv–vi) expression in whole mounts at E13.5, E15.5, and E18.5.

(C) Histogram and FACS contour plot showing the distribution of cells expressing CD24 and CD49f in the LIN$^-$ population (DAPI$^-$CD31$^-$CD45$^-$TER119$^-$) in mammary glands from a nulliparous adult mouse (black) and actin-eGFP E18.5 female embryos (green). Adult eGFP$^-$ mammary and eGFP$^+$ E18.5 fetal mammary cell suspensions were mixed, costained, and analyzed together.

(D) Immunofluorescence analysis of paraffin sections of a regenerated mammary gland from a parous recipient showing casein/K8 (i) and K14/K8 (ii). The inset shows a secondary antibody control.

(E) Representative FACS dot plots showing very similar patterns of expression of CD24 and CD49f in viable lineage-depleted mammary cells from a nulliparous adult mouse (top) and from a mammary gland regenerated by the fMaSCs (bottom).

(F) Representative whole mount of actin-eGFP mammary outgrowth arising from transplantation of the fMaSC population (Lin$^-$CD24^{high}CD49f^{high}) isolated from E18.5 embryos. The mammary glands were harvested from primary (i) and secondary (ii) recipients 12 weeks after transplantation. See also Table 1, Table S1, and Figure S1.

this, flow cytometry demonstrated that most E13.5 and E15.5 rudiment cells expressed CD24 and CD49f (Figure S1B). At E18.5, however, these markers delineate a distinct subpopulation comprising approximately 5% of the cells following the exclusion of endothelial and hematopoietic lineages (Figure 1C and Figure S1B). We then compared the CD24 and CD49f staining profiles of admixed and coprocessed eGFP$^-$ adult and eGFP$^+$ E18.5 fetal mammary cells (Figure 1C). Cells derived from E18.5 rudiments express higher levels of CD24 than the previously reported CD24^{med}CD49f^{high} aMaSC-enriched population and higher levels of CD49f than the CD24^{high}CD49f^{low} adult luminal CFC population (colony-forming cells) (Figure 1C) (Stingl et al., 2006). This direct comparison with adult mammary cells shows that the fetal population is CD24^{high}CD49f^{high}.

The CD24^{high}CD49f^{high} subpopulation contained all fMaSC activity by transplantation analyses (Figure 1A and Table 1).

Table 1. Fetal Mammary Rudiments in Late Embryogenesis Are Found to Be Highly Enriched with Mammary Stem Cells

	MaSC Frequency (95% CI)	
Bulk Mammary Cells	No Matrigel	Coinjected with Matrigel
E13.5[a]	**<1/40,000**	**1/12,000** (1/5000 to 1/30,000), p = 0.2
E15.5[a]	**<1/4,000**	**1/1,800** (1/800 to 1/4,000), p = 0.9
E16.5[a]	ND	**1/200** (1/70 to 1/400), p = 0.68
E18.5[a,b]	**1/830**[c] (1/400 to 1/1,600), p = 0.8	**1/60**[c] (1/40 to 1/80), p = 0.7
Adult[b]	**1/30,000**[d] (1/11,000 to 1/80,000), p = 0.2	**1/300** (1/150 to 1/500), p = 0.2
Fetal MaSC population ($CD24^{high}CD49f^{high}$)	**1/400**[d] **(1/100 to 1/1,700),** p = 0.2	**1/14**[c] (1/10 to 1/18), p = 0.7
Fetal fSTR population ($CD24^{med/low/neg}$)	**<1/9,000**	**<1/5,000**[e]
Adult MaSC population ($CD24^{med}CD49f^{high}$)	**1/800**[d] **(1/200 to 1/2,700),** p = 0.4	**1/50** (1/30 to 1/100), p = 0.98[e]

p > 0.05 for frequency estimates indicates the data are consistent with a single-hit Poisson model. ND is an abbreviation for not determined.
[a] p < 0.001, pairwise group difference.
[b] p < 0.001, pairwise group difference.
[c] p < 0.001, pairwise group difference.
[d] These values represent rough MaSC frequency estimate (see Statistical Analyses for detail).
[e] p < 0.001, pairwise group difference.

The remaining fetal population exhibits lower CD24 levels, is devoid of fMaSC activity, and is enriched in stromal cells as defined by cellular morphology and protein and gene expression analyses (see below). Therefore, we call this population the fetal stroma-enriched population (fSTR). As few as five to ten fMaSCs reproducibly enabled robust mammary gland repopulation, while up to 3,000 cells from the fSTR consistently failed to generate outgrowths, even with Matrigel addition (Figure 1A, Table 1, and Table S1). We estimate the stem cell frequency in the fMaSC-enriched population to be 1 in 14 with Matrigel (Table 1, Table S1, and Figure 1A), which is an ~4-fold enrichment over the aMaSC frequency when immune-compromised hosts were used in the presence of Matrigel (1/50 [aMaSC] versus 1/14 [fMaSC], p < 0.001; Table 1 and Table S1). Importantly, the aMaSC frequency in Matrigel when allotransplantation into immune-compromised hosts was used was similar to that obtained with an immune competent, syngeneic model (C57BL6) in the absence of Matrigel (Tables S1 and S3), and host immune competence did not significantly affect fMaSC frequency in the presence of Matrigel (Table S3).

fMaSCs generated morphologically normal, fully arborized ductal structures that produced casein-positive alveolar structures upon induction of pregnancy (Figure 1D). Mammary outgrowths exhibited the expected localized expression of luminal and myoepithelial keratins (K8 and K14, respectively) (Figure 1D), the phenotypic cellular heterogeneity of wild-type adult mammary glands (Figure 1E), and contained cells able to self-renew based on serial transplantation analyses (Figure 1F). Thus, the fMaSC-enriched population exhibits the multi-lineage cell differentiation and self-renewal characteristics expected of mammary stem cells, but at considerably higher concentration than found in the adult (Shackleton et al., 2006; Stingl et al., 2006).

fMaSCs Are Multipotent and Coexpress Markers of Multiple Mammary Lineages

We evaluated the ability of individual fMaSCs to generate multiple lineages in vitro (Dontu et al., 2003). While fMaSCs had negligible sphere-forming efficiency (SFE) with a conventional nonadherent sphere-forming protocol at low seeding density (1,000 cells per cm^2) (SFE ~0.1%; Figures 2A and 2B), fSTR formed numerous spherical clusters under identical conditions (SFE = 1.4%) (Figures 2A and 2B). However, just as Matrigel profoundly increased transplantation efficiency, even low percentages of Matrigel (2%) enabled the fMaSC population to generate spheres with an SFE of 9.4% when plated at low density (Figures 2A and 2B). The primary fMaSC-derived spheres were morphologically similar to the colonies previously reported for the aMaSC population (Figure 2A) (Stingl et al., 2006). In addition, fMaSC-derived primary spheres expressed markers associated with both myoepithelial (e.g., cytokeratin 14 [K14]) and luminal (cytokeratin 8 [K8]) epithelial lineages of the mammary gland (Figure 2C).

We used two independent strategies to determine whether primary spheres arise from clonal expansion from a single cell or from cell aggregation. First, we seeded single cells from the fMaSC population into individual wells. Primary spheres formed (SFE = 10.7%) in 2% Matrigel, similar to the 9.4% SFE observed when the cells were plated at low density (Figure 2B and Figure S2A). Secondary and tertiary spheres were also formed with similar SFEs (~10%; Figure 2B). The fSTR did not generate spheres in the presence of Matrigel and instead produced cultures of dispersed cells resembling fibroblasts and neurons (Figures 2Aii and 2B). Second, we mixed single-cell suspensions of $eGFP^+$ and $eGFP^-$ cells from the fMaSC population and then grew them at low density (Figure 2B and Figure S2B). As 199 out of 200 spheres were a single color, the vast majority must derive from single cells (Figure S2C). Approximately 60% of fSTR-derived clusters were overtly polyclonal when cultured in nonadherent conditions without Matrigel (Figure 2Aiii), indicating they arise by aggregation. Taken together, these data show that ~10% of the fMaSC population exhibits the stem cell properties of multipotent differentiation and self-renewal in vitro (Figures 2B and 2C).

Coexpression of proteins associated with multiple lineages has been proposed to indicate plasticity in the normal mammary gland and in breast cancers (Creighton et al., 2009; Livasy et al., 2006; Petersen and Polyak, 2010; Sun et al., 2010; Thomas et al., 1999). We detected cells that coexpress K14 and K8 from

as early as E13.5 in the developing mammary gland (Figure S2D and S2E). Approximately 30% of the cells within the fMaSC population were $K14^+K8^+$, and we frequently detected such double-positive cells in fMaSC-derived spheres (Figures 2C and 2D). We also analyzed vimentin expression as it has been associated with the myoepithelial and mesenchymal lineages of the normal mammary gland and with aggressive disease when coexpressed with luminal epithelial markers in breast cancer patients (Creighton et al., 2009; Thomas et al., 1999). Approximately 70% of cells within the $K14^+K8^+$ fMaSC population also expressed vimentin (Figure 2D).

Derivation of fMaSC- and fSTR-Specific Gene Expression Signatures

We performed microarray expression analyses on the fMaSC, fSTR, and aMaSC populations to ascertain molecular pathways with potential relevance to fetal mammary development, fMaSC biology, and breast cancer. We obtained reproducible expression profiles from independent biological pools representing each population and identified differentially expressed genes comprising fMaSC, fSTR, and aMaSC signatures (Figure 3A and Table S4; significance analysis of microarrays; false discovery rate (FDR) < 10%; [Tusher et al., 2001]). We identified 869 unique genes more highly expressed in the fMaSC population (the fMaSC signature) than in the fSTR and 812 unique genes more highly expressed in the fSTR population (the fSTR signature) than in the fMaSC. Among the fMaSC signature genes, ~34% were common to both the fMaSC and aMaSC populations when compared to fSTR, but ~40% were significantly overexpressed in the fMaSC relative to the aMaSC (Figure 3A).

We confirmed the differential gene expression patterns between the fMaSC and fSTR populations on a panel of genes selected from putative stem cell, developmental, and cancer relevant pathways (Figures 3B and 3C). Furthermore, high-throughput single-cell qRT-PCR analyses confirmed expression of a partially overlapping selection of 46 genes in individual cells of the fMaSC population (Figure 3D). This approach also verified that individual fetal cells coexpress luminal keratins, myoepithelial keratins, and vimentin (Figure 3D).

Unique Expression Features of fMaSCs and fSTR

Many well-studied genes were found to be expressed in a manner consistent with the cell types analyzed, indicating the validity of the microarray data (Table S4). However, the fMaSC and fSTR signatures revealed unique gene expression patterns when compared to adult mammary populations or to those isolated earlier in development that we showed to be lower in stem cell content than the fMaSC population (Figures 1 and 4, Table 1, and Tables S4 and S5). For example, qRT-PCR analysis showed significant differences in expression of specific stem cell- and development-related genes between the fMaSC and the E15.5 rudiment (Figure 4A). The gene content in the E18.5 fMaSC and fSTR signatures were also significantly different from those reported for either mouse or human adult mammary populations, or from E12.5 mouse mammary epithelia (ME) and mammary mesenchyme (MM) (Figures 4B–4D and Tables S4 and S5) (Kendrick et al., 2008; Lim et al., 2009; Lim et al., 2010; Pece et al., 2010; Stingl et al., 2006; Wansbury et al., 2011). Although the similarities between the fMaSC and previously reported primary mammary epithelial signatures are statistically significant, the majority of genes in the fMaSC signature are not represented in aMaSC signatures (Figure 4B). Surprisingly, the fSTR signature is similar to the adult mammary stromal signature and to published aMaSC signatures (Figure 4B).

We delineated genes specific to the fMaSC and fSTR populations by comparing their signatures to composite human/mouse adult MaSC or stromal gene lists and to E12.5 ME and MM signatures (Figure 4C). The sets of genes specific to the fMaSC and fSTR signatures are rich in biological content as indicated by their significant correlation with numerous gene ontology (GO) categories (Figure 4D, Figure S4, and Tables S4 and S6). Among these, cell signaling and genes associated with the plasma membrane figure prominently (Figure 4D). A selection of the most highly enriched clusters are detailed in Figure 4D, including several genes previously implicated in mammary stem cell function and breast cancer, such as *ErbB2* and *ErbB4* (Korkaya and Wicha, 2009). Several genes reciprocally expressed in the fSTR and fMaSC populations are suggestive of paracrine signaling and may contribute to stem cell behavior in normal or neoplastic growth. For instance, the fSTR specific signature includes *Nodal* and *Wnt5a*. Nodal is a TGF-β family morphogen that can promote oncogenic phenotypes in mammary cells and has been implicated in breast and other cancers (Strizzi et al., 2009). The Wnt5a protein is a noncanonical Wnt implicated in polarity, migration and stem cell maintenance (Kikuchi et al., 2012). Additional processes found in both the fMaSC and fSTR populations are likely to contribute to the unique properties of mammary cells at this stage. For example, changes in chromatin regulation, augmented synthetic metabolism and cell cycle, and the production of distinctive extracellular matrices may contribute to the robust fMaSC function we observe (Figure S4 and Table S4).

Cellular Interaction Is a Predicted Hallmark of fMaSC Function

GO enrichment analysis of fetal signatures suggested a prominent role for cell-cell and cell-niche interactions, including cell surface receptor signaling in the fMaSC population (Figure 4 and Table S4). We used curated interaction networks in the GeneGo pathway analysis platform to organize the genes comprising the fMaSC and fSTR signatures into potential receptor-ligand interactions. This enabled construction of a hypothetical interaction map based on reported receptor-ligand interactions (Figure 5A).

We determined whether predicted pathways are relevant for fMaSC function in vitro (Figures 5B–5D and Figure S5). We analyzed ErbB and FGF receptors and their ligands given their cancer relevance and that growth of adult mammary epithelial cells in vitro requires either EGF or FGF (Dontu et al., 2003). qRT-PCR validated the differential expression of all four ErbB family members, and the hormone receptors ER and PR (Figure S5A, and data not shown). *ErbB4* was expressed at a low level but exclusively in the fMaSC population (data not shown). *ErbB2* and *ErbB3* were expressed more highly in the fMaSC population than either the fSTR or aMaSC populations (Figure S5A), and ErbB2 protein was detected in situ in CD24+ cells in E18.5 mammary rudiments (Figure S5B).

We examined the requirement for ErbB and FGF signaling by growing fMaSC-derived spheres in 2% Matrigel culture

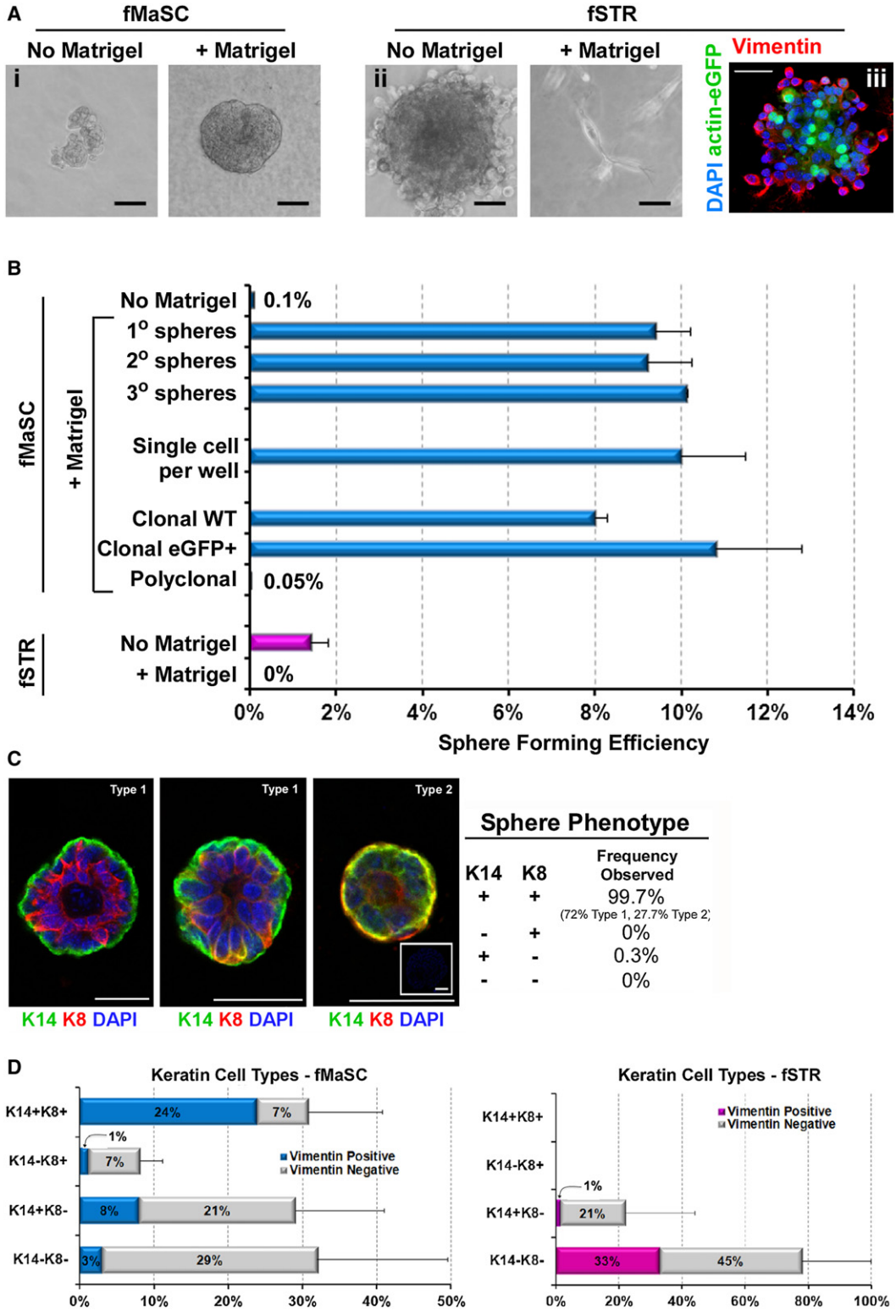

Figure 2. Individual Cells from the fMaSC Population Generate Clonal, Multilineage Spheres that Can Be Serially Propagated and Coexpress Markers of Multiple Lineages

(A) Morphology of structures generated from fMaSC (i) and fSTR (ii) populations grown under nonadherent conditions in vitro in the presence and absence of Matrigel. (iii) Confocal image of an fSTR polyclonal sphere derived by mixing fSTR cells from WT and actin-eGFP transgenic embryos showing vimentin immunofluorescence (red), nuclear counterstain DAPI (blue), and actin-eGFP (green). The scale bar represents 50 μm.

containing or lacking specific ErbB ligands and FGF. Cultures lacking both EGF and FGF produce no spheres, while either EGF or FGF stimulated fMaSC-derived sphere formation (Figure 5B). Heregulin 1 (Hrg1, neuregulin, Neu differentiation factor), an ErbB ligand with a preference for ErbB3 or ErbB4, stimulated sphere growth in the absence of EGF and FGF (Figure 5B) (Britsch, 2007). The effects of these ligands were additive (SFE ~10% for EGF/FGF/HRG), and Hrg1 showed the most dramatic effect on large sphere production (Figure 5B). By contrast, GDNF, which is not represented in the hypothesized interaction network, did not stimulate sphere formation (Figure 5B).

Consistent with the above results, ErbB and FGFR kinase antagonists inhibited fMaSC-derived sphere growth (Figures 5C and 5D and Figure S5). Lapatinib is a reversible and highly specific ErbB1/2 dual tyrosine kinase inhibitor (Rusnak et al., 2001), while neratinib is an irreversible pan-ErbB kinase inhibitor (Rabindran et al., 2004). Lapatinib and neratinib inhibited sphere growth with a similar dose-dependence as observed for Her2 overexpressing human mammary cells with documented sensitivity to these agents (Figures 5C and 5D) (Wang et al., 2006). As these drugs have nonoverlapping potential off-target effects (Karaman et al., 2008; Rabindran et al., 2004; Rusnak et al., 2001), it is most likely that their effects on sphere growth derive from ErbB pathway antagonism. However, it remains to be determined whether inhibition of one specific ErbB receptor accounts for the observed effects on sphere growth or whether redundancy in this family necessitates inhibition of multiple receptors for effective fMaSC growth antagonism in vitro. Altogether, these findings substantiate the importance of ErbB and FGF signaling in fMaSC-derived sphere growth in vitro, and they indicate the presence of functionally relevant gene content in the microarray-derived fetal mammary signatures.

Molecular Links between Fetal Mammogenesis and Breast Cancer

Cancer-associated genes (*ErbB2*, *Met*, *CXCR4*, etc.) were prominent among the fetal signatures and pathway analyses and in unsupervised gene set enrichment analyses (Figures 4 and 5, Figure S5, and Table S4). Therefore, we determined whether fetal gene expression signatures were enriched in particular human breast cancer intrinsic subtypes by using archival tumor microarray data from two independent compendia and human orthologs of the fMaSC and fSTR signature genes (Figure 6, Figure S6, and Table S6) (Ben-Porath et al., 2008; Prat et al., 2010). Enrichment for the fMaSC signature was concentrated among tumors designated as basal-like, which tend to be poorly differentiated and stem-like (Ben-Porath et al., 2008; Mizuno et al., 2010). In addition, many Her2$^+$ tumors showed significant enrichment for the fMaSC signature (Figure 6A and Figure S6A). Enrichment for the fSTR signature often correlated with tumor subtypes characterized by low proliferation and favorable prognoses (Figure 6A and Figure S6A) (Sørlie et al., 2001). However, claudin-low and metaplastic-like tumors, which have also been suggested to be stem-like (Hennessy et al., 2009; Perou et al., 2010; Prat et al., 2010) were generally enriched for the fSTR signature and depleted for the fMaSC signature (Figure 6A and Figure S6A). Breast cancers showing enrichment for fSTR signatures showed nearly identical enrichment patterns for aMaSC signatures (Figure S6B), consistent with the significant gene overlap of aMaSC and fSTR signatures noted above (Figure 4B).

Previous studies have used signatures derived either directly from breast cancer array data or from specific biological contexts, such as serum stimulation of fibroblasts (simulating wound healing), to classify breast cancers into different tumor types with distinct clinical features (Fan et al., 2006). The fMaSC signature exhibits relatively little overlap with these signatures (6.5% of fMaSC genes shared, Figure 6B and Table S7). While, the signatures compared in Figure 6B have significant representation of ER- and/or proliferation-associated genes (Fan et al., 2006; Wirapati et al., 2008), the fMaSC and fSTR signatures have little representation of proliferation genes (Figure 6B) because this is a characteristic they share, leading to exclusion from their comparative profiles. Furthermore, removal of the few residual proliferation-related genes from the fMaSC and fSTR signatures did not markedly alter the observed tumor enrichments (Figure 6A and Figure S6A). We cannot rule out the possibility that fetal-like molecular programs are also invoked by other proliferative states in the mammary gland, for instance at puberty or pregnancy or during outgrowth of transplanted material. Regardless, the fMaSC and fSTR signatures clearly identify a distinct group of genes associating fetal mammary gland biology and fMaSCs with specific molecularly defined breast cancer subtypes (Figure 6B).

Fetal Gene Subsets Identify Patients with Diverse Prognoses in Archival Tumor Samples

The fMaSC and fSTR signatures can be subdivided into gene expression modules showing coordinated expression across multiple tumors with hierarchical clustering analysis (Figures S6C and S6D). We then correlated these subsignatures with breast cancer grade, progression, subtype, or prognosis (Figure 6C and Figures S6 and S7). We subdivided the fMaSC and fSTR signatures into five and four subsignatures, respectively, by using the 96 genes with the greatest variance in the compendium. The genes comprising these enriched subsets represent diverse biological processes previously implicated in cancer, including immune response (fMaSC-iii, fSTR-ii), cell survival (fMaSC-v), and wounding (fSTR-ii, fSTR-iv) (Figures 6C and 6E and Table S7) (Chang et al., 2004; Perou et al., 2000; Rody

(B) Quantification of clonal, primary fMaSC-derived sphere growth, secondary and tertiary sphere growth, and fSTR-derived sphere growth. The error bars indicate standard deviation (SD).
(C) Confocal immunofluorescence analysis of spheres derived from the fMaSC population showing the expression of K8 (red), K14 (green), or both (yellow) with nuclear counterstain DAPI (blue) and tabular summary of sphere types observed. Type 1 spheres consist of cells expressing either K14 (i.e., sphere periphery) or K8 (i.e., middle of the sphere), while type 2 spheres consist mainly of cells coexpressing K8 and K14 (yellow cells). Inset, secondary antibody control. The scale bars represent 25 μm. The inset shows secondary antibody control (the scale bars represent 50 μm).
(D) Summary of the percentage of cells in the fMaSC and fSTR populations expressing K8, K14, and/or vimentin. The error bars indicate standard deviation. See also Figure S2.

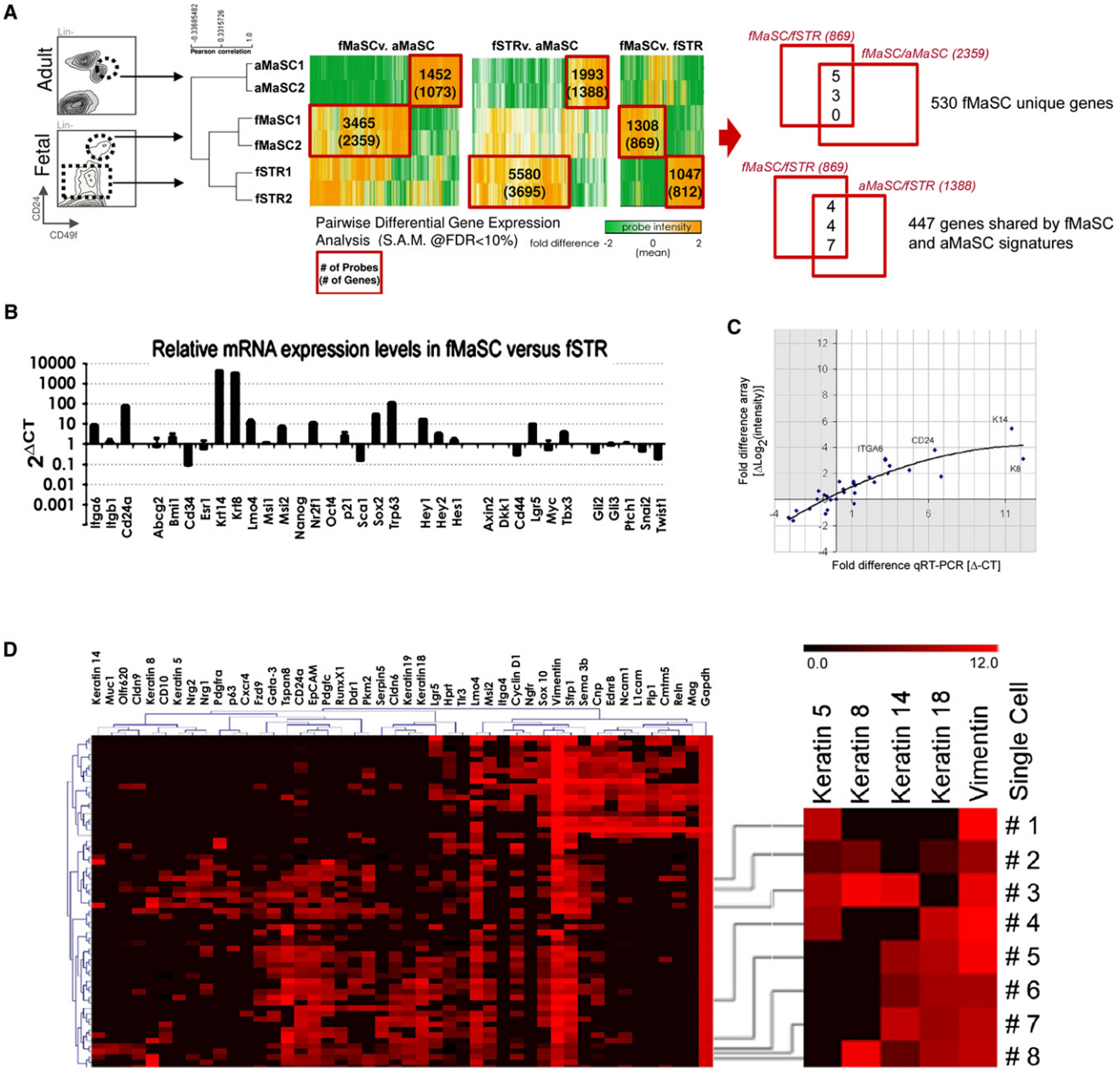

Figure 3. Differential Gene Expression Profiling of fMaSC, fSTR, and aMaSC Populations

(A) Illustration of sorted populations, Pearson correlation among biological replicates for each cell type, and heat maps illustrating the identification of differentially expressed genes (SAM; FDR < 10%).

(B) qRT-PCR analysis of select stem cell and developmental genes in the fMaSC population relative to fSTR. Error bars indicate standard deviation.

(C) Expression levels of a representative selection of genes determined by microarray and by qRT-PCR. The expression level in the fMaSC relative to the fSTR is plotted as the fold difference in expression. Fold differences in gene expression were calculated for RT-PCR assuming ideal amplification (fold change = $2^{\Delta Ct}$) and for Nimblegen array data with the normalized probe intensities (fold change = ΔLog_2(intensity)). Data were normalized to hypoxanthine-guanine phosphoribosyltransferase. Despite differences in the dynamic range of the two techniques, the pattern of differential expression between the fMaSC and fSTR determined by array was consistent with the pattern determined by qRT-PCR.

(D) Microfluidics-based, single-cell, qRT-PCR analyses of cells from the fMaSC population. Right: examples of single cells coexpressing various keratins and the mesenchymal marker, vimentin. See also Table S4.

et al., 2011). Other process such as embryonic morphogenesis (fMaSC-ii, fMaSC-v, and fSTR-iii) and adhesion (fMaSC-iv and fSTR-iii), which have been less extensively investigated in cancer, were also represented (Figure 6E and Table S7).

These fetal subsignatures exhibit prognostic relevance in archival breast cancer array data (Figures 6C and 6D and Figure S7). For instance, enrichment for signature fMaSC-ii or repression of signature fSTR-iv correlated with Her2^{+} and

basal-like tumors, high grade, and reduced probability of patient survival (Figure 6C and Figure S7). This observation is consistent with the predicted outcome of these intrinsic subtypes (Sørlie et al., 2001). In addition, multivariate survival analyses based on enrichment for the fetal subsignatures showed prognostic value beyond commonly used clinical metrics such as ER status, tumor size, grade, and lymph node status (Figure 6D). The biological and prognostic relevance of the signatures described here is a function of their biological origin, as randomized signatures are not enriched in a sufficient number of tumors to enable tumor classification and subsequent survival analysis (Figure 6A and Figure S7). However, it may be possible to derive alternative fetal gene subsignature groupings exhibiting enhanced prognostic value, predictive value, or additional functional biological insight through the use of alternative statistical approaches. As approximately 60% of the genes comprising these fetal subsignatures are specifically upregulated relative to the aMaSC population (Tables S4, S5, and S7), these signatures provide new candidates for therapeutic and prognostic strategies that would probably be missed by deriving signatures from the resting adult gland.

DISCUSSION

The existence of fMaSCs has been inferred from studies demonstrating that intact mammary epithelium obtained from as early as E13.5 can fully reconstitute the mammary gland (Sakakura et al., 1979). However, these studies did not quantify or purify mammary stem cells. This left a substantial gap in our understanding of mammary biology and precluded elucidation of the long predicted molecular and genetic links between fetal mammary development, stem cells, and breast cancer (Howard and Ashworth, 2006). Here, we provide the first quantitative assessment of mammary stem cell activity during fetal mammogenesis, obtain fetal mammary gene expression profiles and evaluate their relationship to breast cancer.

Our studies reveal the surprising finding that mammary rudiments from E15.5 and earlier contain few if any functional fMaSCs. We observed a 200-fold increase in fMaSC activity during the course of fetal mammogenesis that parallels the change in cellular context as the proliferating mammary epithelium begins to invade through the adjacent mesenchyme and interacts with the fat pad microenvironment (Veltmaat et al., 2003). While proliferation during mammogenesis probably contributes to fMaSC abundance, the 9-fold increase in fMaSC frequency between E15.5 and E16.5 is difficult to explain solely by cell division. Instead, we propose that stromal interactions during this interval generate signals that act on precursor cells to engender the stem cell competence we assay by transplantation. It is noteworthy that a recent in vivo lineage tracing study also demonstrated the existence of bipotent mammary stem cells in late embryogenesis and suggested a restriction to unipotent stem/progenitor activity occurring shortly after birth (Van Keymeulen et al., 2011).

The hypothesis that context and extrinsic cues underlie fMaSC functional identity is consistent with studies showing the importance of cell-cell and cell-matrix interactions and locally produced soluble factors for stem cell function (Jones and Wagers, 2008; Spradling et al., 2001). Direct niche interactions also maintain the stem cell state within various tissues and organisms, such as *Drosophila* testes and mammalian hair follicles, bone marrow, testes, and intestines (Spradling et al., 2001). Stem cell niches also produce soluble factors, including Wnt, FGF, TGFβ, and EGF ligands, which promote or maintain the stem cell state (Spradling et al., 2001; Zeng and Nusse, 2010). Importantly, the involvement of multiple ErbB receptors and their ligands in mammary morphogenesis (Jackson-Fisher et al., 2008; Jackson-Fisher et al., 2004; Tidcombe et al., 2003; Wansbury et al., 2011) is consistent with our observations implicating this family in fMaSC function in vitro. The ErbB kinase inhibitor studies reported here and gene knockout studies showing that mammary gland development is impaired to differing degrees in various ErbB knockout mouse strains (Jackson-Fisher et al., 2008; Jackson-Fisher et al., 2004; Tidcombe et al., 2003) suggest that interactions with relevant stromal components and growth factor gradients may be important for inducing stem cell activity during development.

The fMaSC population includes cells that coexpress luminal and myoepithelial markers with vimentin. The expression of vimentin within epithelial cells of the human adult mammary gland is normally restricted to the myoepithelial lineage and has not been reported to occur in concert with luminal keratin expression (Anbazhagan et al., 1998; Mørk et al., 1990). Interestingly, forced coexpression of luminal keratins 8 and 18 with vimentin in human breast cancer cells in vitro increases motility, invasiveness, and proliferation (Hendrix et al., 1997). Similarly, basal-like breast cancers frequently exhibit an undifferentiated phenotype and coexpress myoepithelial and luminal epithelial keratins and vimentin (Livasy et al., 2006). Our data suggest that the coexpression of myoepithelial and luminal keratins and vimentin may typify an uncommitted, embryonic, fMaSC-like state. We suggest that the partial epithelial to mesenchymal transition (EMT) commonly observed during aggressive tumorigenesis may represent a reversion to an embryonic-like state resembling the fMaSC and/or fSTR compartments (Hanahan and Weinberg, 2011). EMT has long been recognized as an essential embryonic process required for development beyond the blastula stage (Hay and Zuk, 1995) and may also promote a stem cell-like state in breast cells (Mani et al., 2008; Thiery and Sleeman, 2006).

fMaSC signatures are derived from cells with a defined biological role and have not been analyzed previously for their relationship to cancer. Other signatures representing biological processes, such as wound healing and immune response, have proven useful for gauging the risk of recurrence in some breast cancer subtypes (Chang et al., 2004; Rody et al., 2011). Thus, we anticipated that our analyses would uncover new genes and pathways related not only to fetal mammary development and fMaSC function but also to breast cancer. Our results suggest that this resource contains new gene sets with prognostic value that may also be useful for predicting which patients will respond to certain treatment strategies. For example, patients receiving ErbB (Her)-targeted therapies, such as herceptin (Trastuzumab) and lapatinib, are selected based on *ErbB2* gene amplification and high-level ErbB2 expression within their tumors (Jacobs et al., 1999). However, in the NSABP-31 clinical trial, some patients confirmed by clinical standards to be ErbB2 negative responded to the ErbB targeted treatment

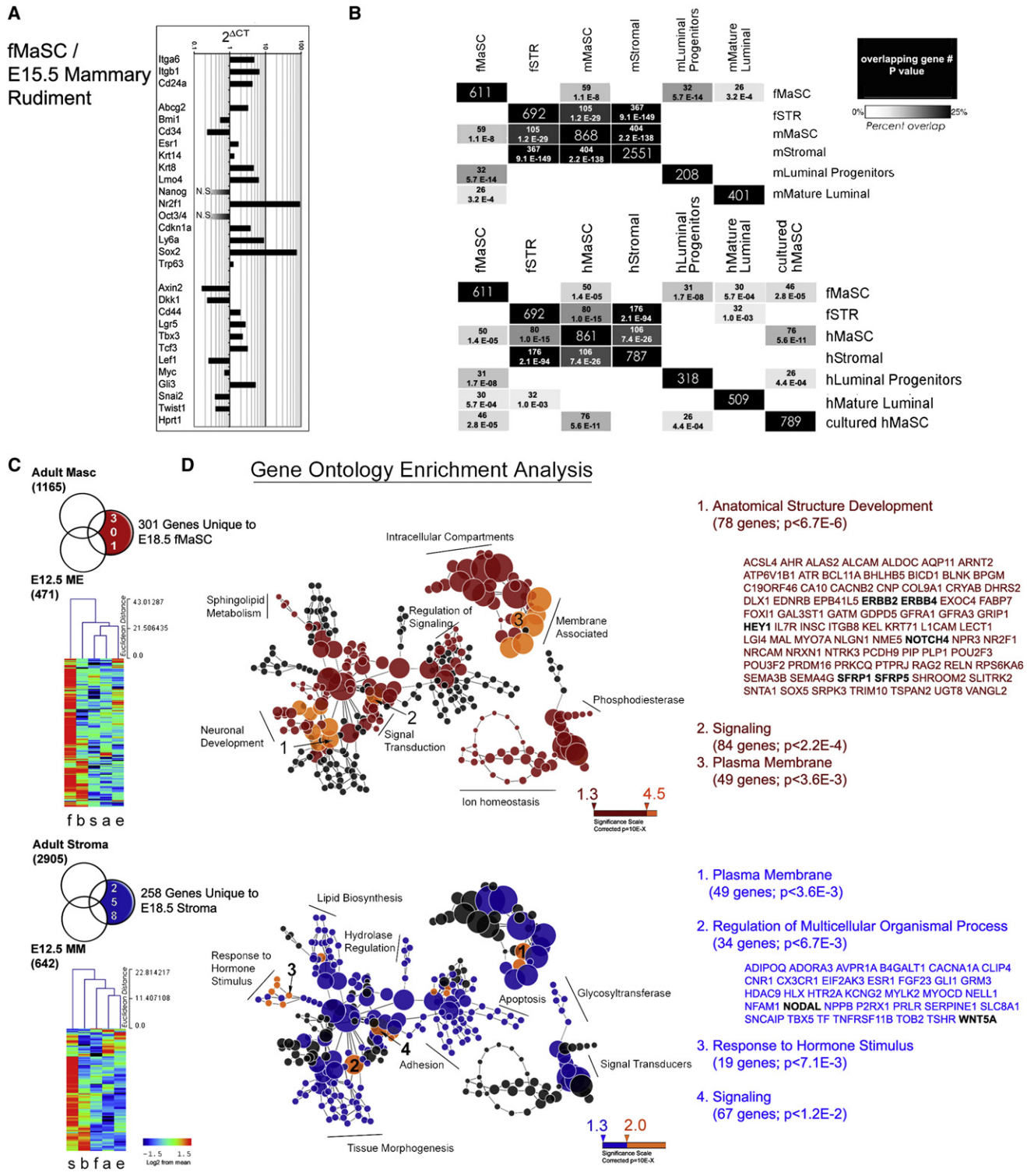

Figure 4. Unique Gene Content in the E18.5 Fetal Mammary

(A) qRT-PCR analysis of select stem cell and developmental genes in fMaSC relative to the E15.5 mammary rudiment. N.S is an abbreviation for no signal.

(B) Overlap of fMaSC signature genes and their orthologs with previously reported normal adult mammary signatures. Upper: mouse signatures (Lim et al., 2009; Pece et al., 2010). Lower: shows hMaSC and hStromal signatures (Lim et al., 2010) and a signature from cultured hMaSCs (Lim et al., 2009; Pece et al., 2010); p values represent the hypergeometric probability based on all 20,309 probes in the mouse array and 19,828 probes in the human arrays.

(C) Identification of genes unique to fMaSC and fSTR populations (Venn diagrams) and clustering of expression array data for these genes for fMaSC (f), fSTR (s), aMaSC (a), E15.5 mammary rudiments (b, buds), and lineage-depleted adult mammary epithelium (e) (heat maps).

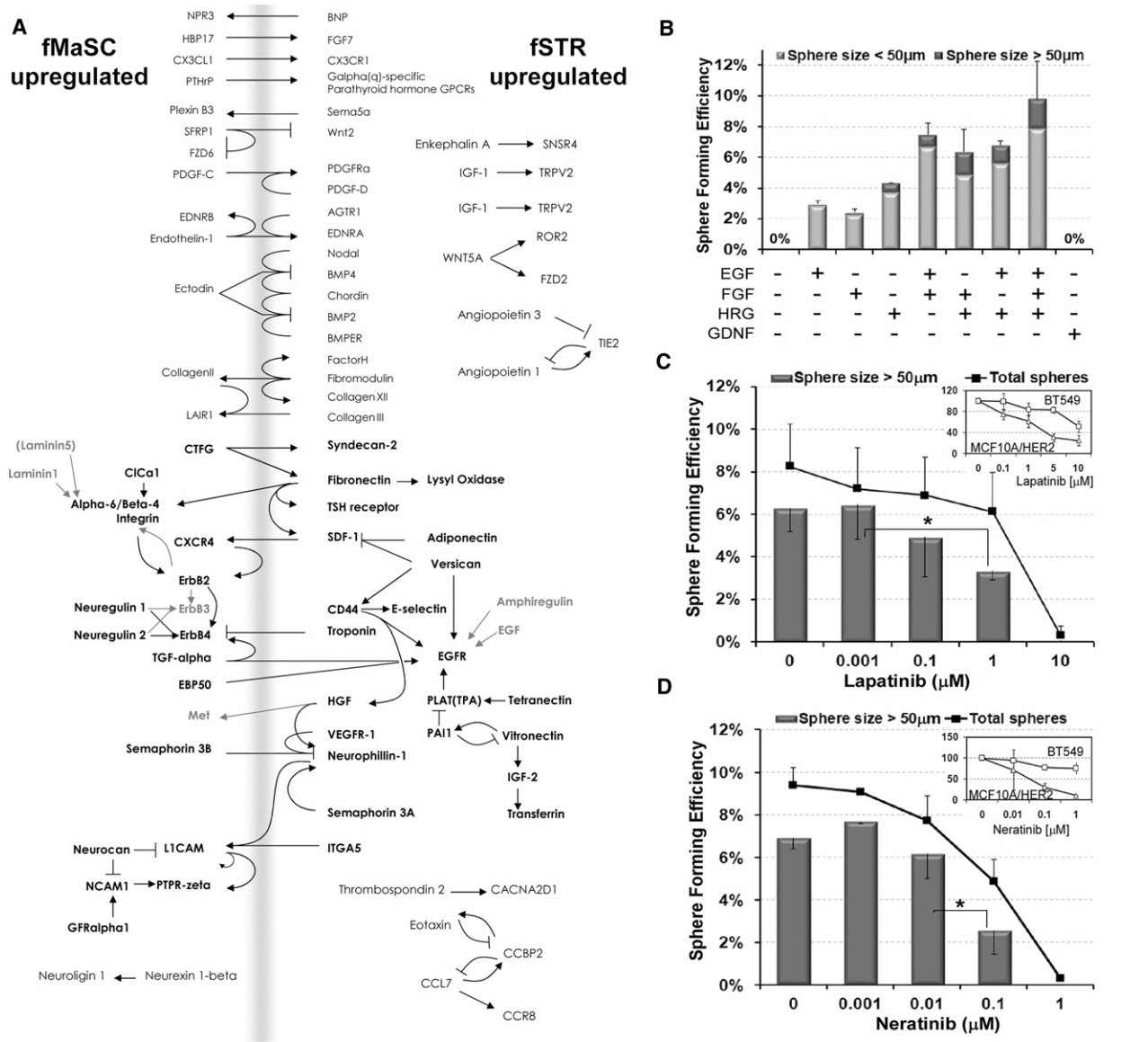

Figure 5. Prediction and Validation of Nonautonomous Signaling in fMaSC Function

(A) A model constructed from fetal gene signatures filtered for receptors and ligands with the GeneGo pathway analysis platform. The model illustrates candidate protein-protein interactions including receptor-ligand pairs expressed reciprocally in the fMaSC (left) and fSTR (right) populations. Additional gene products of interest predicted to interact with the network are also indicated (gray). The map suggests that ErbB signaling, among other pathways, may play a prominent role in fMaSC function.

(B) Quantification of fMaSC-derived spheres in the absence and presence of growth factors suggested by the model in (A).

(C) Quantification of fMaSC-derived sphere growth upon inhibition of ErbB1/2 signaling by either lapatinib or inhibition of ErbB1/2/4 signaling by neratinib. *$p < 0.05$, Student's t test.

(D) Dose-response curves to lapatinib and neratinib in resistant human BT549 and sensitive MCF10A/HER2 cell lines (Wang et al., 2006; Weigelt et al., 2010). All error bars indicate SD. See also Figure S5.

regimen (Paik et al., 2008). Our results show that fMaSCs, which would probably also be designated as ErbB2 negative with accepted clinical guidelines, are sensitive to ErbB pathway inhibitors. We speculate that tumors acquiring an fMaSC-like state will rely on ErbB pathway signaling and, therefore, be sensitive to ErbB antagonists despite being clinically designated as

(D) Gene ontology enrichment analysis of genes unique to the fMaSC and fSTR signatures. Each globe represents an ontological category and the size of the globe represents the number of genes in the category. Significantly enriched categories are color coded in red for fMaSC and blue for fSTR (Benjamini-Hochberg adjusted FDR = 5%). The organic layout algorithm used (Cytoscape) allows visualization of dense ontological data and the observation that many categories are enriched for each signature type. The most highly enriched categories are color coded in orange. The categories with the lowest p values and the gene names contributing to the most enriched "biological process" for each population are listed to the right. See also Figure S4 and Tables S4 and S5.

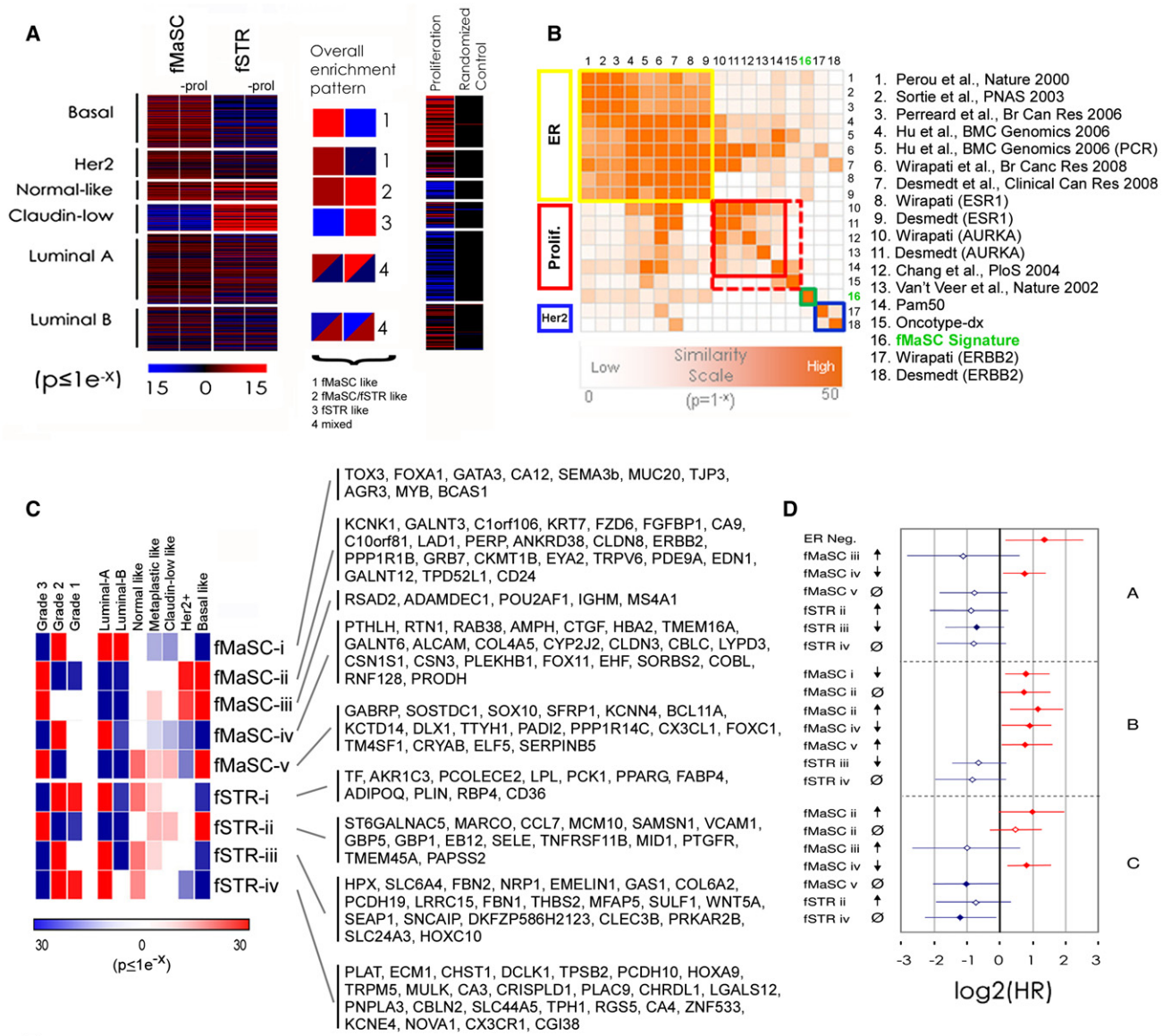

Figure 6. Fetal Mammary Gene Expression Patterns Provide Molecular Links to Human Breast Cancers

(A) Significant correlation between fMaSC and fSTR gene signatures and human breast cancers (n = 337) (Prat et al., 2010) are indicated by horizontal bars, each representing the gene expression profile from an individual tumor sample. Red bars indicate tumors enriched in fetal signature expression; blue bars indicate signature repression. Black bars indicate no significant correlation. Larger colored squares illustrate the trend for each intrinsic subtype. For comparison, a randomized signature of equivalent size and a proliferation signature (Ben-Porath et al., 2008) are shown.

(B) A comparison of several signatures and clinical metrics by significance of gene overlap. Most signatures are closely related and are significantly associated with ER (yellow box) or Proliferation (AURKA; red box) related signatures. Because of its size, the small OncotypeDX signature shows modest significance values for the proliferation group, although it includes several proliferation ER- and Her2-related genes. The fMaSC signature (green box and arrows) is relatively unique and shows no significant overlap with proliferation or ErbB2/Her2-related signatures (blue) and relatively low association with ER-related signatures.

(C) Significance of enrichment for subsignatures among diverse breast cancers in a large microarray compendium (n = 1,211) (Ben-Porath et al., 2008). Enrichments according to subtype and grade are indicated by colored squares that represent probabilities for the percentage of tumors enriched or repressed in each annotation group. Genes comprising each subsignature are listed.

(D) Subsignatures showing significance (◆, $p \leq 0.1$) or trends (◇, $p \leq 0.25$) in multivariate analysis are graphed for models including the following categorical clinical variables: A: ER status, grade, lymph-node status, tumor size; B: grade, lymph-node status and tumor size; C: ER status, lymph-node status and size

ErbB2 negative. The disproportionate representation of proliferation-, ER-, and Her2-related signaling in many existing prognostic signatures may mask less prominent yet critical signaling pathways that can be uncovered by studying normal developmental paradigms such as the fMaSC and fSTR states that are perturbed in cancer.

Links between embryogenesis and tumorigenesis were first proposed in 1838 by Müller as the stem cell origin of cancer and then extended by Durante and Conheim's hypothesis that elements remaining in an undifferentiated embryonic state or that reacquire characteristics of this state generate malignancies (Brewer et al., 2009; Sell, 2010). Subsequent descriptions of onco-fetal proteins, identification of embryonic stem cell genes, splice isoforms, microRNAs, and embryonic metabolism in cancer add credence to this concept (Brewer et al., 2009; Christofk et al., 2008; Powers and Mu, 2008; Sell, 2010). Our identification of a population of fMaSCs and associated stroma with gene expression signatures enriched in different types of breast cancer further support the importance of understanding both components and their interaction during cancer progression.

We suggest that cells resurrecting the programs that govern fetal tissue stem cells and fetal stroma may subsequently fuel tumor progression in the adult. This raises the question of how cells eliciting such programs arise during tumor progression. In some breast cancers, oncogenic lesions, such as loss of p53, may impart developmental plasticity, either directly or through reprogramming of tumor cells to more primitive states, including those resembling fMaSC or fSTR (Mizuno et al., 2010; Spike and Wahl, 2011). In this regard, the gene expression network we report involving both the fMaSCs and their associated stroma will provide a resource for generating new molecular hypotheses linking development and cancer, developing new diagnostic and prognostic metrics, and identifying candidate therapeutic targets.

EXPERIMENTAL PROCEDURES

Mice and Embryos

CD-1 and CB17-SCID were purchased from Charles River. Actin-eGFP mice were maintained on a CD-1 mixed background or pure C57BL/6J (Jackson Labs).

Cell Preparation

Adult mammary glands were dissociated according to the Stem Cell Technologies (SCT) protocol. For fetal mammary glands, collagenase/hyaluronidase digestion time was reduced to 90 min and the trypsin treatment was omitted.

Flow Cytometry

Single-cell suspensions were incubated with 4′,6-diamidino-2-phenylindole (DAPI) and the following antibodies: Fc receptor (2.4G), biotinylated CD31/CD45/TER119 cocktail, CD24-PE (M1/69), CD49f-FITC (all from SCT) and streptavidin-PerCPcy5.5 (BD Biosciences).

Mammary Transplantation

Mammary transplantation (Deome et al., 1959) was carried out with pulled-glass capillaries and mouth pipetting. Transplanted glands were evaluated 6–12 weeks postsurgery.

Immunofluorescence

Whole mounts, paraffin sections, OCT sections, or cytospins were stained with antibodies to: keratin 14 (AF-64, Covance, 1:1,000), keratin 8 (Troma-1, DSHB, 1:100), CD24 (M1/69, BD Biosciences, 1:1,000), CD49f (GoH3, BD Biosciences, 1:1,000), casein (a gift from G. Smith and D. Medina, 1:25), vimentin (AB5733, Chemicon, 1:1,000) and ErbB2 (29D8, Cell Signaling, 1:500).

3D, In Vitro Culture

For suspension mammosphere culture, freshly sorted cells were plated on ultralow-adherence plates (Corning) at 1000 cells/cm^2 in Epicult-B mouse media containing B-supplement, rhEGF, rhbFGF, heparin, and penicillin/streptomycin. For the 2% Matrigel culture, cells were plated on ultralow-adherence plates (Corning) at 1000cells/cm^2 in mammosphere media supplemented with 2% Matrigel (growth factor reduced, BD Biosciences). For the 100% Matrigel culture, freshly sorted cells were seeded on top of a 30 μl bed of Matrigel (growth factor reduced) in 2% Matrigel media. Vehicle (DMSO), lapatinib (LC Laboratories), neratinib (HKI-272, Pfizer) or FGFR inhibitor (PD173074, gift from the Verma lab, Salk Institute) were added at the indicated doses. For the clonal sphere culture, single cells were sorted into 96-well, low-adherence plates at a single cell per well density in 2% Matrigel media. For the eGFP$^+$ and eGFP$^-$ mixing experiments, The eGFP$^+$ and eGFP$^-$ fMaSC and fSTR populations were mixed in a 1:1 or 1:4 ratio. fMaSCs were seeded at low density on top of Matrigel (growth factor reduced) in 2% Matrigel media and fSTR cells were seeded at both low and high densities in mammosphere media.

Microarray and Bioinformatic Analysis

RNA was linearly T7-amplified, and gene expression was measured with a Nimblegen Array (12x135k MM9; Roche Nimblegen). The data are available at the gene expression omnibus at http://www.ncbi.nlm.nih.gov/ under accession GSE27027. Data were RMA normalized and processed with Excel, TIGR-MeV, Genomica, Cytoscape and MedCalc softwares, and the DAVID website. Differential expression was determined with SAM at FDR < 10% (Tusher et al., 2001). Detailed experimental and statistical methods accompany this manuscript as supplemental information.

ACCESSION NUMBERS

The gene expression omnibus accession number for the microarray expression data reported in this paper is GSE27027.

SUPPLEMENTAL INFORMATION

Supplemental Information include six figures and eight tables and can be found with this article online at doi:10.1016/j.stem.2011.12.018.

ACKNOWLEDGMENTS

We dedicate this work to Evelyn Lauder, visionary founder and inspirational leader of the Breast Cancer Research Foundation, and Corinne Mentzelopoulos for generously supporting the Salk Institute. We thank J.C. Izpisua Belmonte, T. Bonnefoix, G. Cunha, C. Eaves, P. Eirew, L. Ellies, P. Gray, J. Green, S. Kinkel, A. Legler, D. Medina, C.M. Perou, A. Prat, A. Rodrigues, B. Vonderhaar, and Wahl Lab members for invaluable advice; Pfizer for neratinib; and BCRF, DOD-BCRP, Susan G. Komen for the Cure, the G. Harold and Leila Y. Mathers Charitable Foundation, T32 training grants CA009523 (B.T.S.), GM07240 (D.D.E.), CA009370 (J.C.L.), and California Breast Cancer Research Program 15GB-0015 (D.D.E.) for funding; and National Cancer Institute 5P30CA014195 for core resource support.

Received: June 11, 2011
Revised: November 22, 2011
Accepted: December 22, 2011
Published online: February 2, 2012

(NKI295). A positive (or red) value indicates a poorer prognosis, while a negative (or blue) value indicates a better prognosis. The associated hazard for ER-negative tumors is shown in model A (versus fMaSC-i) for comparison. The following symbols are used: ↑ = subsignature enrichment; ↓ = subsignature repression; Ø = no significant signature enrichment and/or depletion. Error bars indicate 95th percentile CI.
(E) Biological functions associated with gene constituents of the subsignatures (gene set enrichment $p < 0.05$). See also Figures S6 and S7 and Tables S6–S8.

REFERENCES

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA *100*, 3983–3988.

Anbazhagan, R., Osin, P.P., Bartkova, J., Nathan, B., Lane, E.B., and Gusterson, B.A. (1998). The development of epithelial phenotypes in the human fetal and infant breast. J. Pathol. *184*, 197–206.

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.

Bonnefoix, T., Bonnefoix, P., Verdiel, P., and Sotto, J.J. (1996). Fitting limiting dilution experiments with generalized linear models results in a test of the single-hit Poisson assumption. J. Immunol. Methods *194*, 113–119.

Brewer, B.G., Mitchell, R.A., Harandi, A., and Eaton, J.W. (2009). Embryonic vaccines against cancer: An early history. Exp. Mol. Pathol. *86*, 192–197.

Britsch, S. (2007). The neuregulin-I/ErbB signaling system in development and disease. Adv. Anat. Embryol. Cell Biol. *190*, 1–65.

Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., Montgomery, K., Chi, J.T., van de Rijn, M., Botstein, D., and Brown, P.O. (2004). Gene expression signature of fibroblast serum response predicts human cancer progression: Similarities between tumors and wounds. PLoS Biol. *2*, E7.

Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., Fleming, M.D., Schreiber, S.L., and Cantley, L.C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature *452*, 230–233.

Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I., et al. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc. Natl. Acad. Sci. USA *106*, 13820–13825.

Cunha, G.R., and Hom, Y.K. (1996). Role of mesenchymal-epithelial interactions in mammary gland development. J. Mammary Gland Biol. Neoplasia *1*, 21–35.

Deome, K.B., Faulkin, L.J., Jr., Bern, H.A., and Blair, P.B. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. Cancer Res. *19*, 515–520.

Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin. Cancer Res. *14*, 5158–5165.

Dontu, G., Al-Hajj, M., Abdallah, W.M., Clarke, M.F., and Wicha, M.S. (2003). Stem cells in normal breast development and breast cancer. Cell Prolif. *36* (*Suppl 1*), 59–72.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A.B., van't Veer, L.J., and Perou, C.M. (2006). Concordance among gene-expression-based predictors for breast cancer. N. Engl. J. Med. *355*, 560–569.

Fan, C., Prat, A., Parker, J.S., Liu, Y., Carey, L.A., Troester, M.A., and Perou, C.M. (2011). Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. BMC Med. Genomics *4*, 3.

Haibe-Kains, B., Desmedt, C., Piette, F., Buyse, M., Cardoso, F., Van't Veer, L., Piccart, M., Bontempi, G., and Sotiriou, C. (2008). Comparison of prognostic gene expression signatures for breast cancer. BMC Genomics *9*, 394.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell *144*, 646–674.

Hay, E.D., and Zuk, A. (1995). Transformations between epithelium and mesenchyme: Normal, pathological, and experimentally induced. Am. J. Kidney Dis. *26*, 678–690.

Hendrix, M.J., Seftor, E.A., Seftor, R.E., and Trevor, K.T. (1997). Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior. Am. J. Pathol. *150*, 483–495.

Hennessy, B.T., Gonzalez-Angulo, A.M., Stemke-Hale, K., Gilcrease, M.Z., Krishnamurthy, S., Lee, J.S., Fridlyand, J., Sahin, A., Agarwal, R., Joy, C., et al. (2009). Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Res. *69*, 4116–4124.

Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., et al. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. *8*, R76.

Howard, B., and Ashworth, A. (2006). Signalling pathways implicated in early mammary gland morphogenesis and breast cancer. PLoS Genet. *2*, e112.

Hu, Y., and Smyth, G.K. (2009). ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods *347*, 70–78.

Jackson-Fisher, A.J., Bellinger, G., Ramabhadran, R., Morris, J.K., Lee, K.F., and Stern, D.F. (2004). ErbB2 is required for ductal morphogenesis of the mammary gland. Proc. Natl. Acad. Sci. USA *101*, 17138–17143.

Jackson-Fisher, A.J., Bellinger, G., Breindel, J.L., Tavassoli, F.A., Booth, C.J., Duong, J.K., and Stern, D.F. (2008). ErbB3 is required for ductal morphogenesis in the mouse mammary gland. Breast Cancer Res. *10*, R96.

Jacobs, T.W., Gown, A.M., Yaziji, H., Barnes, M.J., and Schnitt, S.J. (1999). Specificity of HercepTest in determining HER-2/neu status of breast cancers using the United States Food and Drug Administration-approved scoring system. J. Clin. Oncol. *17*, 1983–1987.

Jones, D.L., and Wagers, A.J. (2008). No place like home: Anatomy and function of the stem cell niche. Nat. Rev. Mol. Cell Biol. *9*, 11–21.

Karaman, M.W., Herrgard, S., Treiber, D.K., Gallant, P., Atteridge, C.E., Campbell, B.T., Chan, K.W., Ciceri, P., Davis, M.I., Edeen, P.T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. Nat. Biotechnol. *26*, 127–132.

Kendrick, H., Regan, J.L., Magnay, F.A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M., and Smalley, M.J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. BMC Genomics *9*, 591.

Kikuchi, A., Yamamoto, H., Sato, A., and Matsumoto, S. (2012). Wnt5a: Its signalling, functions and implication in diseases. Acta Physiol. (Oxf.) *204*, 17–33.

Korkaya, H., and Wicha, M.S. (2009). HER-2, notch, and breast cancer stem cells: Targeting an axis of evil. Clin. Cancer Res. *15*, 1845–1847.

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.

Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M.-L., Vaillant, F., Yagita, H., Lindeman, G.J., Smyth, G.K., and Visvader, J.E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast Cancer Res. *12*, R21.

Livasy, C.A., Reading, F.C., Moore, D.T., Boggess, J.F., and Lininger, R.A. (2006). EGFR expression and HER2/neu overexpression/amplification in endometrial carcinosarcoma. Gynecol. Oncol. *100*, 101–106.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Mizuno, H., Spike, B.T., Wahl, G.M., and Levine, A.J. (2010). Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures. Proc. Natl. Acad. Sci. USA *107*, 22745–22750.

Mørk, C., van Deurs, B., and Petersen, O.W. (1990). Regulation of vimentin expression in cultured human mammary epithelial cells. Differentiation *43*, 146–156.

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., et al. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J. Clin. Oncol. *24*, 3726–3734.

Paik, S., Kim, C., and Wolmark, N. (2008). HER2 status and benefit from adjuvant trastuzumab in breast cancer. N. Engl. J. Med. *358*, 1409–1411.

Pal, S.K., Childs, B.H., and Pegram, M. (2011). Triple negative breast cancer: Unmet medical needs. Breast Cancer Res. Treat. *125*, 627–636.

Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P.G., and Di Fiore, P.P. (2010). Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell *140*, 62–73.

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.

Perou, C.M., Parker, J.S., Prat, A., Ellis, M.J., and Bernard, P.S. (2010). Clinical implementation of the intrinsic subtypes of breast cancer. Lancet Oncol. *11*, 718–719, author reply 720–721.

Petersen, O.W., and Polyak, K. (2010). Stem cells in the human breast. Cold Spring Harb Perspect Biol. *2*, a003160.

Powers, S., and Mu, D. (2008). Genetic similarities between organogenesis and tumorigenesis of the lung. Cell Cycle *7*, 200–204.

Prat, A., and Perou, C.M. (2011). Deconstructing the molecular portraits of breast cancer. Mol. Oncol. *5*, 5–23.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*, R68.

Rabindran, S.K., Discafani, C.M., Rosfjord, E.C., Baxter, M., Floyd, M.B., Golas, J., Hallett, W.A., Johnson, B.D., Nilakantan, R., Overbeek, E., et al. (2004). Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. Cancer Res. *64*, 3958–3965.

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.

Rody, A., Karn, T., Liedtke, C., Pusztai, L., Ruckhaeberle, E., Hanker, L., Gaetje, R., Solbach, C., Ahr, A., Metzler, D., et al. (2011). A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res. *13*, R97.

Rusnak, D.W., Lackey, K., Affleck, K., Wood, E.R., Alligood, K.J., Rhodes, N., Keith, B.R., Murray, D.M., Knight, W.B., Mullin, R.J., and Gilmer, T.M. (2001). The effects of the novel, reversible epidermal growth factor receptor/ErbB-2 tyrosine kinase inhibitor, GW2016, on the growth of human normal and tumor-derived cell lines in vitro and in vivo. Mol. Cancer Ther. *1*, 85–94.

Sakakura, T., Nishizuka, Y., and Dawe, C.J. (1979). Capacity of mammary fat pads of adult C3H/HeMs mice to interact morphogenetically with fetal mammary epithelium. J. Natl. Cancer Inst. *63*, 733–736.

Sell, S. (2010). On the stem cell origin of cancer. Am. J. Pathol. *176*, 2584–494.

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.

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.

Sotiriou, C., and Piccart, M.J. (2007). Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care? Nat. Rev. Cancer *7*, 545–553.

Spike, B.T., and Wahl, G.M. (2011). p53, Stem Cells, and Reprogramming: Tumor Suppression beyond Guarding the Genome. Genes Cancer *2*, 404–419.

Spradling, A., Drummond-Barbosa, D., and Kai, T. (2001). Stem cells find their niche. Nature *414*, 98–104.

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.

Strizzi, L., Hardy, K.M., Seftor, E.A., Costa, F.F., Kirschmann, D.A., Seftor, R.E., Postovit, L.M., and Hendrix, M.J. (2009). Development and cancer: At the crossroads of Nodal and Notch signaling. Cancer Res. *69*, 7131–7134.

Sun, P., Yuan, Y., Li, A., Li, B., and Dai, X. (2010). Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. Histochem. Cell Biol. *133*, 213–221.

Thiery, J.P., and Sleeman, J.P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nat. Rev. Mol. Cell Biol. *7*, 131–142.

Thomas, P.A., Kirschmann, D.A., Cerhan, J.R., Folberg, R., Seftor, E.A., Sellers, T.A., and Hendrix, M.J. (1999). Association between keratin and vimentin expression, malignant phenotype, and survival in postmenopausal breast cancer patients. Clin. Cancer Res. *5*, 2698–2703.

Tidcombe, H., Jackson-Fisher, A., Mathers, K., Stern, D.F., Gassmann, M., and Golding, J.P. (2003). Neural and mammary gland defects in ErbB4 knockout mice genetically rescued from embryonic lethality. Proc. Natl. Acad. Sci. USA *100*, 8281–8286.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. *347*, 1999–2009.

Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. Nature *479*, 189–193.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Veltmaat, J.M., Mailleux, A.A., Thiery, J.P., and Bellusci, S. (2003). Mouse embryonic mammogenesis as a model for the molecular regulation of pattern formation. Differentiation *71*, 1–17.

Wang, S.E., Narasanna, A., Perez-Torres, M., Xiang, B., Wu, F.Y., Yang, S., Carpenter, G., Gazdar, A.F., Muthuswamy, S.K., and Arteaga, C.L. (2006). HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. Cancer Cell *10*, 25–38.

Wansbury, O., Mackay, A., Kogata, N., Mitsopoulos, C., Kendrick, H., Davidson, K., Ruhrberg, C., Reis-Filho, J.S., Smalley, M.J., Zvelebil, M., and Howard, B.A. (2011). Transcriptome analysis of embryonic mammary cells reveals insights into mammary lineage establishment. Breast Cancer Res. *13*, R79.

Weigelt, B., Lo, A.T., Park, C.C., Gray, J.W., and Bissell, M.J. (2010). HER2 signaling pathway activation and response of breast cancer cells to HER2-targeting agents is dependent strongly on the 3D microenvironment. Breast Cancer Res. Treat. *122*, 35–43.

Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schütz, F., et al. (2008). Meta-analysis of gene expression profiles in breast cancer: Toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res. *10*, R65.

Zeng, Y.A., and Nusse, R. (2010). Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. Cell Stem Cell *6*, 568–577.