, ,
, & ,
(2013)
Molecular profiling of human mammary gland links breast cancer risk to a p27(+) cell population with progenitor characteristics.
Cell Stem Cell, *13*(1), pp. 117-130.

Parity and BRCA1/2-associated breast cancer risk linked to hormone-responsive p27^{+} luminal epithelial progenitors

Sibgat Choudhury[1,2,3*], Vanessa Almendro[1,2,3,4*], Reo Maruyama[1,2,3*], Zhenhua Wu[5,6*], Vanessa F. Merino[7*], Ying Su[1,2,3], Filipe C. Martins[1,8,9], Mary Jo Fackler[7], Marina Bessarabova[10], Adam Kowalczyk[11,12,13], Thomas Conway[11,13], Bryan Beresford-Smith[11,13], Geoff Macintyre[11,13], Yu Kang-Cheng[5,6], Zoila Lopez-Bujanda[7], Antony Kaspi[14], Rong Hu[2], Judith Robens[15,16], Tatiana Nikolskaya[10], Vilde D. Haakensen[17], Stuart J. Schnitt[15,16], Pedram Argani[18], Gabrielle Ethington[19], Laura Panos[19], Michael Grant[19], Jason Clark[19], William Herlihy[19], S. Joyce Lin[20], Grace Chew[21], Erik W. Thompson[21,22], April Greene-Colozzi[23], Andrea Richardson[15,23], Gedge D. Rosson[24], Malcolm Pike[25], Yuri Nikolsky[10], Joanne Blum[19], Rulla M. Tamimi[2,6], Franziska Michor[5,6], Izhak Haviv[11,14,26,27], X. Shirley Liu[5,6], Saraswati Sukumar[7], Kornelia Polyak[1,2,3,28]

Departments of [1]Medical Oncology and [5]Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. Departments of [2]Medicine and [23]Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA. Departments of [3]Medicine and [15]Pathology, Harvard Medical School, Boston, MA 02115, USA. [6]Harvard School of Public Health, Boston, MA 02115, USA. [4]Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. Departments of [7]Oncology, [18]Pathology, and [23]Plastic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA. [8]Obstetrics and Gynaecology Department, Coimbra University Hospital, Portugal. [9]Programme for Advanced Medical Education, Portugal. [10]Thomson Reuters Healthcare & Science, Encinitas, CA 92024, USA. [11]NICTA Victoria Research Laboratory, and Departments of [12]Electrical and Electronic Engineering, [13]Computer Science and Software Engineering, [20]Pathology, [21]Surgery, and [26]Biochemistry, The University of Melbourne, Parkville, Victoria, Australia. [14]Bioinformatics and System Integration, Baker IDI Heart and Diabetes Institute, Prahran, Victoria, Australia. [16]Department of Pathology, Beth-Israel Deaconess Medical Center, Boston, MA 02115, USA. [17]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. [19]Baylor-Charles A. Sammons Cancer Center, Dallas, TX 75246, USA. [25]University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA 90089, USA. [22]St. Vincent's Institute, Fitzroy, Victoria, Australia. [27]Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. [28]Harvard Stem Cell Institute, Cambridge, MA 02138, USA.

[*]These authors contributed equally to this work.

Running title: Parity-related changes in the normal breast

Contact: Kornelia Polyak, Dana-Farber Cancer Institute, 450 Brookline Ave., D740C, Boston, MA 02215; Phone: 617-632-2106, Fax: 617-582-8490, Email: Kornelia_Polyak@dfci.harvard.edu; X. Shirley Liu, Dana-Farber Cancer Institute, 450 Brookline Ave., CLS-11022, Boston, MA 02215; Phone: 617-632-2472, Fax: 617-632-2444, Email: xsliu@jimmy.harvard.edu; Saraswati Sukumar, Johns Hopkins University School of Medicine, MD 21231; Phone: 410-614-2479, Fax: 410-614-4073, Email: saras@jhmi.edu.

SUMMARY

Early full-term pregnancy is one of the most effective natural protections against breast cancer; yet the underlying mechanism is poorly understood. To investigate this, we characterized the global gene expression and epigenetic profiles of multiple cell types from normal breast tissue of nulliparous and parous women. We found that the most significant differences are in cells with progenitor features where the levels of many stem cell-related genes and pathways including $p27^{Kip1}$ and TGFβ are lower in parous women except in BRCA1 and BRCA2 mutation carriers. The numbers of $p27^{+}$ cells are also significantly decreased in control parous women, the majority of these cells are estrogen receptor (ER^{+}) and phospho-Smad2 ($pSmad2^{+}$) positive, and their numbers inversely correlate with ovarian hormone-induced proliferation during the menstrual cycle and pregnancy. Inhibition of TGFβ signaling in tissue explant cultures decreased the numbers of $p27^{+}$ cells with a concomitant increase in cell proliferation. Our results imply that $p27^{+}$ mammary epithelial cells are hormone-responsive progenitors, their numbers correlate with breast cancer risk, and targeting of pathways that control their numbers and proliferation can be explored for cancer prevention.

HIGHLIGHTS

- **Parity-related molecular alterations in control but not BRCA1/2 normal breast**
- **Stem cell-related pathways are decreased in progenitors from parous tissues**
- **Number of hormone-responsive $p27^{+}$ progenitors correlate with cancer risk**
- **p27 and TGFβ signaling are key regulators of breast epithelial progenitors**

INTRODUCTION

A single full-term pregnancy in early adulthood decreases the risk of estrogen receptor positive (ER+) postmenopausal breast cancer, the most common form of the disease (Colditz et al., 2004). Age at first pregnancy is critical, as the protective effect decreases after the mid 20s, and women aged >35 at first birth have increased risk of both ER+ and ER- breast cancer. Parity-associated risk is also influenced by germline variants as BRCA1 and BRCA2 mutation carriers do not experience the same decrease in risk reduction as do women in the general population (Cullinane et al., 2005). These human epidemiological data suggest that pregnancy induces long-lasting effects in the normal breast epithelium and that ER+ and ER- tumors might have a different cell-of-origin.

The protective effect of pregnancy is also observed in animal models, where it can be mimicked by hormonal factors in the absence of gestation (Ginger and Rosen, 2003; Russo et al., 2005; Sivaraman and Medina, 2002). The cellular and molecular mechanisms that underlie pregnancy and hormone-induced refractoriness to carcinogens are largely undefined. Several hypotheses have been proposed including the induction of differentiation, decreased susceptibility to carcinogens, a decrease in cell proliferation and in the number of mammary epithelial stem cells, an altered systemic environment due to a decrease in circulating growth hormone and other endocrine factors, and permanent molecular changes leading to alterations in cell fate (Ginger and Rosen, 2003; Russo et al., 2005; Sivaraman and Medina, 2002).

Almost all studies investigating pregnancy-induced changes and the breast cancer-preventative effects of pregnancy have been conducted in rodent models and most of them have focused only on the mammary gland. Global gene expression profiling of mammary glands from virgin and parous rats identified changes in TGFβ and IGF signaling, and in the expression of extracellular matrix proteins (Blakely et al., 2006; D'Cruz et al., 2002). Related studies conducted in humans also identified consistent differences in gene expression profiles between

nulliparous and parous women (Asztalos et al., 2010; Belitskaya-Levy et al., 2011; Russo et al., 2008; Russo et al., 2011). Because all these studies have used total mammary gland or mammary organoids, which are composed of multiple cell types, the cellular origin of these gene expression differences remains unknown.

Emerging data indicate that mammary epithelial progenitor or stem cells are the cell-of-origin of breast carcinomas and factors that influence breast cancer risk may alter the number and/or properties of these cells (Visvader, 2011). Studies assessing changes in mammary epithelial stem cells following pregnancy have been conducted only in mice and thus far have been inconclusive (Asselin-Labat et al., 2010; Britt et al., 2009; Siwko et al., 2008). Thus, the effect of pregnancy on the number and functional properties of murine mammary epithelial progenitors is still elusive and it has not yet been analyzed in humans.

Here we describe the detailed molecular characterization of luminal and myoepithelial cells, lineage-negative (lin-) cells with progenitor features, and stromal fibroblasts from nulliparous and parous women, the identification of cell-type-specific differences related to parity, functional validation of selected parity-related pathways, and the relevance of these to breast cancer risk and survival.

RESULTS

Parity-related differences in gene expression patterns in multiple cell types

To investigate parity-associated differences in the normal human breast, first we defined three distinct mammary epithelial cell populations by FACS (fluorescence-activated cell sorting) for cell surface markers previously associated with luminal epithelial (CD24), myoepithelial (CD10), and progenitor features ($lin^-/CD44^+$) (Bloushtain-Qimron et al., 2008; Mani et al., 2008; Shipitsin et al., 2007) that showed minimal overlap both in nulliparous and parous tissues (Figure S1A-B). The relative fraction of $CD44^+$ cells was slightly higher in parous relative to nulliparous samples, which is likely due to the more developed lobular structures composed of many $CD44^+$ cells in these tissues that we observed in sections (Figure S1B-C). We also performed multicolor immunofluorescence analyses for these cell surface markers and genes specific for luminal (e.g., GATA3) and myoepithelial (e.g., SMA) cells, to further confirm the identity of the cells (Figure S1D).

To investigate parity-related differences in global gene expression profiles, we analyzed immuno-magnetic bead purified (Bloushtain-Qimron et al., 2008; Shipitsin et al., 2007) $CD24^+$, $CD10^+$, and $CD44^+$ cells (captured sequentially, thus, $CD44^+$ fraction is $CD24^-CD10^-CD44^+$), and fibroblast-enriched stroma from multiple nulliparous and parous women using SAGE-seq (Serial Analysis of Gene Expression applied to high-throughput sequencing) (Maruyama et al., 2011; Maruyama et al., 2010; Wu et al., 2010). To minimize interpersonal variability unrelated to parity status, women were closely matched for age, the number of pregnancies, time at first and since last pregnancy, and ethnicity (Table S1). The expression of known cell type-specific genes (e.g., KRT8 and MUC1 luminal, ACTG2 and CNN1 myoepithelial, ZEB2 and TWIST1 progenitor cell markers) was consistently observed in each of the three respective epithelial cell types both from nulliparous and parous samples based on SAGE-seq confirming the purity and identity of the cells (Figure S1E). Comparison of each cell type between nulliparous and parous samples

revealed the most pronounced differences in $CD44^+$ cells, where the number of significantly ($p<0.05$) differentially expressed genes and the fold differences were the largest between groups (Figure 1A and Table S2). The degrees of differences were smaller and similar in $CD10^+$ and $CD24^+$ cells, whereas stromal fibroblasts had the fewest differentially expressed genes (Tables S2). Further examination of parity-related differences in expression patterns using principal component analysis (PCA) confirmed that $CD24^+$ and $CD10^+$ cells and fibroblasts from nulliparous and parous women were highly similar, whereas $CD44^+$ cells formed very distinct nulliparous and parous clusters (Figure 1B and Figure S1F).

To validate our parity-related gene expression differences in an independent cohort, we analyzed the levels of our differentially expressed genes (in all cell types or only in CD44+ cells) in gene expression data from breast biopsies of a cohort of Norwegian women (Haakensen et al., 2011a; Haakensen et al., 2011b) matched to our samples for age (<40) and parity (P2). Clustering analysis using our differentially expressed gene sets divided these samples into a distinct nulliparous and a mixed parous/nulliparous group (Figure SGF). Using genes differentially expressed in all four cell types (i.e., $CD24^+$, $CD10^+$, $CD44^+$ cells, and fibroblasts) combined or only in $CD44^+$ cells gave identical results, supporting our hypothesis that changes in $CD44^+$ cells are the most significant and physiologically relevant. Interestingly, the two nulliparous groups displayed significant differences in serum estradiol levels (SEL) with the group more similar to parous cases having low SEL; all parous samples also had low SEL (Figure S1H). Because these were all premenopausal women and SEL is known to be higher in the luteal phase of the menstrual cycle, when breast epithelial cell proliferation is also higher, our findings imply that breast tissues of nulliparous and parous women may be more distinct in the luteal phase potentially due to differences in the activity of signaling pathways driving cell proliferation or the number of cells that respond to these stimuli.

The expression of selected genes was also validated in additional samples by quantitative

RT-PCR (qRT-PCR) using $CD44^+$ cells from multiple nulliparous and parous cases. Despite some interpersonal variability, we detected statistically significant differences between nulliparous and parous groups that overall correlated with SAGE-seq data (Figure S1I).

Based on these findings in gene expression profiles, we focused our follow up studies on $CD44^+$ and $CD24^+$ cells.

Lack of parity-associated differences in BRCA1 and BRCA2 mutation carriers

To strengthen our hypothesis that the parity-associated differences we detected in $CD44^+$ cells might be related to subsequent breast cancer risk, we analyzed the gene expression profiles of $CD44^+$ cells from parous BRCA1 and BRCA2 mutation carriers, whose risk is not decreased by parity (Cullinane et al., 2005). $CD44^+$ cells from parous BRCA1/2 mutation carriers were highly related to $CD44^+$ cells from nulliparous controls (Figure 1C), suggesting that parity-associated changes observed in control parous women may not occur in these high risk women.

To determine if the lack of parity-associated changes in CD44+ cells from BRCA1/2 women could be due to differences in the cell populations identified by the three cell surface markers, we performed FACS analysis of multiple tissue samples from control and BRCA1/2 women. The relative ratio of $CD44^+$ to $CD24^+$ cells was slightly higher in control and BRCA1/2 parous compared to nulliparous control samples, whereas the relative frequency of $CD10^+$ cells was about the same in all groups (Figure 1D). The increase in the relative frequency of $CD44^+$ to $CD24^+$ cells in parous samples could potentially be due to the increased number of lobulo-alveolar relative to ductal structures observed in parous women (Russo et al., 2001) or may reflect the presence of parity-induced stem cells described in murine mammary glands (Matulka et al., 2007; Wagner et al., 2002).

Biological pathways and networks affected by parity-related gene expression changes

Because the ultimate goal of our study is to identify targets for chemoprevention that would mimic the cancer-protective effects of parity, we investigated which signaling pathways might be affected by parity-related molecular changes. As early pregnancy specifically decreases the risk of ER+ breast tumors, we first explored our differentially expressed gene lists in $CD44^+$ cells for candidate mediators of this effect. We found several genes that may change the response of breast tissue to steroid hormones by altering metabolism (e.g., HSD17B11, HSD17B12, and HSD17B14) or by modulating nuclear receptors (e.g., NCOR1, NCOR2, NCOA4, and NCOA7) (Tables S2). Interestingly, androgen receptor (AR) and one of its key targets PSA (KLK3) were highly expressed in nulliparous $CD44^+$ cells, implying active androgen signaling pathway that is decreased following pregnancy. Among genes highly expressed in $CD44^+$ parous cells were a number of known tumor suppressors, such as Hakai/CBLL1 (Gong et al., 2010), CASP8 (Cox et al., 2007), SCRIB and LLGL2 (Humbert et al., 2008), and DNA repair-related genes (e.g., PRKDC, FANCB) suggesting that these cells may be more resistant to transformation in parous women.

In order to determine overall activation of specific biological functions due to parity in the cell types analyzed, we performed pathway enrichment, network, and protein interactome analyses using the MetaCore platform (Bessarabova et al., 2011; Ekins et al., 2006; Nikolsky et al., 2009). We found that parity has similar global effects on three of the four cell types analyzed, as pathways built on expression patterns in $CD10^+$ and $CD44^+$ cells and stroma cluster together for parous and nulliparous states (Figure 1E). The most significant pathways highly active in parous samples in all of these three cell types included apoptosis, survival, and immune response, whereas stem cells and development-related pathways were enriched only in $CD44^+$ nulliparous cells (Figure 1F and Table S3). Pathways highly active in parous stroma were enriched in energy metabolism, fatty acid metabolism and adipocyte differentiation from stem cells, which is consistent with adipose tissue development and a decrease in breast

density following pregnancy (Boyd et al., 2009). The functional categories of genes affected by parity were similar in all four cell types with receptors and enzymes representing the most enriched groups (Figure S2A and Table S4).

We focused our further analysis on $CD44^+$ cells that showed the most pronounced differences between parous and nulliparous states. Pathways highly active in nulliparous samples are related to major developmental and tumorigenic pathways including cytoskeleton remodeling, chemokines and cell adhesion, and WNT signaling, whereas pathways more active in parous samples include PI3K/AKT signaling and apoptosis (Table S3). Importantly, the highest scored pathway for genes highly expressed in nulliparous samples is four orders of magnitude more statistically significant than those for the genes highly expressed in parous samples, suggesting that downregulation of protumorigenic developmental pathways is a prominent feature of $CD44^+$ parous cells. Interactome analysis also demonstrated a much larger number of overconnected proteins in nulliparous than in parous state in all four cell types, but particularly in $CD44^+$ cells (Figure S2A). As the relative number of interactions (connectivity) is directly related to the functional activity of the dataset (Nikolsky et al., 2008), this result suggests that parous cells are overall substantially less active than nulliparous ones.

Because pregnancy-induced protection against breast cancer is also observed in rodents, we investigated if pathways altered by parity are conserved across species. We compared pathways in $CD44^+$ cells to that generated based on genes differentially expressed between virgin and parous rats (Blakely et al., 2006; D'Cruz et al., 2002). We found significant overlap between pathways highly active in nulliparous and virgin samples with top ranked pathways, including cytoskeleton remodeling and cell adhesion, highly relevant in stem cells (Figure S2B-C). Thus, pregnancy appears to induce similar alterations in the mammary epithelium regardless of species.

Cell type-specific epigenetic patterns related to parity and their functional relevance

Reduction of breast cancer risk in postmenopausal women conferred by full-term pregnancy in early adulthood implies the induction of long-lasting changes such as alterations in cell type-specific epigenetic patterns. To investigate this hypothesis we analyzed the comprehensive DNA methylation and histone H3 lysine 27 trimethylation (K27) profiles of $CD24^+$ and $CD44^+$ cells from nulliparous and parous women using MSDK-seq (Methylation-Specific Digital Karyotyping (Hu et al., 2005) applied to high-throughput sequencing) and ChIP-seq (Chromatin Immunoprecipitation applied to high-throughput sequencing) (Maruyama et al., 2011), respectively. Comparison of MSDK-seq libraries of nulliparous and parous samples within each cell type showed a higher number of significantly ($p<0.05$) differentially methylated regions (DMRs) in $CD44^+$ cells and in both cell types more DMRs were hypermethylated in nulliparous than in parous cells (Figure 2A and Table S5). The differences in DNA methylation of selected genes were validated in additional samples by quantitative methylation-specific PCR (qMSP) using $CD44^+$ cells from multiple nulliparous and parous cases. Despite some interpersonal variability, we detected statistically significant differences between nulliparous and parous groups that overall correlated with MSDK-seq data (Figure S2D).

To investigate pathways affected by parity-related epigenetic alterations, we analyzed pathways enriched by genes associated with gene body or promoter DMRs in $CD44^+$ cells from nulliparous and parous samples and found very little overlap among the four distinct categories (Figure 2B). The fraction of transcription factors (TFs) among differentially methylated genes is 2-3 fold higher than expected and what was observed among differentially expressed genes, implying that promoter methylation might be a preferred control mechanism of their expression (Figure 2C-D). Similar to the expression data, DMRs in nulliparous samples had higher numbers of overconnected objects than in parous ones. Gene body DMRs in $CD44^+$ nulliparous cells had the highest number of overconnected objects and transcription factors represented a significant

fraction of overconnected objects in promoter hypermethylated DMRs in $CD44^+$ nulliparous cells (Figure 2D).

We also analyzed associations between differential gene expression and presence of DMRs in $CD44^+$ and $CD24^+$cells, but did not find any significant correlation at the global scale (data not shown) potentially due to the complex relationship between DNA methylation and transcript levels as DNA methylation can both positively (e.g., in gene body) and negatively (e.g., in promoters) regulate gene expression, depending on the location relative to transcription start site (Jones, 1999; Maruyama et al., 2011; Zhang et al., 2006). However, the expression of several transcription factors with key roles in stem cells (e.g., HES7, STAT1) was correlated with the degree of promoter or gene body DNA methylation (Figure 2E). These findings are in agreement with recent data in multiple tissue types where most differentially expressed genes did not show significant differences in DNA methylation with the exception of transcription factors (Bock et al., 2012).

Analysis of the H3K27me3 profiles of $CD44^+$ or $CD24^+$ cells from nulliparous and parous samples did not detect significant parity-related differences in either cell types, although cell type-specific differences were clearly present (Figure S2E and data not shown). However, genes highly expressed in $CD44^+$ or $CD24^+$ cells from nulliparous women were not K27-enriched in either parous or nulliparous cases implying the potential lack of their regulation by the PRC2 complex that establishes this histone mark (Table S6).

Overall it appears that pregnancy may have a more pronounced long-term effect on DNA methylation than on K27 patterns and that parity-associated differences in DNA methylation only affect the expression of a limited number of transcription factors with key roles in development and differentiation.

Persistent parity-related decrease in $p27^+$ breast epithelial cells

CDKN1B encoding for p27, was one of the most significantly differentially expressed genes in $CD44^+$ cells from nulliparous and parous and also from control and BRCA1/2 parous tissues (Table S2). p27 has been reported to affect the number and proliferation of stem cells and progenitors in several organs in mice (Cheng et al., 2000; Muraoka et al., 2001; Oesterle et al., 2011). Thus, the higher expression of p27 in $CD44^+$ cells from nulliparous and BRCA1/2 women may indicate higher numbers of mammary epithelial progenitors in these samples. To investigate this hypothesis, we performed immunofluorescence analysis for p27 alone and in combination with CD24 and CD44 cell surface and Ki67 proliferation markers. We analyzed both premenopausal and postmenopausal tissues to confirm that the parity-related differences we detected by global profiling of premenopausal women are maintained after menopause. Using this approach we observed that the expression level of p27 and the number of $p27^+$ cells was significantly lower in parous compared to nulliparous samples from both pre- and postmenopausal women (Figure 3A-B). The frequency of $Ki67^+$ cells displayed similar differences between nulliparous and parous cases and $Ki67^+$ cells were only rarely $p27^+$ (Figure 3B-C).

To strengthen the link between the frequency of $p27^+$ cells and parity-related decrease in postmenopausal ER+ breast cancer risk, we also analyzed postmenopausal nulliparous and parous women with or without breast cancer. Although cancer-free nulliparous postmenopausal women had higher fraction of $p27^+$ cells than parous ones, the frequency of these cells was the highest in parous postmenopausal women with breast cancer (Figure S3A).

These data support our hypothesis that p27 may mark quiescent mammary epithelial progenitors with proliferative potential and that the numbers of these cells correlate with the risk of breast cancer, especially in postmenopausal parous women.

Link between parity-related differences and mammographic density

Mammographic density is one of the most significant risk factors for breast cancer and it is higher in nulliparous compared to parous women (Boyd et al., 2009). Thus, parity-related differences that are the most relevant to breast cancer risk may also correlate with differences in mammographic density. To test this hypothesis, we analyzed the expression of p27 and Ki67 in biopsy samples obtained from high and low density areas of the same breast (Lin et al., 2011). The overall expression of p27 and Ki67 were not significantly different between low and high-density areas, but the number of $p27^+$ cells was higher in high-density areas (Figure 3D). Thus, the numbers of $p27^+$ cells might correlate with both parity and mammographic density-related breast cancer risk.

$p27^+$ breast epithelial cells are quiescent hormone-responsive progenitors

The mutually exclusive expression of Ki67 and p27 in breast epithelial cells with their concomitant decrease in parous compared to nulliparous women implied that they may represent proliferating and quiescent progenitors, respectively. Ovarian hormones are the best-understood regulators of breast epithelial cell proliferation and also breast cancer risk (Brisken and O'Malley, 2010). Correlating with this, our gene expression data indicated a decrease in AR and AR targets in $CD44^+$ cells from parous women (Table S2) and prior studies implied a decrease in ER^+ breast epithelial cells in parous compared to nulliparous women (Taylor et al., 2009).

To explore the potential hormonal regulation of $p27^+$ breast epithelial cells, we analyzed the expression of p27, ER, and AR, in breast tissue samples from women varying parity and hormonal status. These included control nulliparous and parous women, parous BRCA1/2 mutation carriers, breast biopsy tissues from women in early (8-10 weeks) and late (22-26 weeks) stage of pregnancy, in follicular and luteal phase of the menstrual cycle or undergoing ovarian hyperstimulation prior to oocyte collection for in vitro fertilization. Interestingly, we found

that nearly all $p27^+$ cells were also ER^+ and their numbers were the highest in parous BRCA1/2 mutation carriers and the lowest in biopsy samples from pregnant women and after ovarian hyperstimulation, where both ovarian hormone and hCG (human choriogonadotropin) levels are the highest (Figure 4A). The frequencies of $p27^+$, ER^+, and $p27^+ER^+$ cells were also higher in control nulliparous compared to parous women and in follicular relative to luteal phase of the menstrual cycle (Figure 4A). Similar observations were made for AR, although the overlap between p27 and AR was less pronounced compared to that between p27 and ER (Figure 5B).

To further investigate the relationship between the numbers of $p27^+$ cells and ovarian hormone-induced breast epithelial cell proliferation, we performed immunofluorescence analysis for p27 and Ki67 in tissue samples with the highest differences in hormone levels. Correlating with prior data (Chung et al., 2011; Going et al., 1988; Meyer, 1977), the frequency of $Ki67^+$ cells was the highest in luteal phase of the menstrual cycle when both estrogen and progesterone levels are high (Figure 4C). Samples from early pregnancy had a lower fraction of proliferating $Ki67^+$ cells and the numbers of these cells was the lowest in follicular phase. The frequency of $p27^+$ cells displayed inverse correlation with that of $Ki67^+$ cells: it was the highest in follicular phase and lowest in biopsies from oocyte donors (Figure 4C). Interestingly, a low but detectable fraction of $p27^+$ cells was also $Ki67^+$ in luteal phase and early pregnancy, potentially marking proliferating progenitors in early G1 phase of the cell cycle. The differences in the frequency of $p27^+$ and $Ki67^+$ cells between follicular and luteal phase was less significant in parous compared to nulliparous women in part due to the lower overall fractions of these cells in parous cases (Figure S3A).

These results suggest that a subset of $p27^+$ cells may represent quiescent hormone-responsive luminal progenitors and that the numbers of these cells correlate with the risk of breast cancer.

Functional validation of parity-related differences in signaling pathways

Several signaling pathways less active in $CD44^+$ cells from parous women were related to stem cell maintenance and cell proliferation (Figure 1F). To investigate if inhibition of these pathways affect the number of $p27^+$ and proliferating cells, we incubated normal breast tissues in a tissue explant culture model with inhibitors or agonists of selected pathways (e.g., cAMP, EGFR, Cox2, Hh, TGFβ, Wnt, and IGFR) for 8-10 days after which we assessed the number of $p27^+$ cells and cellular proliferation based on bromodeoxyuridine (BrdU) incorporation and Ki67 expression.

Tissue architecture and cellular viability were maintained and $p27^+$, $Ki67^+$, and $BrdU^+$ cells were detected in all conditions (Figure 5A,B). The frequency of $p27^+$ cells most pronouncedly decreased following TGFBR and IGFR inhibitor treatment (Figure 5C). Treatment with TGFBR inhibitor significantly ($p<0.05$) increased mammary epithelial cell proliferation and the number of $BrdU^+$ cells, whereas inhibition of cAMP, EGFR, Cox2, Hh, and IGFR signaling had the opposite effects. Inhibition of EGFR and Cox2, and to a lesser degree Wnt and IGFR, decreased the fraction of $Ki67^+$ cells, while stimulation with Shh ligand increased it (Figure S3B).

To demonstrate that the numbers of and proliferation of $p27^+$ cells is regulated by ER and estrogen signaling, we analyzed the fraction of $p27^+$ and $Ki67^+$ cells in tissue slices treated with varying concentrations of ovarian hormones or tamoxifen. To be able to correlate the tissue slices data with that we observed in physiologic conditions (Figure 4), we used estrogen, progesterone, and prolactin hormone levels that mimic serum levels in follicular and luteal phase of the menstrual cycle or mid-pregnancy. The numbers of $p27^+$ cells was high in sections treated with concentrations of estrogen present in follicular phase and also following tamoxifen treatment, whereas it decreased in cultures incubated with luteal phase and pregnancy level hormones (Figure 5D and Figure SB). These data further support our hypothesis that a subset of $p27^+$ cells are hormone-responsive progenitors.

To establish a direct link between $p27^+$ cells and the signaling pathways analyzed, we confirmed that the selected pathways were active in $p27^+$ cells (Figure 6A) and that the compounds effectively inhibited their activity in these cells (Figure 6B-C). Most importantly, the expression of phosphoSmad2 (pSmad2), a key mediator of TGFβ signaling, demonstrated a nearly complete overlap with p27 both in tissue slices (Figure 6A) and in uncultured tissue samples (Figure 6D). The frequency of $p27^+pSmad2^+$ cells also fluctuated according to hormone levels displaying inverse correlation in mammary epithelial cell proliferation during menstrual cycle and pregnancy (Figures 6D and 4A) These results imply a key role for TGFβ signaling in maintaining mammary epithelial progenitors in a quiescent state via modulating the expression of p27 (Polyak et al., 1994).

Thus, these data suggest that the decreased frequency of $p27^+$ and $Ki67^+$ cells in parous women is a reflection of the decreased activity of stem cell-related signaling pathways after pregnancy identifying these pathways as potential targets for cancer preventive interventions.

Parity-associated decrease in mammary epithelial progenitors and breast tumor initiation

Our data support the hypothesis that a decrease in the number and proliferative potential of luminal progenitors in parous women directly relate to a decrease in breast cancer risk and that this effect is dependent on the age at first full-term pregnancy. As this hypothesis cannot be addressed experimentally in humans, we designed a mathematical model of the dynamics of proliferating mammary epithelial cells that can accumulate the changes leading to cancer initiation (Figure S4). We considered two types of cells: a self-renewing population of stem cells and a population of proliferating hormone-responsive luminal progenitors that result from the differentiation of these stem cells. We initiated our simulations at menarche and continued it until breast cancer initiation or death. We tested the effect of pregnancy on breast cancer initiation at varying times after menarche and compared these results with the risk of tumor

initiation in nulliparous women. We tested the robustness of our simulation over varying numbers of stem cells per terminal end duct, additional proliferative capacities resulting from pregnancy, and rates of asymmetric stem cell division. We then compared the relative likelihood of cancer initiation with pregnancy occurring at four different time points during childbearing years as compared to nulliparous simulations (see Supplementary Information for details). We found that the probability of cancer initiation in a duct increases as the age at first pregnancy increases (Figure 7A and S5). Furthermore, our simulations imply that differences in the numbers of luminal epithelial progenitors with proliferative potential are the most probable explanation of differences in breast cancer risk due to reproductive (e.g., parity) and genetic (e.g., BRCA1/2 germline mutation) factors.

Prognostic relevance of parity in breast cancer patients

Our gene expression and immunofluorescence data suggest that the numbers and properties of luminal progenitors are different in nulliparous and parous women. If these cells participate in breast cancer initiation then tumors developing in parous and nulliparous women might also be different and this might impact their gene expression profiles and clinical outcome. To test these hypotheses, we first investigated the effect of parity on breast cancer-specific survival in the Nurses' Health Study (NHS), a cohort including 121,700 women followed since 1976 (Colditz et al., 1995). Overall, Kaplan Meier curves show that there is no significant association between parity and breast cancer-specific survival (p=0.29). However, when the analysis was limited to ER+ tumors, we found that nulliparous women had a suggestive worse survival compared with parous women (Figure 7B). In multivariate analysis there was still a marginally significant association among women with ER+ tumors, with nulliparous women having a nearly 30% increased risk of death from their disease (HR: 1.29, 95% CI: 0.98, 1.70; p=0.06). Assessing associations between age at first pregnancy and number of pregnancies gave similar results

(data not shown). In contrast, among women with ER- tumors, parity was not associated with breast cancer-specific survival (P=0.51). Thus, parity may also influence the clinical outcome of breast cancer patients with ER+ tumors.

Because pregnancy may not induce the same epigenetic and gene expression changes in all women due to germline variations, we next investigated if our parity-related gene expression signature (PAGES) in $CD44^+$ cells might be a more useful prognostic marker than parity status alone. Thus, we analyzed the expression of PAGES in public breast cancer gene expression data with clinical outcome. We applied the supervised principle component analysis (SPCA) on one of the cohorts (Wang et al., 2005) as a training set to identify the subset of the PAGES with prognostic value followed by validation in three other cohorts (Desmedt et al., 2007; Sotiriou et al., 2006; van de Vijver et al., 2002) (Figure 7B and Figure S4A-C). In each dataset we selected ER^+ tumors, the tumor subtype affected by parity, and cases without systemic therapy in order to avoid differences due to treatment. All patients in the training set had small (<2cm), lymph node negative tumors at the time of diagnosis. Using this approach, we identified parity-related gene signatures that split patients into two distinct groups with significant survival difference (Table S7). Interestingly, such prognostic signature was found among genes highly expressed in both nulliparous and parous samples and each set of genes could be further separated into good and bad signatures (Figure 7B and Figure S4A-C). These results reflect the complex relationship between pregnancy and breast cancer that involves both protective and tumor-promoting effects (Schedin, 2006).

DISCUSSION

Our comprehensive analysis of multiple cell types in normal human breast tissue from nulliparous and parous women implies that parity has the most pronounced effect on luminal progenitors enriched in $CD44^+$ cell fraction. Most of the differences relate to transcriptional

repression and downregulation of genes and pathways important for stem cell function including EGF, IGF, Hh, and TGFβ signaling. Correlating with our findings, high circulating IGF-1 levels have been associated with increased risk of ER+ breast cancer (Key et al., 2010). Similarly, germline polymorphism in members of the TGFβ signaling pathway influence breast cancer susceptibility (Scollen et al., 2011).

One of the intriguing findings of our study is the high numbers of $p27^{+}$ cells in breast tissues of nulliparous women and parous BRCA1/2 mutation carriers with high risk of breast cancer, which seems paradoxical as CDKN1B/$p27^{kip1}$ is a bona fide tumor suppressor and potent inhibitor of cell cycle progression. However, p27 has been shown to play an important role in stem and progenitor cells, best characterized in the murine hematopoietic and nervous system, where loss of p27 increases the number of transit amplifying progenitors but not that of stem cells (Cheng et al., 2000; Mitsuhashi et al., 2001; Oesterle et al., 2011). In the mouse mammary gland, the consequences of p27 deficiency have been controversial. The role of p27 in mouse breast epithelium has been assessed based on mammary transplant assays (Muraoka et al., 2001) due to infertility and hormonal defects of female p27/Cdkn1b -/- mice (Fero et al., 1996; Kiyokawa et al., 1996). Using this approach, in one study p27 deficiency was associated with hypoplasia and impaired ductal branching and lobulo-alveolar differentiation (Muraoka et al., 2001), a phenotype consistent with a putative role for p27 in regulating the number and proliferation of breast epithelial progenitors (although this was not investigated). In contrast, another study using the same strain of mice found increased cell proliferation but no defects in ducto-alveolar branching and differentiation (Davison et al., 2003).

Similarly, the effect of p27 deletion on mammary tumorigenesis has been analyzed by crossing MMTV-neu and p27+/+, +/-, and -/- mice (Muraoka et al., 2002). The number of neu-induced tumors increased in p27 +/- background compared to p27 +/+, whereas the number of tumors in MMTV-neu/p27-/- mice was decreased. The authors attributed these findings to the

cell autonomous effects of p27 on cyclin D-cdk4 complex formation and activity, and the potential role of p27 in the regulation of the numbers of tumor initiating cells was not investigated. Nevertheless, these data are consistent with our hypothesis that p27 plays a key role in mammary epithelial progenitors. Correlating with this, our analysis of stem, luminal progenitor and mature luminal cells from mammary glands of virgin and retired breeder (i.e., parous) mice demonstrated highest p27 expression in progenitors and higher levels in virgin compared to parous mice (Huh et al, manuscript in preparation).

Based on our data, we hypothesize, that p27 regulates the proliferation and pool size of hormone-responsive luminal progenitors with proliferative potential. Thus, the lower numbers of these $p27^+$ cells in control parous women may contribute to their decrease in breast cancer risk. High p27 levels and quiescence are maintained in these cells by TGFβ signaling as implied by the co-expression of pSmad2 with p27 and the increase in BrdU incorporation with concomitant decrease in p27. Correlating with the presumed importance of TGFβ signaling and p27 in regulating hormone-responsive progenitors, recent whole-genome sequencing studies detected *CDKN1B* and TGFβ pathway mutations specifically in ER+ breast tumors (Stephens, 2012).

In summary, here we describe that the numbers of $p27^+$ cells in the normal breast might be used as a biomarker to predict breast cancer risk and to monitor the efficacy of cancer preventive strategies. Furthermore, the pathways we identified, especially TGFβ, might be exploited for breast cancer prevention as their modulation could deplete $p27^+$ luminal progenitors and consequently decrease breast cancer risk.

EXPERIMENTAL PROCEDURES

Tissue samples, cell purification, and genomic profiling

Fresh normal breast tissue specimens were collected at Harvard-affiliated hospitals, at Johns Hopkins University School of Medicine, and Baylor-Charles A. Sammons Cancer Center using Institutional Review Board-approved protocols. For organ cultures thin (~1mm) slices of epithelium-enriched breast tissue were cultured for 8 days in 6-well plates with co-culture inserts in M87A medium (Bloushtain-Qimron et al., 2008; Garbe et al., 2009). Detailed protocols for cell purification and the generation of SAGE-seq, MSDK-seq, and ChIP-seq libraries are posted on our web-site (http://research4.dfci.harvard.edu/polyaklab/protocols_linkpage.php). Genomic data were analyzed as described before (Kowalczyk et al., 2011; Maruyama et al., 2011; Wu et al., 2010). Semi-quantitative and quantitative RT-PCR and qMSP analyses were performed on cells purified from at least 15-20 samples of nulliparous and parous breast tissue as previously reported (Bloushtain-Qimron et al., 2008; Hu et al., 2005; Shipitsin et al., 2007). Details of genomic data analyses and tissue slice cultures are included in Supplemental Information.

FACS, immunofluorescence, and immunohistochemical analyses

Single-cell suspension of human breast epithelial cells was obtained essentially as described (Bloushtain-Qimron et al., 2008; Shipitsin et al., 2007). Cells were stained with propidium iodine, PE/Cy7-CD10 (Biolegend, Clone HI10a), APC-CD24 (Biolegend, clone ML5), and Zenon Alexa 405 labeled CD44 (BD, Clone 515). Immunohistochemical and immunofluorescence analyses were performed essentially as described (Marotta et al., 2011); detailed protocols are included in Supplemental Information.

Nurses' Health Study data

The Nurses' Health Study (NHS) is a prospective cohort study established in 1976. This study was approved by the Human Subjects Committee at Brigham and Women's Hospital in Boston, Massachusetts. Breast cancer cases were followed from the date of diagnosis until January 1, 2008 or death, whichever came first. Ascertainment of deaths included reporting by next of kin or postal authorities or searching the National Death Index. Approximately 98% of deaths in the NHS have been identified by these methods. Cause of death was ascertained from death certificates and physician review of medical records. Information on estrogen receptor (ER) status was extracted from the medical record and pathology reports. If data were missing for ER status, we used scoring from immunohistochemical staining for ER on 5 μm paraffin sections cut from tissue microarray blocks (Tamimi et al., 2008). There were 8,055 women with invasive breast cancer diagnosed after return of the 1976 baseline through 2006 questionnaire. One woman was excluded due to missing information on parity. Thus, our final analysis included 8,054 women with invasive breast cancer and information on parity. Survival curves were estimated by the Kaplan-Meier method and statistical significance was assessed with the log-rank test. Multivariate cox proportional hazards regression models were used to evaluate the relationship between parity and breast cancer-specific mortality after adjusting for age at diagnosis, aspirin use, date of diagnosis, disease stage, grade, radiation treatment, chemotherapy and hormonal treatment. All analyses were performed using SAS version 9.1. All statistical tests were two sided and $P<0.05$ was considered statistically significant.

ACCESSION NUMBERS

Raw data files and methodological details have been submitted to GEO with accession number GSE32017.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and seven tables.

ACKNOWLEDGEMENTS

We thank the BC Cancer Agency Genome Sciences Centre and the Wistar Institute for sequencing, Lisa Cameron in the DFCI Confocal and Light Microscopy Core Facility and members of Dr. Massimo Loda's lab for technical assistance, and Dr. Jenny Cawson and her team at BreastScreen Victoria and Dr. Prue Hill from St. Vincent's Pathology, Melbourne for their help in accruing the breast mammographic density samples, and Drs. Sally Knox, Jeffrey Lamont, and Dao Tuoc from Baylor University Medical Center- Baylor Sammons Cancer Center, Dallas, TX for their help with collecting tissue samples from BRCA1/2 patients. We thank members of our laboratories for their critical reading of this manuscript and useful discussions. We thank Jonathan Yingling (Eli Lilly) for providing the LY2109761 TGFBR kinase inhibitor. Samples from the Susan G. Komen for the Cure Tissue Bank at the IU Simon Cancer Center were used in this study. We thank contributors, including Indiana University who collected samples used in this study, as well as donors and their families, whose help and participation made this work possible. We thank the participants and staff of the Nurses' Health Study (NHS) cohort, for their valuable contributions. We thank the following state cancer registries for their help with the NHS: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The study sponsors had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication. This work was supported by the Avon Foundation (K.P. and S.S), the National Cancer Institute P50 CA89383 and P01 CA080111 (S.S., K.P., and A.R.), CA116235-04S1 (K.P.), and CA087969 (R.T.), the Susan G. Komen Foundation (R.M., Y.S., I.H., and K.P.), the

Terri Brodeur Foundation (S.C.), US Army Congressionally Directed Research W81XWH-07-1-0294 (K.P.), the Victorian Breast Cancer Research Consortium (S.J.L., G.L.C., and E.W.T.), the St. Vincent's Hospital Research Endowment Fund and the Victorian Government's Operational Infrastructure Support Program (E.W.T.), and the Cellex Foundation (V.A.). The Programme for Advanced Medical Education (F.C.M.) is sponsored by Fundação Calouste Gulbenkian, Fundação Champalimaud, Ministério da Saúde e Fundação para a Ciência e Tecnologia, Portugal.

REFERENCES

Asselin-Labat, M.L., Vaillant, F., Sheridan, J.M., Pal, B., Wu, D., Simpson, E.R., Yasuda, H., Smyth, G.K., Martin, T.J., Lindeman, G.J.*, et al.* (2010). Control of mammary stem cell function by steroid hormone signalling. Nature *465*, 798-802.

Asztalos, S., Gann, P.H., Hayes, M.K., Nonn, L., Beam, C.A., Dai, Y., Wiley, E.L., and Tonetti, D.A. (2010). Gene expression patterns in the human breast after pregnancy. Cancer Prev Res (Phila) *3*, 301-311.

Belitskaya-Levy, I., Zeleniuch-Jacquotte, A., Russo, J., Russo, I.H., Bordas, P., Ahman, J., Afanasyeva, Y., Johansson, R., Lenner, P., Li, X.*, et al.* (2011). Characterization of a genomic signature of pregnancy identified in the breast. Cancer Prev Res (Phila) *4*, 1457-1464.

Bessarabova, M., Pustovalova, O., Shi, W., Serebriyskaya, T., Ishkin, A., Polyak, K., Velculescu, V.E., Nikolskaya, T., and Nikolsky, Y. (2011). Functional synergies yet distinct modulators affected by genetic alterations in common human cancers. Cancer Res *71*, 3471-3481.

Blakely, C.M., Stoddard, A.J., Belka, G.K., Dugan, K.D., Notarfrancesco, K.L., Moody, S.E., D'Cruz, C.M., and Chodosh, L.A. (2006). Hormone-induced protection against mammary tumorigenesis is conserved in multiple rat strains and identifies a core gene expression signature induced by pregnancy. Cancer Res *66*, 6421-6431.

Bloushtain-Qimron, N., Yao, J., Snyder, E.L., Shipitsin, M., Campbell, L.L., Mani, S.A., Hu, M., Chen, H., Ustyansky, V., Antosiewicz, J.E.*, et al.* (2008). Cell type-specific DNA methylation patterns in the human breast. Proc Natl Acad Sci U S A *105*, 14076-14081.

Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol Cell *47*, 633-647.

Boyd, N.F., Martin, L.J., Yaffe, M., and Minkin, S. (2009). Mammographic density. Breast Cancer Res *11 Suppl 3*, S4.

Brisken, C., and O'Malley, B. (2010). Hormone action in the mammary gland. Cold Spring Harb Perspect Biol *2*, a003178.

Britt, K.L., Kendrick, H., Regan, J.L., Molyneux, G., Magnay, F.A., Ashworth, A., and Smalley, M.J. (2009). Pregnancy in the mature adult mouse does not alter the proportion of mammary epithelial stem/progenitor cells. Breast Cancer Res *11*, R20.

Cheng, T., Rodrigues, N., Dombkowski, D., Stier, S., and Scadden, D.T. (2000). Stem cell repopulation efficiency but not pool size is governed by p27(kip1). Nat Med *6*, 1235-1240.

Chung, K., Hovanessian-Larsen, L.J., Hawes, D., Taylor, D., Downey, S., Spicer, D.V., Stanczyk, F.Z., Patel, S., Anderson, A.R., Pike, M.C.*, et al.* (2011). Breast epithelial cell proliferation is markedly increased with short-term high levels of endogenous estrogen secondary to controlled ovarian hyperstimulation. Breast Cancer Res Treat.

Colditz, G.A., Hankinson, S.E., Hunter, D.J., Willett, W.C., Manson, J.E., Stampfer, M.J., Hennekens, C., Rosner, B., and Speizer, F.E. (1995). The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. N Engl J Med *332*, 1589-1593.

Colditz, G.A., Rosner, B.A., Chen, W.Y., Holmes, M.D., and Hankinson, S.E. (2004). Risk factors for breast cancer according to estrogen and progesterone receptor status. J Natl Cancer Inst *96*, 218-228.

Cox, A., Dunning, A.M., Garcia-Closas, M., Balasubramanian, S., Reed, M.W., Pooley, K.A., Scollen, S., Baynes, C., Ponder, B.A., Chanock, S.*, et al.* (2007). A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet *39*, 352-358.

Cullinane, C.A., Lubinski, J., Neuhausen, S.L., Ghadirian, P., Lynch, H.T., Isaacs, C., Weber, B., Moller, P., Offit, K., Kim-Sing, C.*, et al.* (2005). Effect of pregnancy as a risk factor for breast cancer in BRCA1/BRCA2 mutation carriers. Int J Cancer *117*, 988-991.

D'Cruz, C.M., Moody, S.E., Master, S.R., Hartman, J.L., Keiper, E.A., Imielinski, M.B., Cox, J.D., Wang, J.Y., Ha, S.I., Keister, B.A.*, et al.* (2002). Persistent parity-induced changes in growth factors, TGF-beta3, and differentiation in the rodent mammary gland. Mol Endocrinol *16*, 2034-2051.
Davison, E.A., Lee, C.S., Naylor, M.J., Oakes, S.R., Sutherland, R.L., Hennighausen, L., Ormandy, C.J., and Musgrove, E.A. (2003). The cyclin-dependent kinase inhibitor p27 (Kip1) regulates both DNA synthesis and apoptosis in mammary epithelium but is not required for its functional development during pregnancy. Mol Endocrinol *17*, 2436-2447.
Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M.S.*, et al.* (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res *13*, 3207-3214.
Ekins, S., Nikolsky, Y., Bugrim, A., Kirillov, E., and Nikolskaya, T. (2006). Pathway mapping tools for analysis of high content data. In High Content Screening (Humana Press), pp. 319-350.
Fero, M.L., Rivkin, M., Tasch, M., Porter, P., Carow, C.E., Firpo, E., Polyak, K., Tsai, L.H., Broudy, V., Perlmutter, R.M.*, et al.* (1996). A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice. Cell *85*, 733-744.
Garbe, J.C., Bhattacharya, S., Merchant, B., Bassett, E., Swisshelm, K., Feiler, H.S., Wyrobek, A.J., and Stampfer, M.R. (2009). Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. Cancer Res *69*, 7557-7568.
Ginger, M.R., and Rosen, J.M. (2003). Pregnancy-induced changes in cell-fate in the mammary gland. Breast Cancer Res *5*, 192-197.
Going, J.J., Anderson, T.J., Battersby, S., and MacIntyre, C.C. (1988). Proliferative and secretory activity in human breast during natural and artificial menstrual cycles. Am J Pathol *130*, 193-204.
Gong, E.Y., Park, E., and Lee, K. (2010). Hakai acts as a coregulator of estrogen receptor alpha in breast cancer cells. Cancer Sci *101*, 2019-2025.
Haakensen, V.D., Bjoro, T., Luders, T., Riis, M., Bukholm, I.K., Kristensen, V.N., Troester, M.A., Homen, M.M., Ursin, G., Borresen-Dale, A.L.*, et al.* (2011a). Serum estradiol levels associated with specific gene expression patterns in normal breast tissue and in breast carcinomas. BMC Cancer *11*, 332.
Haakensen, V.D., Lingjaerde, O.C., Luders, T., Riis, M., Prat, A., Troester, M.A., Holmen, M.M., Frantzen, J.O., Romundstad, L., Navjord, D.*, et al.* (2011b). Gene expression profiles of breast biopsies from healthy women identify a group with claudin-low features. BMC medical genomics *4*, 77.
Hu, M., Yao, J., Cai, L., Bachman, K.E., van den Brule, F., Velculescu, V., and Polyak, K. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. Nat Genet *37*, 899-905.
Humbert, P.O., Grzeschik, N.A., Brumby, A.M., Galea, R., Elsum, I., and Richardson, H.E. (2008). Control of tumourigenesis by the Scribble/Dlg/Lgl polarity module. Oncogene *27*, 6888-6907.
Jones, P.A. (1999). The DNA methylation paradox. Trends Genet *15*, 34-37.
Key, T.J., Appleby, P.N., Reeves, G.K., and Roddam, A.W. (2010). Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. Lancet Oncol *11*, 530-542.
Kiyokawa, H., Kineman, R.D., Manova-Todorova, K.O., Soares, V.C., Hoffman, E.S., Ono, M., Khanam, D., Hayday, A.C., Frohman, L.A., and Koff, A. (1996). Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27(Kip1). Cell *85*, 721-732.

Kowalczyk, A., Bedo, J., Conway, T., and Beresford-Smith, B. (2011). The poisson margin test for normalization-free significance analysis of NGS data. J Comput Biol *18*, 391-400.
Lin, S.J., Cawson, J., Hill, P., Haviv, I., Jenkins, M., Hopper, J.L., Southey, M.C., Campbell, I.G., and Thompson, E.W. (2011). Image-guided sampling reveals increased stroma and lower glandular complexity in mammographically dense breast tissue. Breast Cancer Res Treat *128*, 505-516.
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.
Marotta, L.L., Almendro, V., Marusyk, A., Shipitsin, M., Schemme, J., Walker, S.R., Bloushtain-Qimron, N., Kim, J.J., Choudhury, S.A., Maruyama, R.*, et al.* (2011). The JAK2/STAT3 signaling pathway is required for growth of CD44CD24 stem cell-like breast cancer cells in human tumors. J Clin Invest *121*, 2723-2735.
Maruyama, R., Choudhury, S., Kowalczyk, A., Bessarabova, M., Beresford-Smith, B., Conway, T., Kaspi, A., Wu, Z., Nikolskaya, T., Merino, V.F.*, et al.* (2011). Epigenetic regulation of cell type-specific expression patterns in the human mammary epithelium. PLoS genetics *7*, e1001369.
Maruyama, R., Shipitsin, M., Choudhury, S., Wu, Z., Protopopov, A., Yao, J., Lo, P.K., Bessarabova, M., Ishkin, A., Nikolsky, Y.*, et al.* (2010). Breast Cancer Special Feature: Altered antisense-to-sense transcript ratios in breast cancer. Proc Natl Acad Sci U S A.
Matulka, L.A., Triplett, A.A., and Wagner, K.U. (2007). Parity-induced mammary epithelial cells are multipotent and express cell surface markers associated with stem cells. Dev Biol *303*, 29-44.
Meyer, J.S. (1977). Cell proliferation in normal human breast ducts, fibroadenomas, and other ductal hyperplasias measured by nuclear labeling with tritiated thymidine. Effects of menstrual phase, age, and oral contraceptive hormones. Hum Pathol *8*, 67-81.
Mitsuhashi, T., Aoki, Y., Eksioglu, Y.Z., Takahashi, T., Bhide, P.G., Reeves, S.A., and Caviness, V.S., Jr. (2001). Overexpression of p27Kip1 lengthens the G1 phase in a mouse model that targets inducible gene expression to central nervous system progenitor cells. Proc Natl Acad Sci U S A *98*, 6435-6440.
Muraoka, R.S., Lenferink, A.E., Law, B., Hamilton, E., Brantley, D.M., Roebuck, L.R., and Arteaga, C.L. (2002). ErbB2/Neu-induced, cyclin D1-dependent transformation is accelerated in p27-haploinsufficient mammary epithelial cells but impaired in p27-null cells. Mol Cell Biol *22*, 2204-2219.
Muraoka, R.S., Lenferink, A.E., Simpson, J., Brantley, D.M., Roebuck, L.R., Yakes, F.M., and Arteaga, C.L. (2001). Cyclin-dependent kinase inhibitor p27(Kip1) is required for mouse mammary gland morphogenesis and function. J Cell Biol *153*, 917-932.
Nikolsky, Y., Kirillov, E., Zuev, R., Rakhmatulin, E., and Nikolskaya, T. (2009). Functional analysis of OMICs data and small molecule compounds in an integrated "knowledge-based" platform. Methods Mol Biol *563*, 177-196.
Nikolsky, Y., Sviridov, E., Yao, J., Dosymbekov, D., Ustyansky, V., Kaznacheev, V., Dezso, Z., Mulvey, L., Macconaill, L.E., Winckler, W.*, et al.* (2008). Genome-wide functional synergy between amplified and mutated genes in human breast cancer. Cancer Res *68*, 9532-9540.
Oesterle, E.C., Chien, W.M., Campbell, S., Nellimarla, P., and Fero, M.L. (2011). p27(Kip1) is required to maintain proliferative quiescence in the adult cochlea and pituitary. Cell Cycle *10*, 1237-1248.
Polyak, K., Kato, J.Y., Solomon, M.J., Sherr, C.J., Massague, J., Roberts, J.M., and Koff, A. (1994). p27Kip1, a cyclin-Cdk inhibitor, links transforming growth factor-beta and contact inhibition to cell cycle arrest. Genes Dev *8*, 9-22.

Russo, J., Balogh, G.A., and Russo, I.H. (2008). Full-term pregnancy induces a specific genomic signature in the human breast. Cancer Epidemiol Biomarkers Prev *17*, 51-66.
Russo, J., Lynch, H., and Russo, I.H. (2001). Mammary gland architecture as a determining factor in the susceptibility of the human breast to cancer. Breast J *7*, 278-291.
Russo, J., Moral, R., Balogh, G.A., Mailo, D., and Russo, I.H. (2005). The protective role of pregnancy in breast cancer. Breast Cancer Res *7*, 131-142.
Russo, J., Santucci-Pereira, J., de Cicco, R.L., Sheriff, F., Russo, P.A., Peri, S., Slifker, M., Ross, E., Mello, M.L., Vidal, B.C.*, et al.* (2011). Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. Int J Cancer.
Schedin, P. (2006). Pregnancy-associated breast cancer and metastasis. Nat Rev Cancer *6*, 281-291.
Scollen, S., Luccarini, C., Baynes, C., Driver, K., Humphreys, M.K., Garcia-Closas, M., Figueroa, J., Lissowska, J., Pharoah, P.D., Easton, D.F.*, et al.* (2011). TGF-beta signaling pathway and breast cancer susceptibility. Cancer Epidemiol Biomarkers Prev *20*, 1112-1119.
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.
Sivaraman, L., and Medina, D. (2002). Hormone-induced protection against breast cancer. J Mammary Gland Biol Neoplasia *7*, 77-92.
Siwko, S.K., Dong, J., Lewis, M.T., Liu, H., Hilsenbeck, S.G., and Li, Y. (2008). Evidence that an early pregnancy causes a persistent decrease in the number of functional mammary epithelial stem cells--implications for pregnancy-induced protection against breast cancer. Stem Cells *26*, 3205-3209.
Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B.*, et al.* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst *98*, 262-272.
Stephens, P.e.a. (2012). The landscape of cancer genes and mutational processes in breast cancer Nature *in press*.
Tamimi, R.M., Baer, H.J., Marotti, J., Galan, M., Galaburda, L., Fu, Y., Deitz, A.C., Connolly, J.L., Schnitt, S.J., Colditz, G.A.*, et al.* (2008). Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. Breast Cancer Res *10*, R67.
Taylor, D., Pearce, C.L., Hovanessian-Larsen, L., Downey, S., Spicer, D.V., Bartow, S., Pike, M.C., Wu, A.H., and Hawes, D. (2009). Progesterone and estrogen receptors in pregnant and premenopausal non-pregnant normal human breast. Breast Cancer Res Treat *118*, 161-168.
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.
Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314-322.
Wagner, K.U., Boulanger, C.A., Henry, M.D., Sgagias, M., Hennighausen, L., and Smith, G.H. (2002). An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. Development *129*, 1377-1386.
Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J.*, et al.* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671-679.
Wu, Z.J., Meyer, C.A., Choudhury, S., Shipitsin, M., Maruyama, R., Bessarabova, M., Nikolskaya, T., Sukumar, S., Schwartzman, A., Liu, J.S.*, et al.* (2010). Gene expression profiling of human breast tissue samples using SAGE-Seq. Genome Res *20*, 1730-1739.
Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

FIGURE LEGENDS

Figure 1. Cell type-specific differences in gene expression according to parity and BRCA1/2 mutation status. (A) Genome-wide view of genes differentially expressed between nulliparous (N) and parous (P) samples in the four cell types analyzed. Each dot represents a gene. Fold differences between averaged N and P samples and their corresponding p-values are plotted on the y and x-axis, respectively. Green vertical lines indicate p=0.05, numbers indicate genes differentially expressed at $p<0.05$. (B) Three-dimensional projection of the gene expression data onto the first three principal components. Each ball is a different sample; cell type and parity are indicated. (C) Hierarchical clustering of $CD44^+$ cells from nulliparous and parous control women and parous BRCA1 and BRCA2 mutation carriers. (D) Frequency of $CD44^+$, $CD24^+$, and $CD10^+$ cells relative to total human breast epithelial cells. 10 samples each from nulliparous and parous groups were analyzed. Each dot represents an individual sample. Error bars represent mean ± SEM. (E) Dendrogram depicting hierarchical clustering of signaling pathways significantly high in parous or nulliparous samples in any of the four cell types analyzed. (F) Heatmap depicting unsupervised clustering of signaling pathways significantly down or upregulated in parous compared to nulliparous samples in any of the four cell types analyzed. Color scale indicates –log p value of enrichment. Orange rectangles highlight cell type-specific or common altered pathways.

Figure 2. Epigenetic differences between nulliparous and parous tissues. (A) Genome-wide view of differentially methylated genes in $CD24^+$ and $CD44^+$ cells between nulliparous and parous samples. All MSDK sites are plotted on the x-axis in the order of p-values of the difference between nulliparous and parous samples in $CD44^+$ or $CD24^+$ cells. Log ratios of averaged MSDK counts in three N and three P samples are plotted on the y-axis. Green vertical lines indicate p=0.01 and the numbers of significant DMRs ($p<0.01$) are shown. (B) Pathways

enriched with genes in $CD44^+$ cells with the indicated difference in DNA methylation between nulliparous and parous women. (C) Genes with promoter and gene body DMRs in $CD44^+$ cells from nulliparous and parous samples were analyzed for relative enrichment with the indicated protein classes and for relative connectivity. X-axes indicate –log10 p-values for enrichment with the listed protein classes and the number of overconnected objects, defined as proteins with higher than expected numbers of interactions, in each functional category within each group, respectively. (D) Pie charts depicting the relative % of genes in different functional categories with the indicated gene expression and DNA methylation pattern in CD44+ cells from nulliparous and parous women. (E) Scatter plot for MSDK-seq and SAGE-seq data to depict correlations between differential promoter methylation and differential gene expression for transcription factors. Genes with transcription regulator activity were selected based on Gene Ontology (GO0030528) and plotted in a scatter (or starburst) plot for comparison of transcriptional and epigenetic differences between parous and nulliparous samples. Each point represents a gene with a MSDK-seq site in certain region (promoter: 5kb upstream to 2kb downstream from TSS, gene body: 2kb downstream from TSS to the end of gene), and log10(p value) is plotted for difference of DNA methylation (x-axis) and expression (y-axis) between parous and nulliparous samples. If a MSDK site is hypermethylated or a gene is higher expressed in parous, -1 is multiplied to log10 (p value), providing positive values. MSDK-seq sites that are significantly ($p<0.05$) hypo- or hypermethylated in parous or nulliparous samples are highlighted in blue.

Figure 3. Expression of p27 in normal breast tissue samples. Representative examples of multicolor immunofluorescence analyses of normal mammary epithelium. (A) Expression of p27, CD24, and CD44 in breast tissue of premenopausal nulliparous (NP) and parous (P) women. Graphs show the quantification of p27 staining intensity in multiple samples. p-value of

difference between nulliparous and parous groups are indicated. Immunofluorescence staining for p27 and Ki67 in breast tissue from premenopausal (B) and postmenopausal (C) nulliparous and parous women. Graph shows frequencies of $p27^+$ and $Ki67^+$ cells in nulliparous (NP) and parous (P) samples. p-values of differences between nulliparous and parous groups are indicated. (D) Expression of p27 in $CD44^+$ and $CD24^+$ cells and the frequencies of $p27^+$ and $Ki67^+$ cells in the breast epithelium of high and low density areas of the same breast from parous women. White arrows mark a subpopulation of $p27^+CD44^+$ cells.

Figure 4. Hormonal factors and the expression of p27 in normal breast tissues. (A) Representative double immunofluorescence staining for p27 and ER in breast tissue from the indicated groups of women. Graph shows frequencies of $p27^+$, ER^+, and $p27^+ER^+$ cells in each group of samples. (B) Representative double immunofluorescence staining for p27 and AR in breast tissue from premenopausal nulliparous and parous women, and in BRCA1 mutation carriers. Graphs show frequencies of $p27^+$ and AR^+ cells in each set of samples. (C) Representative double immunofluorescence staining for p27 and Ki67 in breast tissue from the indicated groups of women. Graphs show frequencies of $p27^+$, $Ki67^+$, and $p27^+Ki67^+$ cells in each group of samples.

Figure 5. Modulation of $p27^+$ breast epithelial cells and proliferation by hormonal and parity-related pathways. (A) Representative hematoxilin and eosin staining depicting morphology of breast tissue after 8 days in culture. (B) Representative examples of multicolor immunofluorescence analyses of $BrdU^+$, $p27^+$, and $Ki67^+$ cells in control and tissues treated with inhibitors of the indicated pathways. (C) Frequency of $Ki67^+$, $BrdU^+$, and $p27^+$ cells in each of the indicated conditions. Asterisks indicate significant ($p<0.05$) differences between control and treated groups. (D) Representative images of immunofluorescence analysis of p27 and graph

depicting the frequency of $p27^+$ cells in tissue slices treated with hormones mimicking the indicated physiologic levels in women.

Figure 6. Signaling pathways regulating $p27^+$ breast epithelial cells. (A) Representative examples of multicolor immunofluorescence analyses of pSMAD2, pEGFR, Axin2+, and p27 cells in control and tissues treated with inhibitors of the indicated pathways. (B) Quantitation of differences in the expression of markers reflecting pathway activity between control and inhibitor-treated tissues (inhibitor). (C) RGB spectra demonstrating overlap between the expression of p27 and the indicated marker. (D) Double immunofluorescence staining for p27 and pSmad2 in breast tissues from the indicated women. Graph shows frequencies of $p27^+$, $pSmad2^+$, and $p27^+pSmad2^+$ cells in each group of samples.

Figure 7. Relevance of parity-related gene expression changes to breast cancer. (A) Kaplan-Meier plot depicting the probability of breast cancer-specific survival among women with invasive ER+ or ER- breast cancer by parity in the Nurses' Health Study (1976-2006). The P value was calculated with use of the log-rank test. (B) Significant association of the presence of a parity-related gene signature with overall survival in a cohort of breast cancer patients with ER+ tumors. Kaplan-Meier curves and their corresponding log-rank test p-values are shown. Overall survival is defined by death from any cause. Top heat map shows the signature from down regulated genes after pregnancy and the bottom heat map from up group genes. The color bar above heat map indicates the two distinct patients groups separated by the co-expression of the signature (Green: better survival group. Red: worse survival group). The color bar at the right side of heat map indicates effect of pregnancy on genes in breast cancer progression. Pink color indicates pregnancy induces gene expression level change in the same trend as breast cancer progression. Blue color indicates pregnancy induces gene expression

level change in the opposite trend as breast cancer progression. Black bar indicates death.