

# Molecular Profiling of Human Mammary Gland Links Breast Cancer Risk to a p27<sup>+</sup> Cell Population with Progenitor Characteristics

Sibgat Choudhury,<sup>1,5,7,32</sup> Vanessa Almendro,<sup>1,5,7,10,32</sup> Vanessa F. Merino,<sup>11,32</sup> Zhenhua Wu,<sup>2,9,32</sup> Reo Maruyama,<sup>1,5,7,32</sup> Ying Su,<sup>1,5,7</sup> Filipe C. Martins,<sup>1,14,15</sup> Mary Jo Fackler,<sup>11</sup> Marina Bessarabova,<sup>16</sup> Adam Kowalczyk,<sup>17,18,19</sup> Thomas Conway,<sup>17,19</sup> Bryan Beresford-Smith,<sup>17,19</sup> Geoff Macintyre,<sup>17,19</sup> Yu-Kang Cheng,<sup>2,9</sup> Zoila Lopez-Bujanda,<sup>11</sup> Antony Kaspi,<sup>23</sup> Rong Hu,<sup>5</sup> Judith Robens,<sup>8,24</sup> Tatiana Nikolskaya,<sup>16</sup> Vilde D. Haakensen,<sup>25</sup> Stuart J. Schnitt,<sup>8,24</sup> Pedram Argani,<sup>12</sup> Gabrielle Ethington,<sup>26</sup> Laura Panos,<sup>26</sup> Michael Grant,<sup>26</sup> Jason Clark,<sup>26</sup> William Herlihy,<sup>26</sup> S. Joyce Lin,<sup>20</sup> Grace Chew,<sup>21</sup> Erik W. Thompson,<sup>21,22,27</sup> April Greene-Colozzi,<sup>3</sup> Andrea L. Richardson,<sup>3,6,8</sup> Gedge D. Rosson,<sup>13</sup> Malcolm Pike,<sup>28</sup> Judy E. Garber,<sup>1,4,5,7</sup> Yuri Nikolsky,<sup>16</sup> Joanne L. Blum,<sup>26</sup> Alfred Au,<sup>29</sup> E. Shelley Hwang,<sup>29</sup> Rulla M. Tamimi,<sup>5,9</sup> Franziska Michor,<sup>2,9</sup> Izhak Haviv,<sup>17,20,23,30</sup> X. Shirley Liu,<sup>2,9</sup> Saraswati Sukumar,<sup>11,\*</sup> and Kornelia Polyak<sup>1,5,7,31,\*</sup>

<sup>1</sup>Department of Medical Oncology

<sup>2</sup>Department of Biostatistics and Computational Biology

<sup>3</sup>Department of Cancer Biology

<sup>4</sup>Center for Clinical Cancer Genetics

Dana-Farber Cancer Institute, Boston, MA 02215, USA

<sup>5</sup>Department of Medicine

<sup>6</sup>Department of Pathology

Brigham and Women's Hospital, Boston, MA 02115, USA

<sup>7</sup>Department of Medicine

<sup>8</sup>Department of Pathology

Harvard Medical School, Boston, MA 02115, USA

<sup>9</sup>Harvard School of Public Health, Boston, MA 02115, USA

<sup>10</sup>Department of Medicine, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer, 08036 Barcelona, Spain

<sup>11</sup>Department of Oncology

<sup>12</sup>Department of Pathology

<sup>13</sup>Department of Plastic Surgery

Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA

<sup>14</sup>Obstetrics and Gynaecology Department, Coimbra University Hospital, 3000-354 Coimbra, Portugal

<sup>15</sup>Programa Gulbenkian de Formação Médica Avançada, 1067-001 Lisboa, Portugal

<sup>16</sup>Thomson Reuters Healthcare & Science, Encinitas, CA 92024, USA

<sup>17</sup>NICTA Victoria Research Laboratory

<sup>18</sup>Department of Electrical and Electronic Engineering

<sup>19</sup>Department of Computer Science and Software Engineering

<sup>20</sup>Department of Pathology

<sup>21</sup>Department of Surgery

<sup>22</sup>Department of Biochemistry

The University of Melbourne, Parkville, VIC 3010, Australia

<sup>23</sup>Bioinformatics and System Integration, Baker IDI Heart and Diabetes Institute, Prahran 3004, VIC Australia

<sup>24</sup>Department of Pathology, Beth-Israel Deaconess Medical Center, Boston, MA 02115, USA

<sup>25</sup>Department of Genetics, Institute for Cancer Research and the KG Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, University of Oslo, Oslo University Hospital, 0424 Oslo, Norway

<sup>26</sup>Baylor–Charles A. Sammons Cancer Center, Dallas, TX 75246, USA

<sup>27</sup>St. Vincent's Institute, Fitzroy 3065, VIC Australia

<sup>28</sup>Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089, USA

<sup>29</sup>Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA

<sup>30</sup>Peter MacCallum Cancer Centre, East Melbourne 3002, VIC, Australia

<sup>31</sup>Harvard Stem Cell Institute, Cambridge, MA 02138, USA

<sup>32</sup>These authors contributed equally to this work

\*Correspondence: saras@jhmi.edu (S.S.), kornelia\_polyak@dfci.harvard.edu (K.P.)

<http://dx.doi.org/10.1016/j.stem.2013.05.004>

## SUMMARY

Early full-term pregnancy is one of the most effective natural protections against breast cancer. To investigate this effect, we have characterized the global

gene expression and epigenetic profiles of multiple cell types from normal breast tissue of nulliparous and parous women and carriers of BRCA1 or BRCA2 mutations. We found significant differences in CD44<sup>+</sup> progenitor cells, where the levels of many



stem cell-related genes and pathways, including the cell-cycle regulator p27, are lower in parous women without BRCA1/BRCA2 mutations. We also noted a significant reduction in the frequency of CD44<sup>+</sup>p27<sup>+</sup> cells in parous women and showed, using explant cultures, that parity-related signaling pathways play a role in regulating the number of p27<sup>+</sup> cells and their proliferation. Our results suggest that pathways controlling p27<sup>+</sup> mammary epithelial cells and the numbers of these cells relate to breast cancer risk and can be explored for cancer risk assessment and prevention.

## INTRODUCTION

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

The protective effect of pregnancy is also observed in animal models and can be mimicked by hormonal factors (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 tumorigenesis are largely undefined. Hypotheses proposed include induction of differentiation, decreased susceptibility to carcinogens, reduction in cell proliferation and in stem cell number, and altered systemic environment due to a decrease in circulating growth hormone and other endocrine factors (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 rodents and mostly focused on the mammary gland. Global gene expression profiling of mammary glands from virgin and parous rats identified changes in TGF- $\beta$  and IGF signaling and in the expression of extracellular matrix proteins (Blakely et al., 2006; D'Cruz et al., 2002). Related studies in humans also identified consistent differences in gene expression profiles between nulliparous and parous women (Asztalos et al., 2010; Belitskaya-Lévy et al., 2011; Russo et al., 2008, 2012). Nevertheless, because these studies have used mammary gland or 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 normal cell of origin of breast carcinomas, and breast cancer risk factors 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 so 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 remains elusive and has not yet been analyzed in humans.

Here, we describe the detailed molecular characterization of luminal and myoepithelial cells, lineage-negative (lin<sup>-</sup>) cells with progenitor features, and stromal fibroblasts from nulliparous and parous women including BRCA1/BRCA2 mutation carriers, the identification of cell-type-specific differences related to parity, functional validation of hormonal factors and selected parity-related pathways on the proliferation of mammary epithelial cells, and the relevance of these to breast cancer risk.

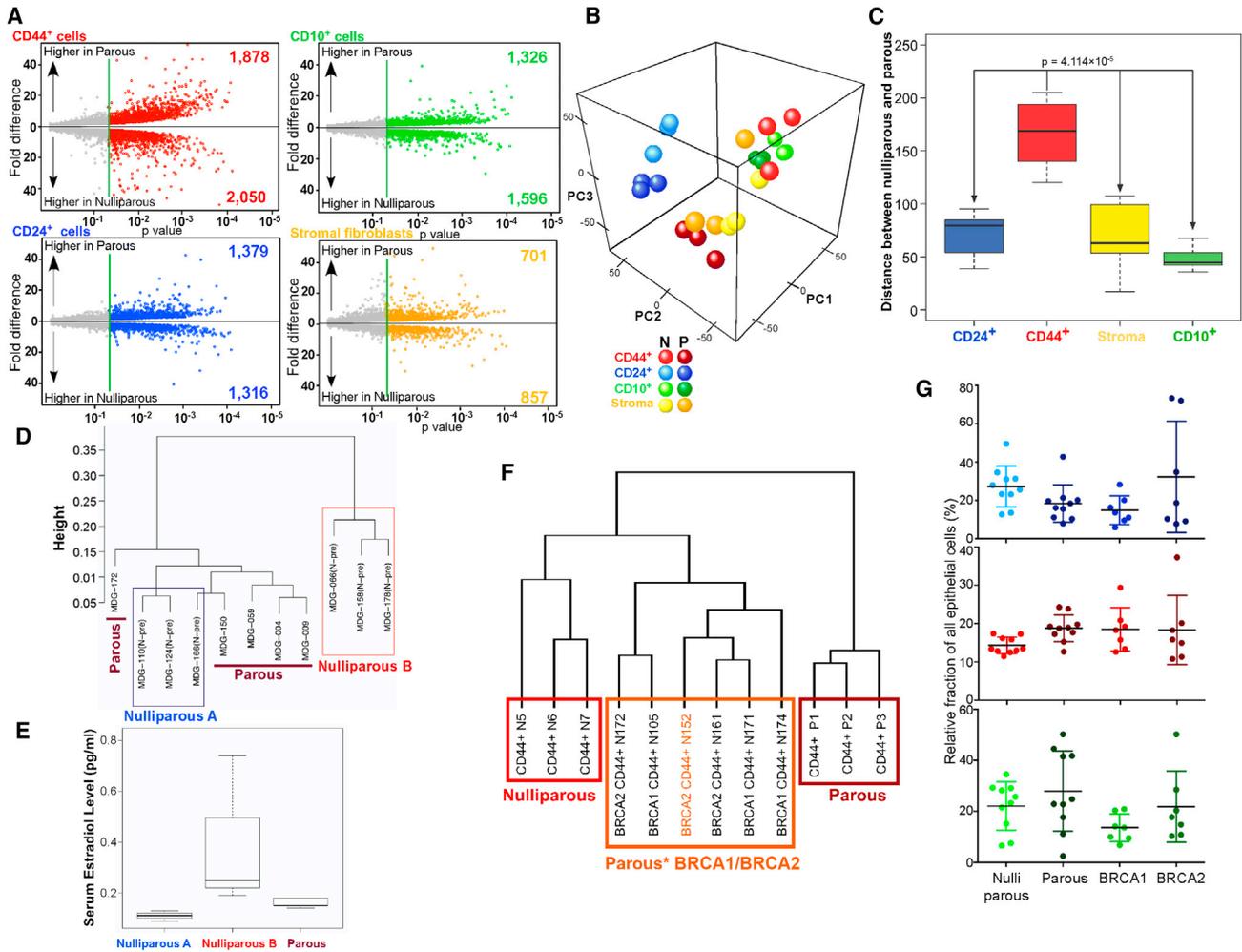
## RESULTS

### Parity-Related Differences in Gene Expression Patterns

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 (CD24), myoepithelial (CD10), and progenitor features (lin<sup>-</sup>/CD44<sup>+</sup>) (Bloushtain-Qimron et al., 2008; Mani et al., 2008; Shipitsin et al., 2007). Cells stained for these markers showed minimal overlap both in nulliparous and parous tissues, with CD24<sup>+</sup> and CD44<sup>+</sup> fractions being especially distinct (Figures S1A and S1B available online). The fraction of CD44<sup>+</sup> cells was slightly higher in parous compared to nulliparous samples, likely due to the more-developed lobulo-alveolar structures in parous women (Russo et al., 2001) that appear to contain many CD44<sup>+</sup> cells (Figures S1B and S1C). We also performed multicolor immunofluorescence analyses for these three 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 gene expression profiles, we analyzed immunomagnetic bead-purified (Bloushtain-Qimron et al., 2008; Shipitsin et al., 2007) CD24<sup>+</sup>, CD10<sup>+</sup>, and CD44<sup>+</sup> cells (captured sequentially in this order; thus, CD44<sup>+</sup> fraction is CD24<sup>-</sup>CD10<sup>-</sup>CD44<sup>+</sup>, but the CD24<sup>+</sup> fraction may contain CD24<sup>+</sup>CD44<sup>+</sup> cells) 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., 2012). To minimize variability unrelated to parity status, women were closely matched for age, number of pregnancies, age at first and time since last pregnancy, and ethnicity (Table S1). The expression of known cell-type-specific genes was consistently observed in each cell type from nulliparous and parous samples based on SAGE-seq confirming the purity and identity of the cells (Figure S2A).

Comparison of each cell type between nulliparous and parous samples revealed the most pronounced differences in CD44<sup>+</sup> cells, where the numbers of significantly ( $p < 0.05$ ) differentially expressed genes and the fold differences were the largest (Figure 1A; Table S2). The degrees of differences were smaller and similar in CD10<sup>+</sup> and CD24<sup>+</sup> cells, whereas stromal fibroblasts had the fewest differentially expressed genes (Table S2). Further examination using principal component analysis (PCA) confirmed that CD24<sup>+</sup> and CD10<sup>+</sup> cells and fibroblasts from



**Figure 1. Cell-type-Specific Differences in Gene Expression According to Parity and BRCA1/BRCA2 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 and numbers indicate  $p = 0.05$  and genes differentially expressed at  $p < 0.05$ , respectively.

(B) 3D 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) Bar plot of the paired Euclidean distance for each of the four cell types. p value indicates the significance of difference (Kolmogorov-Smirnov test) between parous and nulliparous groups in CD44<sup>+</sup> and other cell types.

(D) Hierarchical clustering of Norwegian cohort based on Pearson correlation using genes differentially expressed in CD44<sup>+</sup> cells.

(E) SELs for the samples corresponding to (D).

(F) Hierarchical clustering of CD44<sup>+</sup> cells from nulliparous and parous control women and parous BRCA1/BRCA2 mutation carriers with the exception of one BRCA2 sample (N152, highlighted in orange) that was nulliparous.

(G) Relative frequency of CD44<sup>+</sup>, CD24<sup>+</sup>, and CD10<sup>+</sup> cells. Ten samples were analyzed from each of the indicated groups. Each dot represents an individual sample. Error bars represent mean  $\pm$  SEM.

Boxes in (C) and (E) correspond to the first (Q1) to the third (Q3) quartile, the line within the box is the median, and whiskers are from Q1 – 1.5  $\times$  IQR (interquartile range) = Q3 – Q1) to Q3 + 1.5  $\times$  IQR. See also Figures S1 and S2, and Tables S1, S2, and S3.

nulliparous and parous women were highly similar, whereas CD44<sup>+</sup> cells formed very distinct nulliparous and parous clusters (Figure 1B). In line with this, CD44<sup>+</sup> cells demonstrated the largest distance in gene expression patterns between nulliparous and parous samples (Figure 1C).

To validate our findings in an independent data set, we analyzed the levels of our differentially expressed genes in a Norwegian cohort (Haakensen et al., 2011a, 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 1D). Using genes differentially expressed in all four cell types combined or only in CD44<sup>+</sup> cells gave identical results. Interestingly, nulliparous samples that formed a distinct cluster (nulliparous B) or were closer to parous cases (nulliparous A) displayed significant differences in serum estradiol levels (SELs) with samples more similar to parous cases having low SELs (Figure 1E). Because SEL and breast epithelial cell proliferation are

higher in the luteal phase of the menstrual cycle, our findings imply that cells from nulliparous and parous women may be more distinct in luteal phase.

The expression of selected genes was validated in additional samples by quantitative RT-PCR (qRT-PCR) using CD44<sup>+</sup> cells from multiple nulliparous and parous cases to confirm SAGE-seq data (Figure S2B). Based on these findings in gene expression profiles, we focused our follow-up studies on CD44<sup>+</sup> and CD24<sup>+</sup> cells.

### Lack of Parity-Associated Differences in BRCA1/BRCA2 Mutation Carriers

To strengthen our hypothesis that the parity-associated differences we detected in CD44<sup>+</sup> cells might be related to the risk for breast cancer, we analyzed the gene expression profiles of CD44<sup>+</sup> cells from parous BRCA1/BRCA2 mutation carriers (Table S3), whose risk is not decreased by parity (Cullinane et al., 2005). CD44<sup>+</sup> cells from parous BRCA1/BRCA2 mutation carriers clustered with CD44<sup>+</sup> cells from nulliparous controls (Figure 1F), suggesting that parity-associated changes observed in non-BRCA1/BRCA2 carriers may not occur in these high-risk women.

To determine if the lack of parity-associated changes in CD44<sup>+</sup> cells from BRCA1/BRCA2 women is due to differences in the distribution of cell populations, we performed FACS analysis of multiple tissue samples from BRCA1/BRCA2 and noncarrier women. The relative frequency of CD44<sup>+</sup> was slightly higher in parous compared to nulliparous samples, which was associated with a slight decrease in CD24<sup>+</sup> cells, whereas the relative frequency of CD10<sup>+</sup> cells was about the same in all groups (Figure 1G). The increase in ratio of CD44<sup>+</sup>-to-CD24<sup>+</sup> cells in parous samples could potentially be due to the increased number of lobulo-alveolar structures observed in parous women (Figure S1B) (Russo et al., 2001) or due to the loss of CD24<sup>+</sup> cells during involution.

### 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. Because early pregnancy specifically decreases the risk for ER<sup>+</sup> breast tumors, we first explored our differentially expressed gene lists in CD44<sup>+</sup> cells for candidate mediators of this effect. We found several genes that may change cellular response to steroid hormones by altering metabolism (e.g., *HSD17B11*) or by modulating nuclear receptors (e.g., *NCOR1*) (Table S2). Androgen receptor (*AR*) and one of its key targets PSA (*KLK3*) were highly expressed in nulliparous CD44<sup>+</sup> cells, implying active androgen signaling that is decreased by parity. Among genes highly expressed in parous CD44<sup>+</sup> cells were a number of known tumor suppressors, including *CASP8* (Cox et al., 2007), *SCRIB* (Humbert et al., 2008), and DNA repair genes (e.g., *PRKDC*).

To determine overall activation of specific biological functions due to parity, we performed pathway enrichment, network, and protein interactome analyses using the MetaCore platform (Nikolsky et al., 2009). We found that parity has similar global effects on three of the four cell types analyzed because pathways built on expression patterns in CD10<sup>+</sup> and CD44<sup>+</sup> cells and

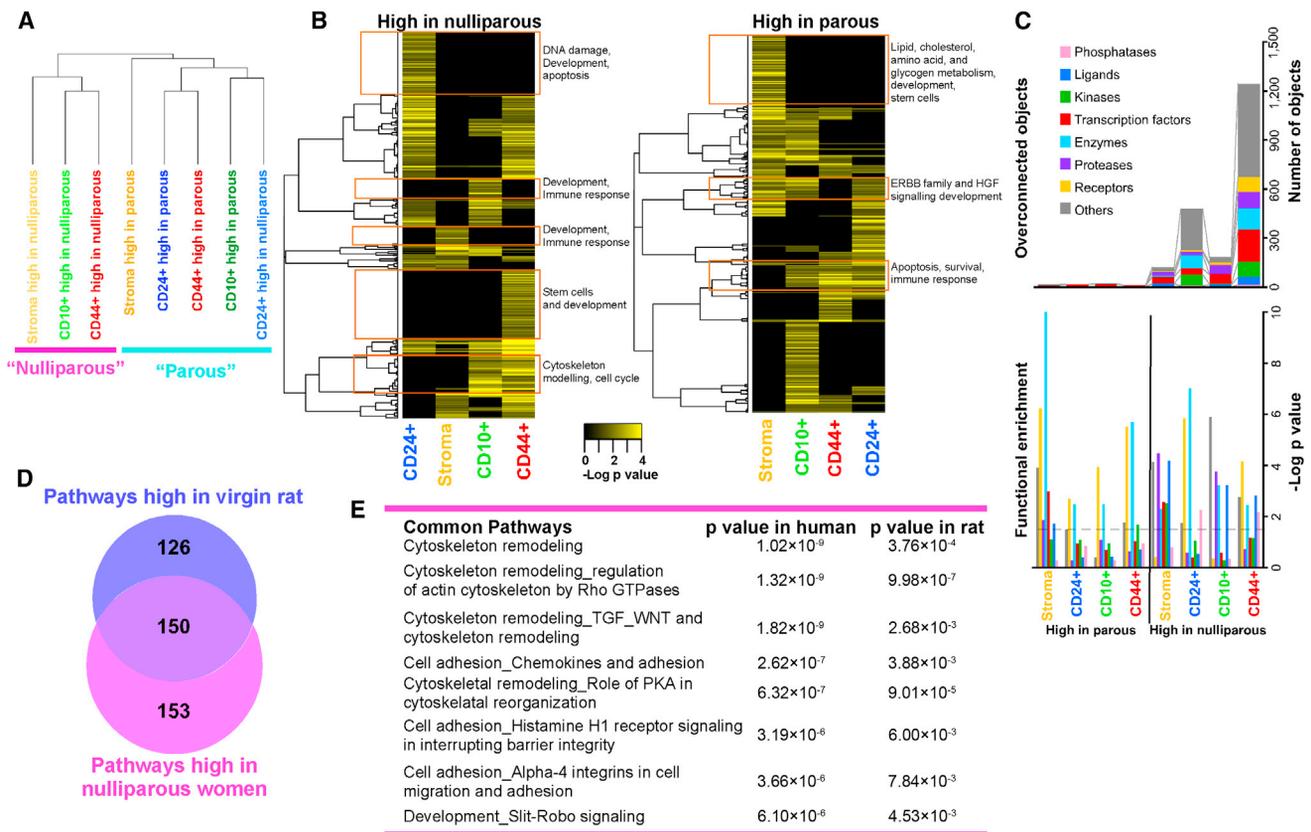
stroma cluster together for parous and nulliparous states (Figure 2A). The most significant pathways highly active in parous samples in all three cell types included apoptosis, survival, and immune response, whereas stem cells and development-related pathways were enriched only in CD44<sup>+</sup> cells from nulliparous women (Figure 2B; Table S4). Pathways highly active in parous stroma were enriched in fatty acid metabolism and adipocyte differentiation, 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 2C; Table S5).

We focused our further analysis on CD44<sup>+</sup> 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, DNA methylation, and WNT signaling, whereas pathways more active in parous samples include PI3K/AKT signaling and apoptosis (Table S4). Importantly, the highest-scored pathway for high-in nulliparous genes is four orders of magnitude more statistically significant than those for high-in-parous genes, suggesting that downregulation of protumorigenic developmental pathways is a prominent feature of CD44<sup>+</sup> cells from parous women. 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<sup>+</sup> cells (Figure 2C). Because the relative number of interactions (connectivity) is directly related to the functional activity of a data set (Nikolsky et al., 2008), these results suggest that parous cells are substantially less active than nulliparous ones. The most overconnected (and overexpressed) transcription factor (TF) in nulliparous CD44<sup>+</sup> cells is SCMH-1, a component of PRC1, which is required for the repression of many genes during development and for the maintenance of hematopoietic stem cells (Ohtsubo et al., 2008). The causal network assembled from the top-scored pathways and overconnected genes high in nulliparous CD44<sup>+</sup> cells included a number of tumorigenic pathways. The network's key "triggers" (i.e., secreted ligands) comprise IL-6, VEGFA, PDGF-B, CCL-2, NOTCH, and BMP4, signaling through major hubs such as PI3K, GSKβ, β-catenin, RhoA/Rac1, MEK3, and MEK4. The network activated in CD44<sup>+</sup> cells from parous women featured IL-10, IL-23, TGF-β2 as ligands, β3 adrenergic receptor, and TFs STAT1, STAT4, STAT5, and NF-κB.

We also explored pathways significantly different in CD44<sup>+</sup> cells from control parous and BRCA1/BRCA2 mutation carriers. Very few pathways were common between cells from BRCA1/BRCA2 mutation carriers, implying that although both are distinct compared to control parous women, they are also different from one another. Several of the top-scoring pathways in cells from BRCA1 mutation carriers relate to DNA damage, cell cycle, and apoptosis, whereas in cells of BRCA2 mutation carriers, many significantly high-scoring pathways are involved in stem cells such as WNT, Slit-Robo, and IGF signaling (Table S4).

### Conservation of Parity-Associated Pathways across Species

Because pregnancy-induced protection against mammary tumors is also observed in rodents, we investigated whether



**Figure 2. Signaling Pathways Affected by Parity-Related Differences in Gene Expression Patterns**

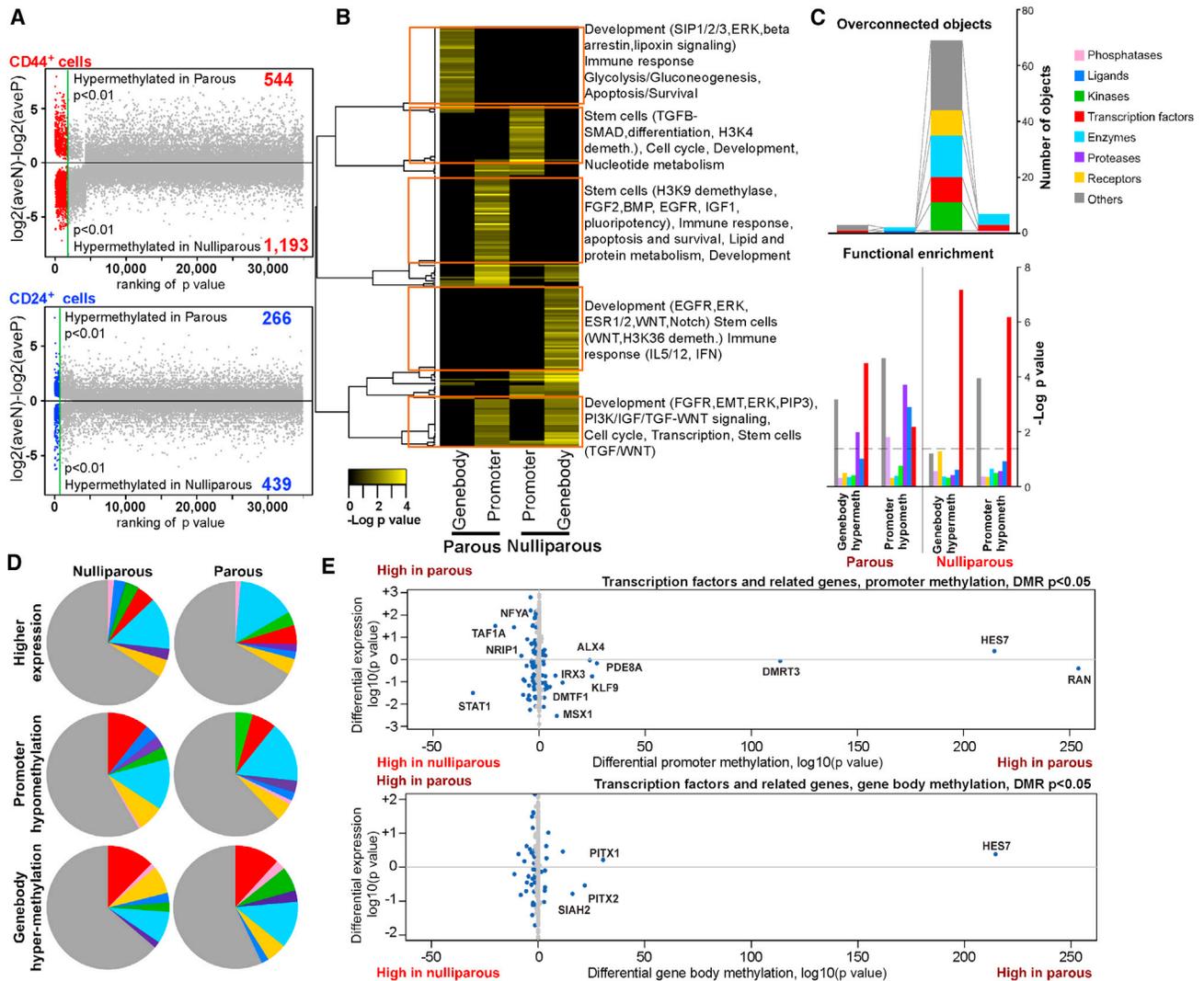
(A) Dendrogram depicting hierarchical clustering of signaling pathways significantly high in parous or nulliparous samples in any of the four cell types analyzed. (B) Heatmap depicting unsupervised clustering of signaling pathways significantly down- or upregulated in parous compared to nulliparous samples 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. (C) Genes differentially expressed between nulliparous and parous samples in each of the four cell types were analyzed for relative enrichment with the indicated protein classes (lower panel) and for relative connectivity (upper panel). y axes indicate  $-\log p$  values for enrichment with the listed protein classes or the number of overconnected objects. (D) Venn diagram depicting the number of unique and common pathways high in CD44+ cells from nulliparous women and in mammary glands of virgin rats, respectively. (E) List of top common pathways downregulated in CD44+ cells and mammary glands from parous women and rats, respectively. Name of pathways and p values of enrichment are indicated. See also Tables S4 and S5.

pathways altered by parity are conserved across species. We compared pathways in CD44+ cells to those generated based on genes differentially expressed between virgin and parous rats (Blakely et al., 2006). We found a significant overlap between pathways highly active in nulliparous and virgin samples, with top-ranked pathways, including cytoskeleton remodeling and cell adhesion, known to be highly relevant in stem cells (Figures 2D and 2E). A network built of the common pathways included a complete NOTCH pathway, IGF, EGF, CD44, CD9, and ITGB1 as triggers (i.e., ligands and receptors), c-Src, PKC, and FAK as major signaling kinases, and c-Jun, p53, SNAIL1, and LEF as TFs. Thus, pregnancy appears to induce similar alterations in the mammary gland 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 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 applied to high-throughput sequencing) (Hu et al., 2005) 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. In both cell types, more DMRs were hypermethylated in nulliparous than in parous cells (Figure 3A; Table S6). The differences in methylation of selected genes were validated in additional samples by quantitative methylation-specific PCR (qMSP) using CD44+ cells from multiple nulliparous and parous cases to confirm MSDK-seq data (Figure S3A).



**Figure 3. Epigenetic Differences between Nulliparous and Parous Tissues**

(A) Genome-wide view of differentially methylated genes in CD24<sup>+</sup> and CD44<sup>+</sup> 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<sup>+</sup> or CD24<sup>+</sup> 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 number of significant DMRs ( $p < 0.01$ ) is shown.

(B) Pathways enriched with genes in CD44<sup>+</sup> cells with the indicated difference in DNA methylation between nulliparous and parous women.

(C) Genes with promoter and gene body DMRs in CD44<sup>+</sup> cells from nulliparous and parous samples were analyzed for relative enrichment with the indicated protein classes and for relative connectivity. y Axes indicate  $-\log p$  values for enrichment with the listed protein classes or the number of overconnected objects.

(D) Pie charts depicting the relative percentage (%) of genes in different functional categories with the indicated gene expression and DNA methylation pattern in CD44<sup>+</sup> cells from nulliparous and parous women.

(E) Scatterplot for MSDK-seq and SAGE-seq data to depict correlations between differential promoter methylation and differential gene expression for TFs. Each point represents a gene with a MSDK-seq site in a certain region (promoter,  $-5$  to  $+2$  kb from TSS; gene body,  $+2$  kb from TSS to the end of gene), and  $\log_{10} p$  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 by  $\log_{10} p$  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.

See also Figure S3, and Tables S5, S6, and S7.

To investigate pathways affected by parity-related epigenetic alterations, we analyzed pathways enriched by genes associated with gene body or promoter DMRs in CD44<sup>+</sup> cells from nulliparous and parous samples and found very little overlap among the four distinct categories (Figure 3B). Promoter hypermethylation in parous CD44<sup>+</sup> cells involved stem cell (e.g., H3K9 demethylase, FGF2, BMP, EGFR, pluripotency) and apoptosis

and survival pathways, whereas promoter hypermethylation in nulliparous CD44<sup>+</sup> cells involved stem cell (e.g., TGF- $\beta$ /SMAD, H3K4 demethylase) pathways, cell cycle, and nucleotide metabolism.

The fraction of TFs among differentially methylated genes is 2- to 3-fold higher than expected and than observed for differentially expressed genes, implying that promoter methylation is a

preferred control mechanism of their expression (Figures 3C and 3D). Similar to the expression data, DMRs in nulliparous samples had higher numbers of overconnected objects than in parous ones. Gene body DMRs in nulliparous CD44<sup>+</sup> cells had the highest number of overconnected objects, and TFs represented a significant fraction of overconnected objects in promoter hypermethylated DMRs in nulliparous CD44<sup>+</sup> cells (Figure 3C).

We also analyzed associations between differential gene expression and presence of DMRs in CD44<sup>+</sup> and CD24<sup>+</sup> cells. However, we found no significant correlation at the global scale (data not shown), potentially due to the complex relationship between DNA methylation and transcription because DNA methylation can have both positive and negative effects on gene expression, depending on the location relative to the transcription start site (Jones, 1999). Even so, the expression of several TFs with key roles in stem cells (e.g., HES7, STAT1) was correlated with the degree of promoter or gene body DNA methylation (Figure 3E). Our results correlate with recent findings that most differentially expressed genes in multiple tissue types do not show significant differences in DNA methylation except for TFs (Bock et al., 2012).

Because of the importance of TF networks in parity-associated epigenetic changes, we also explored the potential of DMRs to modulate the action of specific TFs relevant to development and tumorigenesis by searching for enrichment of specific TF binding sites (TFBSs) within nulliparous or parous-specific DMRs using the ENCODE Regulation Supertrack (Birney et al., 2007). Of the 144 documented TFs, 45 exhibited significant (adjusted  $p < 0.05$ , Benjamini Hochberg corrected) enrichment in the DMRs, and a number of these (e.g., TCF4, STAT3, and CEBPB) were significantly differentially enriched between parous and nulliparous DMRs (Figure S3B), implying that parity-induced epigenetic changes can affect the availability of DNA for specific factors and, therefore, may affect the etiology of breast cancer.

Analysis of the H3K27me3 profiles of CD44<sup>+</sup> or CD24<sup>+</sup> cells from nulliparous and parous samples did not detect significant parity-related differences (Figure S3C; data not shown). However, high-in nulliparous genes in either cell type were never K27 enriched, thereby implying the potential lack of regulation by the PRC2 that establishes this histone mark (Table S7). 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 TFs with key roles in development and differentiation.

### Persistent Parity-Related Decrease in p27<sup>+</sup> Breast Epithelial Cells

*CDKN1B* encoding for p27 was one of the most significantly differentially expressed genes in CD44<sup>+</sup> cells between nulliparous and parous (high in nulliparous) and also control and *BRCA1/BRCA2* (high in *BRCA1/BRCA2*, Table S3) comparisons. p27 is known to affect the number and proliferation of stem cells and progenitors in several organs in mice (Cheng et al., 2000; Oesterle et al., 2011). Thus, higher expression of p27 in CD44<sup>+</sup> cells from nulliparous control and parous *BRCA1/BRCA2* mutation carrier women may indicate higher numbers of mammary epithelial progenitors in these samples. To investigate this

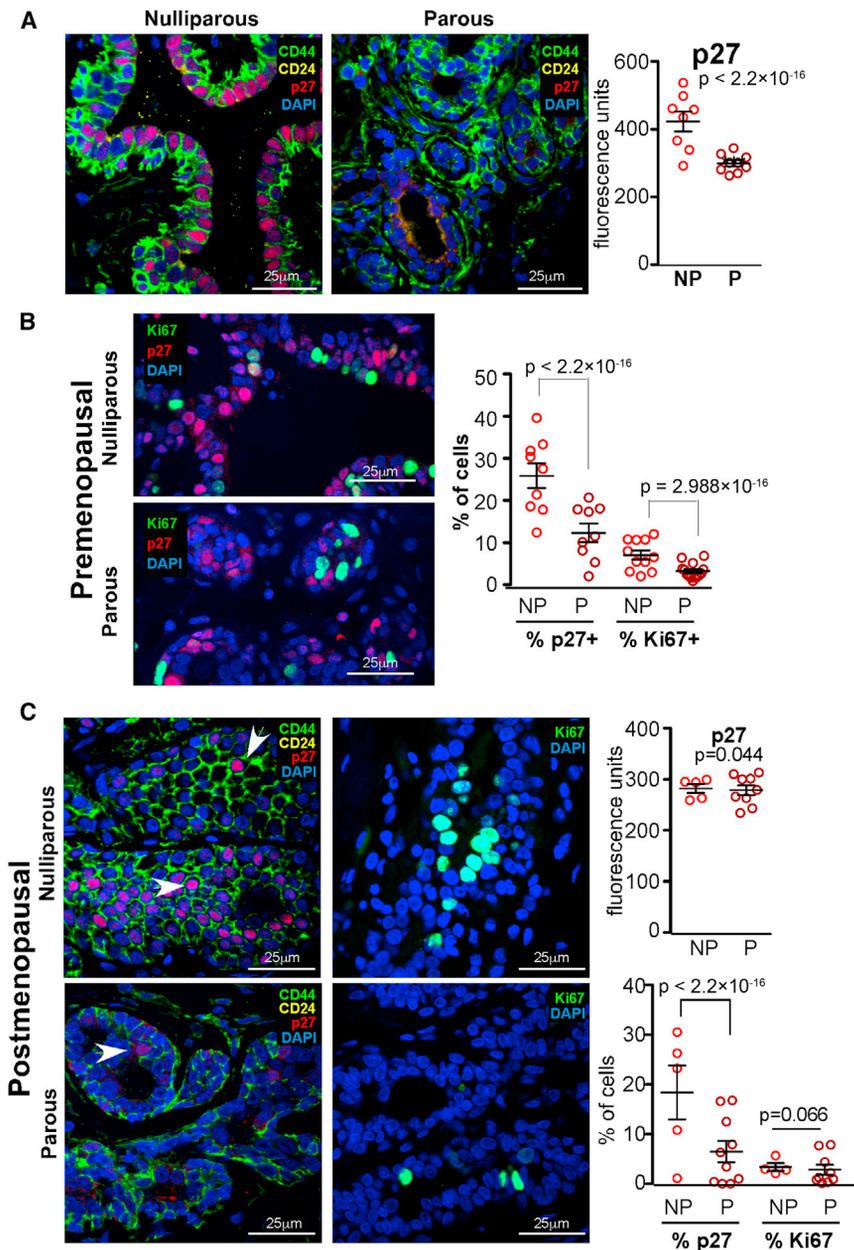
hypothesis, we performed immunofluorescence analysis for p27 alone and in combination with CD24 and CD44 and Ki67 proliferation markers in both premenopausal and postmenopausal tissues to confirm that the parity-related differences we detected by global profiling of premenopausal women are maintained after menopause. We observed that p27 expression and the number of p27<sup>+</sup> cells were significantly lower in parous compared to nulliparous samples from both pre- and postmenopausal women (Figures 4A–4C). The frequency of Ki67<sup>+</sup> cells was also significantly higher in nulliparous than parous cases, and Ki67<sup>+</sup> cells were rarely p27<sup>+</sup> (Figures 4B and 4C).

To strengthen the link between the frequency of p27<sup>+</sup> cells and parity-related decrease in postmenopausal breast cancer risk, we analyzed postmenopausal nulliparous and parous women with or without breast cancer. Although cancer-free nulliparous postmenopausal women exhibited a higher fraction of p27<sup>+</sup> cells than parous ones, the frequency of these cells was the highest in parous postmenopausal women with breast cancer (Figure S4A). Interestingly, the difference in the frequency of p27<sup>+</sup> cells between control women and patients with breast cancer was pronounced only in the postmenopausal parous group (Figure S4A).

### p27<sup>+</sup> Cells Are Quiescent Hormone-Responsive Cells with Progenitor Features

The mutually exclusive expression of Ki67 and p27 in breast epithelial cells and their concordant decrease in parous women implied that they might represent cycling and quiescent cells with proliferative potential, respectively. Ovarian hormones are the best-understood regulators of breast epithelial cell proliferation and also breast cancer risk (Briskin and O'Malley, 2010). Our gene expression data indicated a decrease in AR signaling in CD44<sup>+</sup> cells from parous women (Table S2), and prior studies implied a decrease in ER<sup>+</sup> breast epithelial cells in parous compared to nulliparous women (Taylor et al., 2009).

To explore the potential hormonal regulation of p27<sup>+</sup> breast epithelial cells, we analyzed the expression of p27, ER, and AR in breast tissue samples from women of varying parity and hormonal status. These included nulliparous and parous women, *BRCA1/BRCA2* mutation carriers, women in early (8–10 weeks) and late (22–26 weeks) stages of pregnancy, and premenopausal women in the follicular and luteal phases of the menstrual cycle or subject to ovarian hyperstimulation prior to oocyte collection for in vitro fertilization. Multiple different regions of the tissue samples were examined to account for tissue heterogeneity. We found that most p27<sup>+</sup> cells were also ER<sup>+</sup>, and their numbers were highest in *BRCA1/BRCA2* mutation carriers and the lowest in pregnant women and after ovarian hyperstimulation, where both ovarian hormones and HCG (human chorionic gonadotropin) levels are the highest (Figure 5A). The frequencies of p27<sup>+</sup>, ER<sup>+</sup>, and p27<sup>+</sup>ER<sup>+</sup> cells were also higher in nulliparous compared to parous women and in follicular relative to luteal phase of the menstrual cycle (Figure 5A). Overall, 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). The high fraction of AR<sup>+</sup> cells in *BRCA1* mutation carriers is particularly interesting because AR is a genetic modifier of *BRCA1*-associated breast cancer risk (Rebbeck et al., 1999).



**Figure 4. 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. Graph shows the quantification of p27-staining intensity in multiple samples.

(B and C) Immunofluorescence staining for p27 and Ki67 in breast tissue from premenopausal (B) and postmenopausal (C) nulliparous and parous women. Graphs show frequencies of p27<sup>+</sup> and Ki67<sup>+</sup> cells in nulliparous and parous samples. Arrowheads in (C) point to CD44<sup>+</sup>p27<sup>+</sup> cells.

p values of differences between nulliparous and parous groups are indicated. Error bars represent median ± SEM. See also Figure S4.

marking proliferating progenitors in early G1 phase of the cell cycle when p27 and Ki67 might overlap. The differences in the frequency of p27<sup>+</sup> and Ki67<sup>+</sup> cells between the follicular and luteal phases were less significant in parous compared to nulliparous women in part due to the lower overall fractions of these cells in parous cases (Figure S4A). These results suggest that a subset of p27<sup>+</sup> cells might represent quiescent hormone-responsive progenitors and that their frequency relates to breast cancer risk.

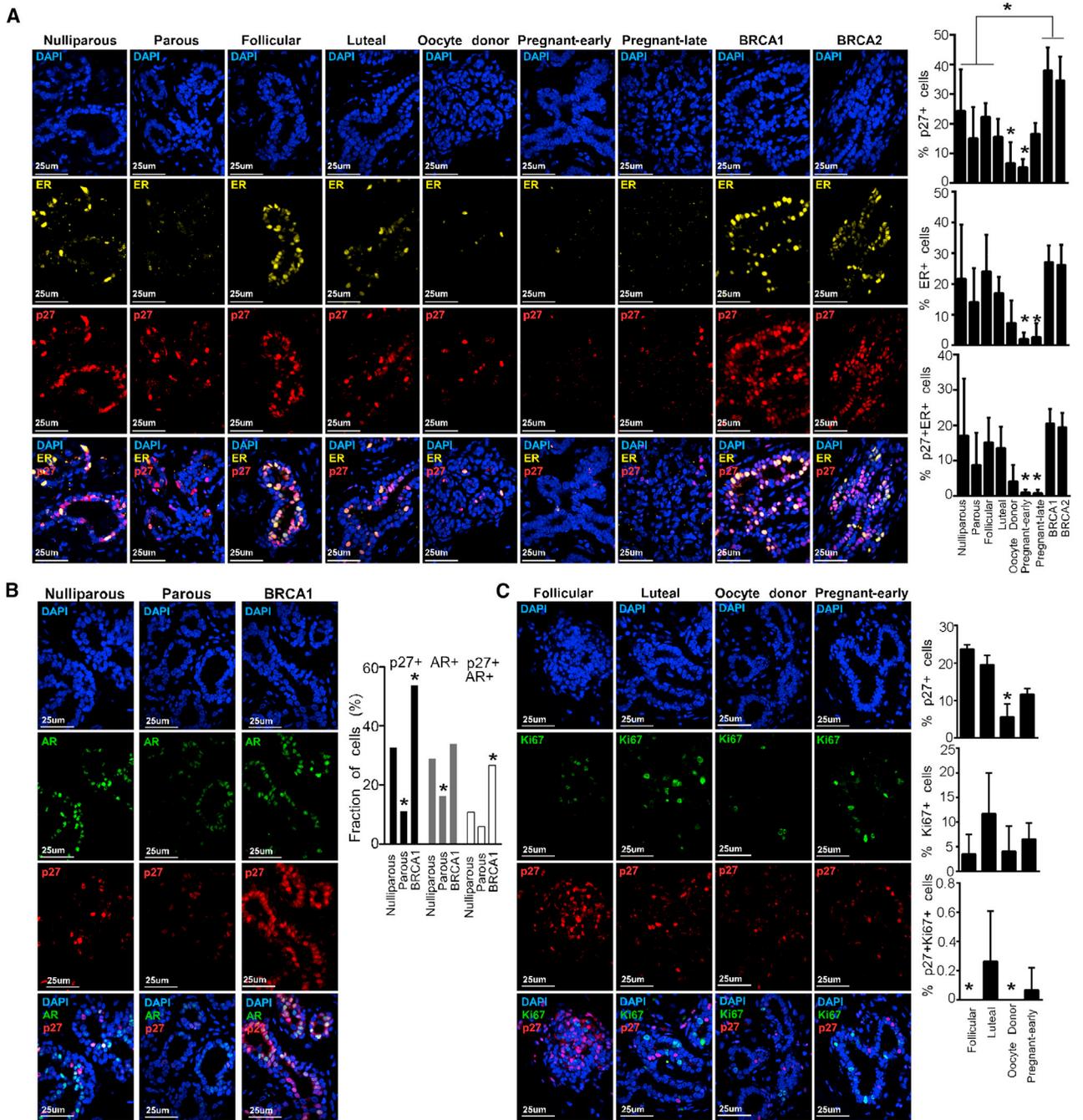
**Functional Validation of Parity-Related Differences in Signaling Pathways**

Several signaling pathways that are less active in CD44<sup>+</sup> cells from parous women were related to stem cells (Figure 2A). To investigate whether inhibition of these pathways affects the number of p27<sup>+</sup> and proliferating cells, we incubated normal breast tissues in a tissue

To further investigate the relationship between the numbers of p27<sup>+</sup> 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 from Chung et al. (2012) and Going et al. (1988), the frequency of Ki67<sup>+</sup> cells was the highest in the luteal phase of the menstrual cycle when both estrogen and progesterone levels are high (Figure 5C). Samples from early pregnancy had a lower fraction of Ki67<sup>+</sup> cells, and the number of these cells was lowest in the follicular phase. The frequency of p27<sup>+</sup> cells displayed an inverse correlation with Ki67<sup>+</sup> cell frequency: it was the highest in the follicular phase and lowest in oocyte donors (Figure 5C). Interestingly, a low but detectable fraction of p27<sup>+</sup> cells was also Ki67<sup>+</sup> in the luteal phase and early pregnancy, potentially

explant culture model with inhibitors or agonists of selected pathways (e.g., EGFR, Hh, TGF-β, and Wnt) for 8–10 days. We tested inhibitors of irrelevant pathways as controls. For each case, we cultured three pieces of breast tissue taken from different regions of the same breast, to minimize variability due to tissue heterogeneity. We then assessed the number of p27<sup>+</sup> cells and cellular proliferation based on bromodeoxyuridine (BrdU) incorporation (S phase cells) and Ki67 (cycling cells).

We found that tissue architecture and cellular viability were maintained, and p27<sup>+</sup>, Ki67<sup>+</sup>, and BrdU<sup>+</sup> cells were detected in all conditions (Figures 6A and 6B). The frequency of p27<sup>+</sup> cells decreased most markedly following TGF-β receptor and IGFR inhibitor treatment (Figure 6C). Treatment with TGF-β receptor inhibitor significantly (p < 0.05) increased, whereas inhibition



**Figure 5. 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. Graphs show frequencies of p27<sup>+</sup>, ER<sup>+</sup>, and p27<sup>+</sup>ER<sup>+</sup> cells in each group of samples.

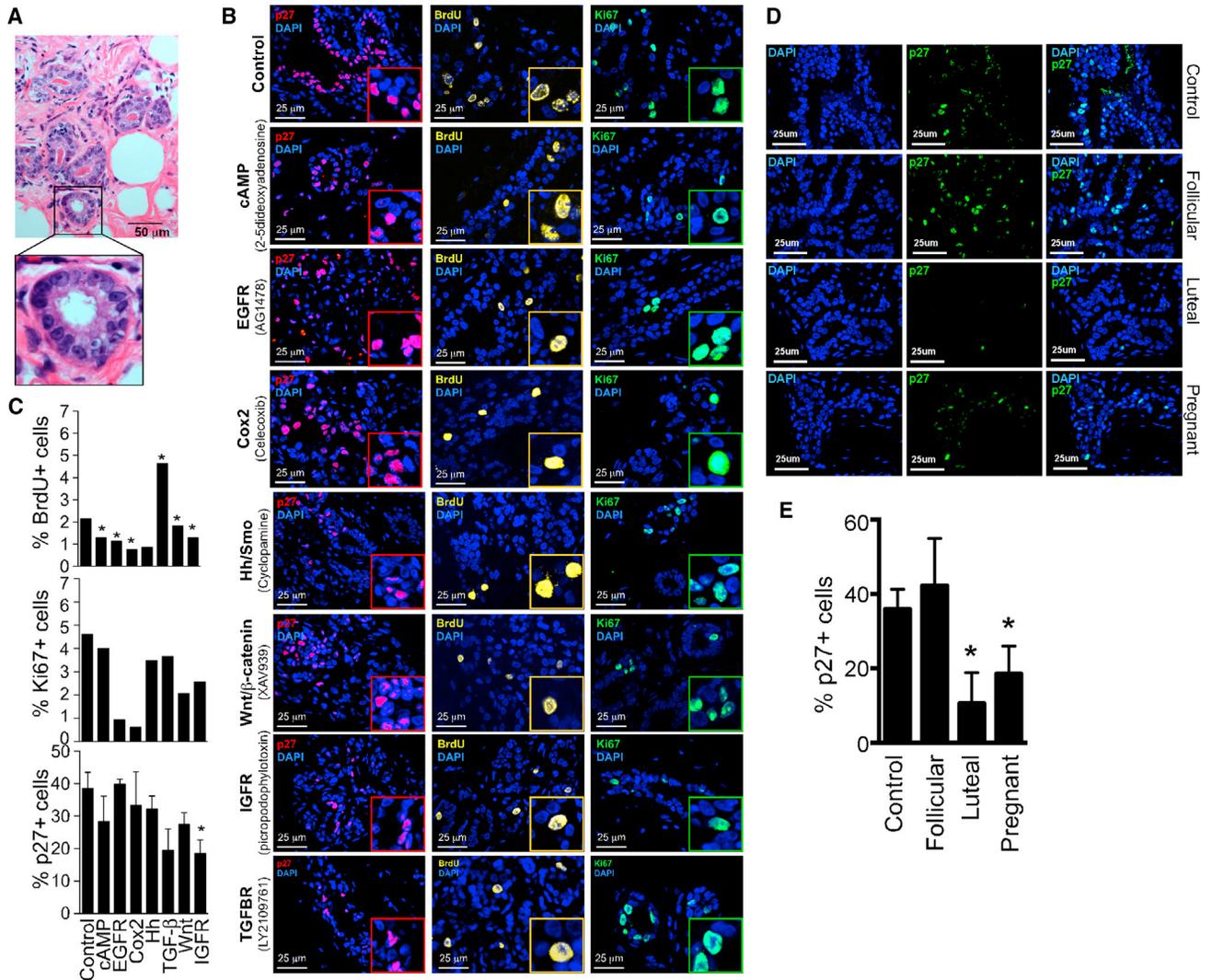
(B) Representative double-immunofluorescence staining for p27 and AR in breast tissue from premenopausal nulliparous and parous women, and in BRCA1 mutation carriers. Graph shows frequencies of p27<sup>+</sup> and AR<sup>+</sup> cells in each set of samples.

(C) Representative double-immunofluorescence staining for p27 and Ki67 in breast tissue from the indicated groups of women. Graphs show frequencies of p27<sup>+</sup>, Ki67<sup>+</sup>, and p27<sup>+</sup>Ki67<sup>+</sup> cells in each group of samples.

Asterisks (\*) indicate significant ( $p \leq 0.05$ , t test or Fisher's exact test) differences between groups of four to eight samples. Error bars represent mean  $\pm$  SD. See also Figure S4.

of cAMP, EGFR, Cox2, Hh, and IGFR signaling decreased the number of BrdU<sup>+</sup> cells, respectively. The fraction of Ki67<sup>+</sup> cells was lower in all inhibitor-treated cultures, with EGFR and

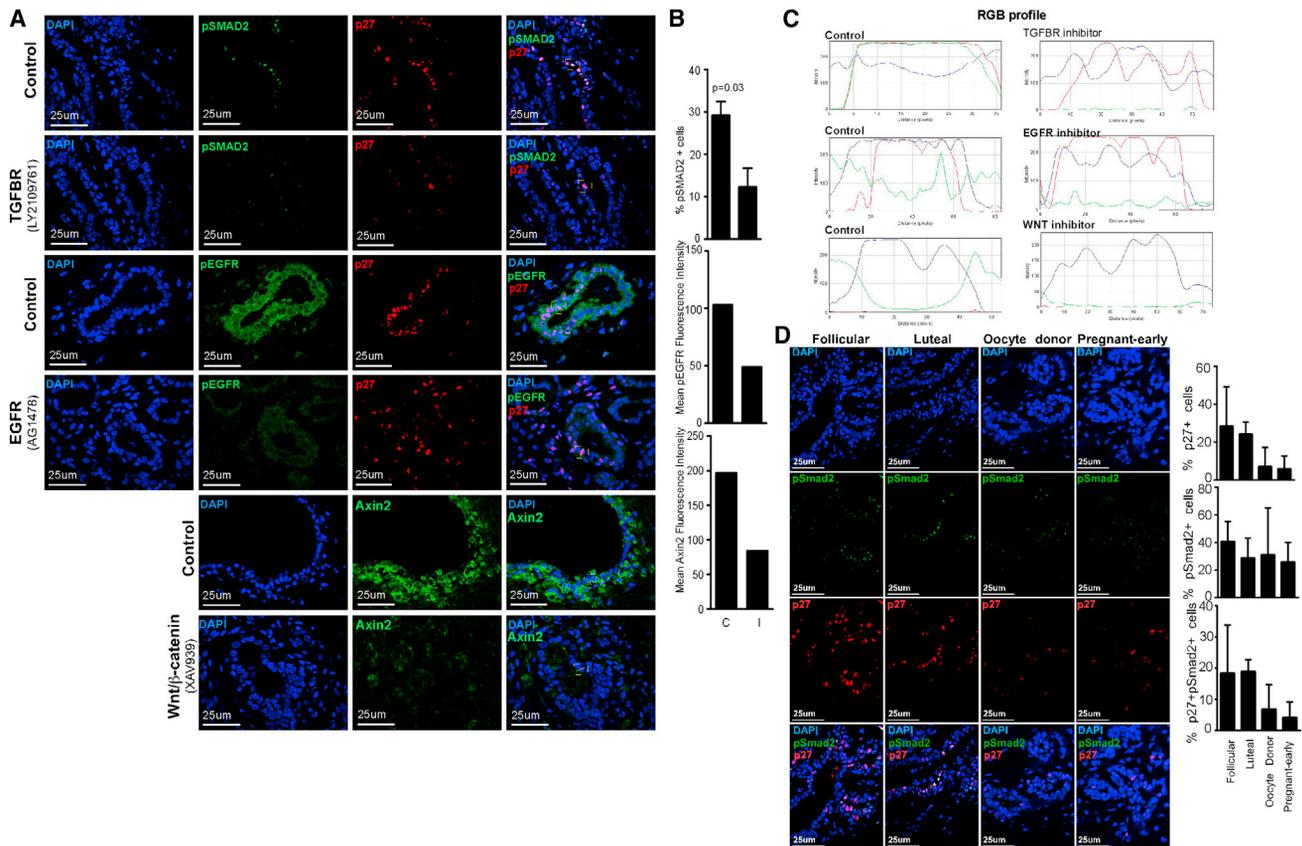
Cox2 inhibition having the most pronounced effects (Figure 6C), whereas stimulation with Shh increased proliferation (Figure S4B).



**Figure 6. Modulation of p27<sup>+</sup> Breast Epithelial Cells and Proliferation by Hormonal and Parity-Related Pathways**  
 (A) Representative hematoxylin and eosin staining depicting morphology of breast tissue after 8 days in culture.  
 (B) Representative examples of multicolor immunofluorescence analyses of BrdU<sup>+</sup>, p27<sup>+</sup>, and Ki67<sup>+</sup> cells in control and tissues treated with inhibitors of the indicated pathways.  
 (C) Frequency of Ki67<sup>+</sup>, BrdU<sup>+</sup>, and p27<sup>+</sup> cells in each of the indicated conditions.  
 (D and E) Representative images of immunofluorescence analysis of p27 and graph depicting the frequency of p27<sup>+</sup> cells in tissue slices from three to four independent cases treated with hormones mimicking the indicated physiologic levels in women.  
 Asterisks indicate significant ( $p \leq 0.05$ ) differences. Error bars represent mean  $\pm$  SD. See also Figure S4.

To determine whether the number and proliferation of p27<sup>+</sup> cells are regulated by estrogen signaling, we analyzed the fraction of p27<sup>+</sup> and Ki67<sup>+</sup> cells in tissue slices treated with varying concentrations of ovarian hormones or tamoxifen. To correlate the tissue slice data with that observed under physiologic conditions (Figure 5), we used estrogen, progesterone, prolactin, and HCG hormone levels that mimic serum levels in the follicular or luteal phases of the menstrual cycle or in midpregnancy. We observed that the number of p27<sup>+</sup> 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 (Figures 6D and S4B). These data further

support our hypothesis that a subset of p27<sup>+</sup> cells is hormone-responsive cells. In search of a direct link between p27<sup>+</sup> cells and the signaling pathways analyzed, we confirmed that the selected pathways were active in p27<sup>+</sup> cells (Figure 7A) and that the compounds effectively inhibited their activity in these cells (Figures 7B and 7C). Most importantly, phospho-Smad2 (pSmad2), a key mediator of TGF-β signaling, demonstrated a significant overlap with p27 both in tissue slices (Figure 7A) and in uncultured patient samples (Figure 7D). The frequency of p27<sup>+</sup>pSmad2<sup>+</sup> cells also fluctuated according to hormone levels displaying inverse correlation with mammary epithelial cell proliferation during the menstrual cycle and pregnancy (Figures 7D and 4A).



**Figure 7. Signaling Pathways Regulating p27<sup>+</sup> 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 (C) and inhibitor-treated (I) tissues.  
 (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 three to four independent cases of the indicated women. Graphs show frequencies of p27<sup>+</sup>, pSmad2<sup>+</sup>, and p27<sup>+</sup>pSmad2<sup>+</sup> cells in each group of samples. Error bars represent mean ± SD.

These results indicate a key role for TGF-β signaling in mammary epithelial cell quiescence state likely via p27 (Polyak et al., 1994). However, alternative hypotheses such as SMAD2-independent paracrine effects of TGF-β cannot be excluded. Thus, these data suggest that the decreased frequency of p27<sup>+</sup> and Ki67<sup>+</sup> 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.

**DISCUSSION**

Our comprehensive data set covering multiple cell types in normal human breast tissue from nulliparous and parous women provides an information resource for future investigation of the way in which parity contributes to a reduction in breast cancer risk. From our analysis of the data, we found that parity has the most pronounced effect on CD44<sup>+</sup> cells enriched for cells with luminal progenitor features. Most of the differences relate to transcriptional repression and downregulation of genes and pathways important for stem cell function including EGF, IGF,

Hh, and TGF-β signaling. Correlating with our findings, high-circulating IGF-1 levels have been associated with increased risk for ER<sup>+</sup> breast cancer (Key et al., 2010). Similarly, germline polymorphism in members of the TGF-β signaling pathway influences breast cancer susceptibility (Scollen et al., 2011).

The gene expression profiles of CD44<sup>+</sup> cells from parous BRCA1/BRCA2 mutation carriers were more similar to nulliparous than to parous noncarriers, implying that parity-related changes may not occur or may be much less apparent in these high-risk women. Our results are consistent with epidemiologic data demonstrating that pregnancy does not decrease breast cancer risk in BRCA1/BRCA2 mutation carriers or only does so after multiple (four or more) pregnancies (Cullinane et al., 2005; Poynter et al., 2010). Interestingly, despite their overall similarity, CD44<sup>+</sup> cells from BRCA1/BRCA2 mutation carriers also displayed significant differences. Pathways related to DNA damage and repair were the highest ranked in BRCA1 cells, correlating with BRCA1's function (Roy et al., 2012). In contrast, top-scoring pathways in BRCA2 cells are involved in stem cell function, development, and differentiation. These results imply that although germline mutations in both genes increase breast

cancer risk, the underlying mechanisms are likely to be distinct, which could also explain the differences in breast tumor subtypes that develop in these high-risk women.

In contrast to the significant parity-related differences in gene expression profiles, cells from nulliparous and parous women showed much less-pronounced differences in the epigenetic patterns we analyzed. With the exception of a subset of TFs, for the majority of differentially expressed genes, differences in transcript levels were not associated with differences in DNA methylation or enrichment for H3K27me3 mark. Although these findings could in part be due to the limitations of the technologies employed, they are consistent with results of studies investigating the epigenetic profiles of hematopoietic and skin stem, progenitor, mature cells in mice (Bock et al., 2012), and our preliminary data in the mouse mammary gland (S.C. and K.P., unpublished data).

One of the intriguing findings of our study is the high number of p27<sup>+</sup> cells in breast tissues of nulliparous women and *BRCA1/BRCA2* mutation carriers with high risk for breast cancer, which seems paradoxical because *CDKN1B/p27<sup>Kip1</sup>* is a tumor suppressor and cell-cycle inhibitor. 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 contrast, the consequences of p27 deficiency in the mouse mammary gland 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<sup>-/-</sup>* 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 mammary 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).

Based on our data, we hypothesize that p27 regulates the proliferation and pool size of hormone-responsive progenitors with proliferative potential. Thus, the lower numbers of these p27<sup>+</sup> cells in control parous women may contribute to their lower breast cancer risk. High p27 and quiescence of these cells are regulated by TGF- $\beta$  signaling, as implied by the colocalization of pSmad2 with p27 and the increase in BrdU incorporation with a concomitant decrease in p27 following TGF- $\beta$  receptor inhibitor treatment. Correlating with the presumed importance of TGF- $\beta$  signaling and p27 in hormone-responsive luminal progenitors, recent whole-genome sequencing studies detected inactivating *CDKN1B* and TGF- $\beta$  pathway mutations in luminal breast tumors (Stephens et al., 2012).

The frequency of p27<sup>+</sup> cells was high in control nulliparous women and even higher in *BRCA1/BRCA2* carriers even though these different groups of women are predisposed to different types of breast cancer. Nulliparous women have increased risk for postmenopausal ER<sup>+</sup> breast cancer (Colditz et al., 2004), whereas *BRCA1* mutation carriers most commonly have

ER<sup>-</sup> basal-like tumors (Maxwell and Domchek, 2012). However, luminal progenitors may serve as potential cell of origin of *BRCA1*-associated breast cancer and other basal-like tumors (Lim et al., 2009; Molyneux et al., 2010). Our data demonstrating increased frequency of hormone-responsive p27<sup>+</sup> cells in all high-risk women support this hypothesis.

In addition to pregnancy itself, the duration of breast-feeding also significantly impacts breast cancer risk, which is especially pronounced for the triple-negative (i.e., ER<sup>-</sup>PR<sup>-</sup>HER2<sup>-</sup>) subtype (Shinde et al., 2010). Thus, because we did not have breast-feeding information on the samples used for our study, further investigation is required to determine whether the changes in molecular profiles and p27<sup>+</sup> cell frequencies in parous women are due to pregnancy itself or are also influenced by length of breast-feeding.

In summary, we here describe global differences in gene expression patterns in human mammary epithelial cells related to parity and identified p27<sup>+</sup> cells with progenitor features as a potential marker of breast cancer risk. The pathways we identified, especially TGF- $\beta$ , might be exploited for breast cancer prevention because their modulation could deplete p27<sup>+</sup> progenitors and decrease breast cancer risk. Analysis of large cohorts with detailed risk factor data and long-term follow-up would be required to conclusively determine the relationships between the frequency of these p27<sup>+</sup> cells, the activity of parity-related signaling pathways, and 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. Each collaborator had their own protocol at their institution, and we had one in DFCI for using these samples. For organ cultures, thin (~1 mm) slices of epithelium-enriched breast tissue were cultured for 8 days in 6-well plates with coculture inserts in M87A medium (Garbe et al., 2009). Detailed protocols for cell purification and the generation of SAGE-seq, MSDK-seq, and ChIP-seq libraries are posted at <http://polyaklab.dfci.harvard.edu/>. Genomic data were analyzed as described before (Kowalczyk et al., 2011; Maruyama et al., 2011; Wu et al., 2010). Semiquantitative and qRT-PCR and qMSP analyses were performed on cells purified from 15 to 20 samples of nulliparous and parous breast tissue as previously reported (Hu et al., 2005). Details are included in the Supplemental Information.

### FACS, Immunofluorescence, and Immunohistochemical Analyses

Single-cell suspension of human breast epithelial cells was obtained essentially as described by Shipitsin et al. (2007). Cells were stained with propidium iodide, PE/Cy7-CD10 (BioLegend; Clone H110a), APC-CD24 (BioLegend; clone ML5), and Zenon Alexa 405-labeled CD44 (BD; Clone 515). Immunohistochemical and immunofluorescence analyses were performed essentially as described by Shipitsin et al. (2007); detailed protocols are included in Supplemental Information.

### ACCESSION NUMBERS

The GEO accession number for the data reported in this paper is GSE32017.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.stem.2013.05.004>.

## ACKNOWLEDGMENTS

We thank Lisa Cameron in the DFCI Confocal and Light Microscopy Core Facility, members of Dr. Massimo Loda's lab for technical assistance, members of our laboratories and Drs. Elgene Lim and David Livingston for their critical reading of this manuscript, Jonathan Yingling (Eli Lilly) for providing the LY2109761 TGF- $\beta$  receptor kinase inhibitor, Drs. Sally Knox, Jeffrey Lamont, and Dao Tuoc (Baylor University Medical Center-Baylor Sammons Cancer Center), Erin Bowlby (University of California San Francisco, San Francisco), and Drs. Eli Golomb and Pikarski (Hadassah Medical Centre) for their help with collecting tissue samples from patients with BRCA1/BRCA2 germline mutation. 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. This work was supported by the Avon Foundation (to K.P. and S.S.), the National Cancer Institute P50 CA89383 and P01 CA080111 (to S.J.S., K.P., and A.L.R.), CA116235-04S1 (to K.P.), and CA087969 (to R.M.T.), the Susan G. Komen Foundation (to R.M., Y.S., I.H., J.E.G., and K.P.), the Terri Brodeur Foundation (to S.C.), US Army Congressionally Directed Research W81XWH-07-1-0294 (to K.P.), the Victorian Breast Cancer Research Consortium (to S.J.L., G.C., and E.W.T.), the St. Vincent's Hospital Research Endowment Fund and the Victorian Government's OIS Program (to E.W.T.), the Programme for Advanced Medical Education funded by Fundação Calouste Gulbenkian (to F.M.), and the Cellex Foundation (to V.A.).

Received: April 30, 2012

Revised: February 11, 2013

Accepted: May 9, 2013

Published: July 3, 2013

## 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., and Visvader, J.E. (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-Lévy, I., Zeleniuch-Jacquotte, A., Russo, J., Russo, I.H., Bordás, 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.
- Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al.; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.
- 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.
- Bloustain-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. USA* **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.
- Briskin, 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. (2012). Breast epithelial cell proliferation is markedly increased with short-term high levels of endogenous estrogen secondary to controlled ovarian hyperstimulation. *Breast Cancer Res. Treat.* **132**, 653–660.
- 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.; Kathleen Cunningham Foundation Consortium for Research into Familial Breast Cancer; Breast Cancer Association Consortium. (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.
- 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.
- 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., and Chodosh, L.A. (2002). Persistent parity-induced changes in growth factors, TGF- $\beta$ 3, and differentiation in the rodent mammary gland. *Mol. Endocrinol.* **16**, 2034–2051.
- 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.
- Haakensen, V.D., Bjørø, T., Lüders, T., Riis, M., Bukholm, I.K., Kristensen, V.N., Troester, M.A., Homen, M.M., Ursin, G., Børresen-Dale, A.L., and Helland, Å. (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., Lüders, 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 Med. Genomics* **4**, 77.

- Hu, M., Yao, J., Cai, L., Bachman, K.E., van den Brûle, 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 tumorigenesis 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.; Endogenous Hormones and Breast Cancer Collaborative Group. (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.
- 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.
- 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.
- 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 Genet.* *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. (2012). Altered anti-sense-to-sense transcript ratios in breast cancer. *Proc. Natl. Acad. Sci. USA* *109*, 2820–2824.
- Maxwell, K.N., and Domchek, S.M. (2012). Cancer treatment according to BRCA1 and BRCA2 mutations. *Nat. Rev. Clin. Oncol.* *9*, 520–528.
- Mitsuhashi, T., Aoki, Y., Eksioğlu, 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. USA* *98*, 6435–6440.
- Molyneux, G., Geyer, F.C., Magnay, F.A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* *7*, 403–417.
- 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., Sviridov, E., Yao, J., Dosymbekov, D., Ustyansky, V., Kaznacheev, V., Dezso, Z., Mulvey, L., Macconail, 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.
- 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.
- 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.
- Ohtsubo, M., Yasunaga, S., Ohno, Y., Tsumura, M., Okada, S., Ishikawa, N., Shirao, K., Kikuchi, A., Nishitani, H., Kobayashi, M., and Takihara, Y. (2008). Polycomb-group complex 1 acts as an E3 ubiquitin ligase for Geminin to sustain hematopoietic stem cell activity. *Proc. Natl. Acad. Sci. USA* *105*, 10396–10401.
- 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.
- Poynter, J.N., Langholz, B., Largent, J., Mellemejaer, L., Bernstein, L., Malone, K.E., Lynch, C.F., Borg, A., Concannon, P., Teraoka, S.N., et al.; WECARE Study Collaborative Group. (2010). Reproductive factors and risk of contralateral breast cancer by BRCA1 and BRCA2 mutation status: results from the WECARE study. *Cancer Causes Control* *21*, 839–846.
- Rebbeck, T.R., Kantoff, P.W., Krithivas, K., Neuhausen, S., Blackwood, M.A., Godwin, A.K., Daly, M.B., Narod, S.A., Garber, J.E., Lynch, H.T., et al. (1999). Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat. *Am. J. Hum. Genet.* *64*, 1371–1377.
- Roy, R., Chun, J., and Powell, S.N. (2012). BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* *12*, 68–78.
- 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., 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., 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. (2012). Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. *Int. J. Cancer* *131*, 1059–1070.
- 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.
- Shinde, S.S., Forman, M.R., Kuerer, H.M., Yan, K., Peintinger, F., Hunt, K.K., Hortobagyi, G.N., Pusztai, L., and Symmans, W.F. (2010). Higher parity and shorter breastfeeding duration: association with triple-negative phenotype of breast cancer. *Cancer* *116*, 4933–4943.
- 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.
- Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., et al.; Oslo Breast Cancer Consortium (OSBREAC). (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* *486*, 400–404.
- 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.
- Visvader, J.E. (2011). Cells of origin in cancer. *Nature* *469*, 314–322.
- 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.