nature medicine

Aberrant luminal progenitors as the candidate target population for basal tumor development in *BRCA1* mutation carriers

Elgene Lim[1,2,9], François Vaillant[1,9], Di Wu[1,2], Natasha C Forrest[1], Bhupinder Pal[1], Adam H Hart[3], Marie-Liesse Asselin-Labat[1], David E Gyorki[1,2], Teresa Ward[1], Audrey Partanen[4], Frank Feleppa[4], Lily I Huschtscha[5], Heather J Thorne[6], kConFab[7], Stephen B Fox[6], Max Yan[6], Juliet D French[8], Melissa A Brown[8], Gordon K Smyth[1], Jane E Visvader[1,9] & Geoffrey J Lindeman[1,2,4,9]

Basal-like breast cancers arising in women carrying mutations in the *BRCA1* gene, encoding the tumor suppressor protein BRCA1, are thought to develop from the mammary stem cell. To explore early cellular changes that occur in *BRCA1* mutation carriers, we have prospectively isolated distinct epithelial subpopulations from normal mammary tissue and preneoplastic specimens from individuals heterozygous for a *BRCA1* mutation. We describe three epithelial subsets including basal stem/progenitor, luminal progenitor and mature luminal cells. Unexpectedly, we found that breast tissue from *BRCA1* mutation carriers harbors an expanded luminal progenitor population that shows factor-independent growth *in vitro*. Moreover, gene expression profiling revealed that breast tissue heterozygous for a *BRCA1* mutation and basal breast tumors were more similar to normal luminal progenitor cells than any other subset, including the stem cell–enriched population. The c-KIT tyrosine kinase receptor (encoded by *KIT*) emerged as a key marker of luminal progenitor cells and was more highly expressed in *BRCA1*-associated preneoplastic tissue and tumors. Our findings suggest that an aberrant luminal progenitor population is a target for transformation in *BRCA1*-associated basal tumors.

Germline mutations in the breast cancer susceptibility gene *BRCA1* are associated with a substantially higher risk of developing basal-like breast cancer[1]. Basal tumors belong to one of the six major breast tumor subtypes identified on the basis of global gene expression analyses and are among the most aggressive[2,3]. The breast stem cell has been suggested to be the 'cell of origin' for tumors arising in *BRCA1* mutation carriers[4]. Whereas multiple functions for BRCA1 have been described, such as in the DNA damage response, X-chromosome inactivation and transcriptional and cell cycle control[5–7], the precise mechanisms by which *BRCA1* mutations lead to tumorigenesis remain to be elucidated. Within the mouse mammary gland, targeted disruption of *Brca1* leads to developmental defects[8,9], and, in mammary epithelial cells, BRCA1 has been implicated in the control of differentiation[10–12].

Dissection of the normal epithelial subtypes that constitute the human mammary hierarchy is a major prerequisite to understanding which cell types are targeted in *BRCA1*-associated and other breast cancers. Human breast tissue is organized as a branching network of ducts and lobules that comprise epithelial cells in the luminal and myoepithelial lineages embedded in a stromal matrix. The distinct phenotypes that arise during culture of primary human mammary cells[13–18] provide evidence for a hierarchical organization of epithelial cells in human breast. Recent *in vivo* studies have shown that human mammary epithelial cells with high aldehyde dehydrogenase-1 (ALDH-1) activity have stem/progenitor cell properties[19] and that transplantation under the renal capsule provides a method for quantifying human mammary stem cells[20]. In this study, we have prospectively isolated discrete epithelial subpopulations that reside within breast tissue from normal individuals and *BRCA1* mutation carriers. The discovery of an aberrant luminal progenitor population in preneoplastic tissue from *BRCA1* mutation carriers, together with molecular profiling analyses, implicate luminal progenitor cells as a probable target population in *BRCA1*-associated and possibly other basal breast tumors.

RESULTS

Identification of three distinct human breast epithelial subsets

To delineate epithelial subpopulations in human mammary tissue, we used an analogous approach to that previously described for the mouse mammary stem cell (MaSC)[21,22]. We depleted hematopoietic and endothelial cells from freshly isolated cell suspensions derived

[1]The Walter and Eliza Hall Institute of Medical Research (WEHI), Parkville, Victoria, Australia. [2]The University of Melbourne, Parkville, Victoria, Australia. [3]Department of Anatomy and Developmental Biology, Monash University, Melbourne, Victoria, Australia. [4]The Royal Melbourne Hospital, Parkville, Victoria, Australia. [5]Children's Medical Research Institute, Westmead, New South Wales, Australia. [6]Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. [7]The Kathleen Cuningham Consortium for Research into Familial Breast Cancer. [8]The School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland, Australia. [9]These authors contributed equally to this work. Correspondence should be addressed to J.E.V. (visvader@wehi.edu.au) or G.J.L. (lindeman@wehi.edu.au).

Received 15 April; accepted 11 June; published online 2 August 2009; doi:10.1038/nm.2000

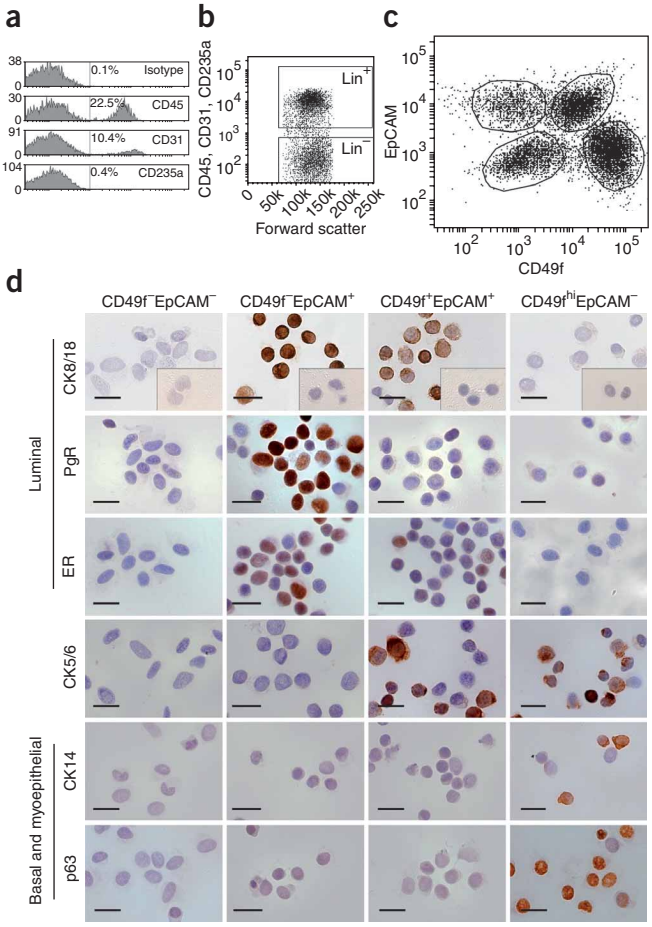

Figure 1 CD49f and EpCAM define distinct subpopulations in the human mammary epithelium. (**a**) Expression analysis of lineage markers CD45 (hematopoietic cells), CD31 (endothelial cells) and CD235a (erythrocyte precursors) in human mammary tissue on live, single-cell gated populations. Percentage of positively labeled cells for each marker, compared to an isotype control antibody, are shown. (**b**) FACS analysis of CD45, CD31 and CD235a expression. Cells negative for all three markers defined the Lineage-negative (Lin⁻) population. (**c**) Expression of CD49f and EpCAM in the Lin⁻ population of a 27-year-old woman. (**d**) Immunohistochemical analysis of cells isolated from normal reduction mammoplasties. Cells were sorted at purities >90% and cytospun for immunohistochemical staining: CK8/18, cytokeratin-8 and cytokeratin-18; PgR, progesterone receptor; ER, estrogen receptor; CK5/6, cytokeratin-5 and cytokeratin-6; CK14, cytokeratin-14; p63. Inset, isotype control. Scale bars, 20 μm.

from reduction mammoplasties by FACS (**Fig. 1a**). We fractionated the resultant lineage-negative (Lin⁻) population (**Fig. 1b**) into four distinct subpopulations using antibodies against CD49f (α_6 integrin) and epithelial cell adhesion molecule (EpCAM; also referred to as CD326 and epithelial-specific antigen), paralleling previous results[18,20] (**Fig. 1c**). CD140b (platelet-derived growth factor receptor-β), a marker of fibroblasts, was specifically expressed by the CD49f⁻EpCAM⁻ subpopulation (**Supplementary Fig. 1**). Only the CD49f⁺EpCAM⁺ subpopulation contained a high proportion of cells positive for CD133 (prominin-1), a progenitor marker in diverse tissues, and CD24, which is expressed on mouse mammary epithelial cells[21,22] (**Supplementary Fig. 1**).

Three epithelial subsets were revealed through immunohistochemical staining of freshly cytospun cells from the subdivided Lin⁻ populations (at >90% purity). EpCAM was predominantly expressed on luminal cells, whereas high CD49f expression marked basal cells. (**Fig. 1d** and **Table 1**). The CD49f^{hi}EpCAM⁻ subpopulation expressed the basal lineage markers p63, cytokeratin-14 and vimentin but did not express the estrogen receptor or progesterone receptor. In contrast, the CD49f⁻EpCAM⁺ and CD49f⁺EpCAM⁺ subsets expressed luminal lineage markers including cytokeratin-8 and cytokeratin-18 (hereafter referred to as cytokeratin-8/18 because the antibody used recognizes both proteins), cytokeratin-19, GATA-3 (a regulator of luminal cell differentiation) and Mucin-1 (MUC-1) (**Fig. 1d** and **Table 1**). The highest proportion of estrogen receptor– and progesterone receptor–expressing cells was observed in the CD49f⁻EpCAM⁺ subpopulation, indicating that it was enriched for mature luminal cells. Notably, cytokeratin-5/6 was expressed in both the basal and luminal subpopulations (**Fig. 1d** and **Table 1**). The observation of cytokeratin-5/6 expression on luminal epithelial cells is consistent with previous findings by others on whole breast tissue samples[23,24]. Thus, cytokeratin-5/6 is not a sole marker of the basal lineage, a finding with implications for the use of this marker in 'typing' breast tumors. The CD49f⁻EpCAM⁻ cell population comprised stromal fibroblasts that expressed the highest levels of ALDH-1, in contrast to the epithelial populations (**Table 1**).

To determine which epithelial subpopulation was enriched for mammary repopulating activity, we transplanted cells in numbers approximating their frequency in the Lin⁻ population into cleared (de-epithelialized) mammary fat pads of immunocompromised nonobese diabetic–severe combined immunodeficient (NOD-SCID) interleukin-2 receptor-γ (*Il2rg*)-null mice. To optimize the engraftment of donor human mammary cells, we co-injected sorted epithelial cells with telomerase reverse transcriptase–immortalized human mammary fibroblasts, and we implanted a subcutaneous estradiol pellet into recipient mice (adapted from methods used in previous studies[25]). Only the CD49f^{hi}EpCAM⁻ subset showed mammary regenerating capacity, at a frequency of 1 in 21,500 (95% confidence interval, 1 in 38,000 to 1 in 12,000; **Supplementary Table 1** and **Fig. 2a**), and is hereon referred to as the MaSC-enriched population. The low repopulating frequency

Table 1 Immunohistochemical analysis of Lin⁻ populations defined by CD49f and EpCAM

	CD49f⁻ EpCAM⁻	CD49f⁻ EpCAM⁺	CD49f⁺ EpCAM⁺	CD49f^{hi} EpCAM⁻
Vimentin	96.9 ± 1.3	1.4 ± 0.9	11.8 ± 3.8	77.8 ± 5.2
Cytokeratin-8/18	3.9 ± 1.6	98.8 ± 0.2	91.2 ± 2.3	2.3 ± 0.9
Cytokeratin-19	1.9 ± 0.6	80.3 ± 15.4	61.6 ± 9.3	4.0 ± 2.8
Estrogen receptor	0.3 ± 0.3	55.4 ± 11.1	28.0 ± 18.1	0.2 ± 0.2
Progesterone receptor	1.8 ± 0.6	71.0 ± 8.2	2.3 ± 0.3	0.0
MUC-1	1.0 ± 0.6	99.8 ± 0.2	80.3 ± 13.1	1.9 ± 0.9
GATA-3	1.2 ± 0.3	57.3 ± 3.0	35.4 ± 4.2	8.0 ± 3.0
Cytokeratin-14	0.6 ± 0.04	0.3 ± 0.04	6.3 ± 4.1	59.8 ± 10.4
p63	1.6 ± 0.8	0.0	0.1 ± 0.0	76.4 ± 9.0
Cytokeratin-5/6	1.2 ± 0.6	2.8 ± 1.2	49.9 ± 11.6	52.4 ± 11.0
HER2	0.0	0.0	0.0	0.0
EGFR	22.0 ± 0.4	6.2 ± 0.3	45.5 ± 19.6	11.7 ± 15.1
Nestin	76.3 ± 9.2	0.7 ± 0.7	11.1 ± 1.6	89.5 ± 4.0
ALDH-1	20.9 ± 7.8	0.7 ± 0.3	0.4 ± 0.1	0.4 ± 0.1

The mean percentage ± s.e.m. of cells that stained positive for the indicated antibodies is shown. A minimum of three independent normal breast reduction mammoplasty samples was evaluated for each marker. For HER2 staining, breast tumors showing HER2 amplification served as positive controls.

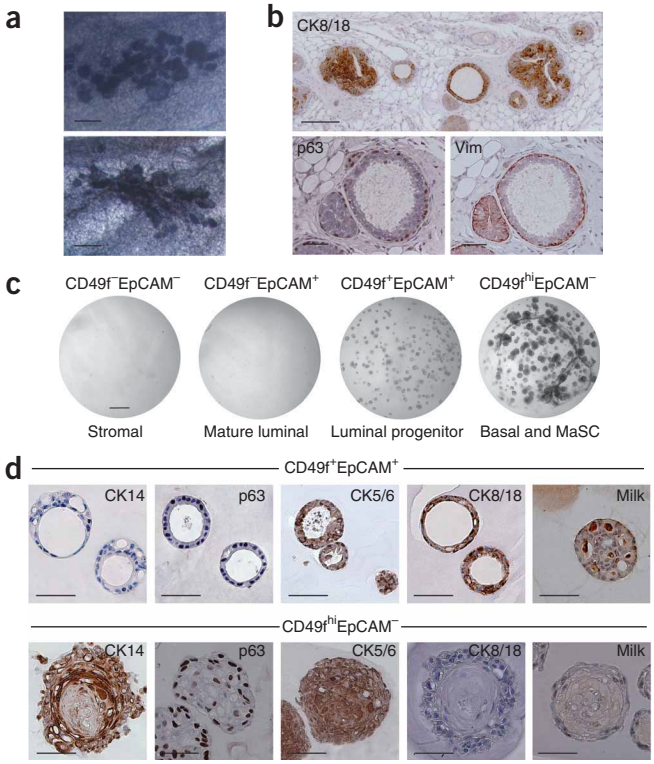

Figure 2 CD49f^{hi}EpCAM^{-} cells have *in vivo* repopulating capacity. (**a**) Outgrowths derived from transplantation of CD49f^{hi}EpCAM^{-} cells (top, 25,000 cells from a 53-year-old woman; bottom, 45,000 cells from a 35-year-old woman). Scale bars, 250 μm. (**b**) Immunostaining of sections of outgrowths with antibodies to cytokeratin-8/18, vimentin and p63. Antibodies to cytokeratin-8/18 and vimentin specifically recognize the human antigens. The top image is a composite of two contiguous fields. Scale bars, 75 μm (cytokeratin-8/18) and 60 μm (p63, Vim). (**c**) Differential *in vitro* growth characteristics of specific epithelial subsets. Cells (1,000) from each of the populations defined in **Figure 1c** were cultured in Matrigel for 14 d. Data are representative of at least 30 independent experiments. Scale bar, 0.5 mm. (**d**) Immunohistological staining of sections of structures arising from the CD49f^{+} populations (top, CD49f^{+}EpCAM^{+}; bottom, CD49f^{hi}EpCAM^{-}) with antibodies to cytokeratin-14, p63, cytokeratin-5/6, cytokeratin-8/18. Far right, milk production in structures after prolactin treatment, detected with a milk-specific antibody. Scale bars, 50 μm.

probably reflects the technical challenges imposed by this orthotopic xenotransplantation assay. We found that inclusion of Matrigel further enhanced the regenerative capacity by three- to fourfold (data not shown). The outgrowths comprised ducts and distinctive lobular-like structures reminiscent of the terminal ductal lobular units that constitute human mammary tissue, and we found them to be donor derived by immunostaining with mouse antibodies specific for human cytokeratin-8/18 and vimentin (**Fig. 2b** and **Supplementary Fig. 2**). Some primary outgrowths derived from the CD49f^{hi}EpCAM^{-} subset showed self renewal, in agreement with a recent report[20] (**Supplementary Fig. 3**).

Two subpopulations show clonogenic activity *in vitro*

To further characterize the epithelial subpopulations, we evaluated progenitor activity with colony-forming cell assays in Matrigel[21,26]. Specifically, we cultured the cells in Matrigel in serum-free medium comprising epidermal growth factor, insulin, hydrocortisone and B27 supplement[27]. We selected B27 supplement because it has been shown to support the growth of mammospheres from human breast tissue[14]. We found B27 supplement to be necessary for colony-forming ability, as few colonies (less than 10%) arose in its absence (**Figs. 2** and **3**). Progenitor activity was only evident in the CD49f^{+}EpCAM^{+} and CD49f^{hi}EpCAM^{-} epithelial subpopulations (**Fig. 2c**). CD49f^{+}EpCAM^{+} cells gave rise to homogeneous colonies that expressed the luminal markers cytokeratin-8/18 and cytokeratin-5/6 (over 60%) but not basal markers (p63 or cytokeratin-14) (**Fig. 2d**) and were capable of undergoing differentiation into milk-producing cells (37.5%) when exposed to a lactogenic stimulus (**Fig. 2d**). These cells therefore fulfill the definition of committed luminal progenitor cells. Conversely, structures derived from the basal CD49f^{hi}EpCAM^{-} population were more complex, comprising ductal-like structures as well as pleomorphic, dense colonies, the majority of which expressed p63, cytokeratin-5/6 and cytokeratin-14 (**Fig. 2c,d** and data not shown). In the presence of a lactogenic stimulus, 5.4% of colonies underwent alveolar differentiation. These findings are compatible with the presence of stem cells and/or bipotent progenitor cells in this subset and parallel observations for the mouse MaSC-enriched population[21,22].

Aberrant progenitors in *BRCA1*-mutant preneoplastic tissue

Because carriers of pathogenic mutations in the *BRCA1* gene are predisposed to the development of basal breast tumors, we evaluated the proportion and *in vitro* growth properties of the various epithelial subpopulations in histologically normal mammary tissue obtained from *BRCA1* mutation carriers (under age 50 years) undergoing prophylactic mastectomy (**Supplementary Table 2**). Analysis of the three epithelial subsets in samples heterozygous for a *BRCA1* mutation (*BRCA1*-mutant samples) ($n = 10$) revealed a substantial reduction in the MaSC-enriched (basal) subset (CD49f^{hi}EpCAM^{-}) ($P < 0.001$) but an increase in the luminal progenitor cell fraction (CD49f^{+}EpCAM^{+}) ($P < 0.05$) relative to age-matched normal breast tissue ($n = 30$, **Fig. 3a,b**). We next assessed the *in vitro* growth properties of *BRCA1*-mutant mammary epithelial cells in Matrigel (**Fig. 3c**). The luminal progenitor population (CD49f^{+}EpCAM^{+}) from *BRCA1* mutation carriers ($n = 11$) showed markedly higher clonogenic activity compared to that from noncarriers in the presence of B27 ($n = 20$, $P < 0.001$, **Fig. 3d**). Immunostaining of sections revealed little difference in the expression of lineage markers on colonies and structures derived from normal versus *BRCA1*-mutant mammary epithelial cells (**Supplementary Fig. 4**). However, when we assessed progenitor activity in the absence of B27 supplement, we observed B27 factor–independent growth of luminal progenitor cells in virtually every *BRCA1*-mutant sample evaluated, a noteworthy difference in the growth properties of the luminal progenitor population (14 of 15 samples, $P < 0.001$; **Supplementary Table 2**).

Notably, the luminal progenitor but not the MaSC-enriched subset showed clonogenic activity in the absence of B27 supplement (**Fig. 3c,d**). Analysis of *BRCA1* messenger RNA expression amongst the various subpopulations by quantitative RT-PCR revealed substantially higher levels in the two luminal subsets than in the MaSC-enriched population, consistent with a role for BRCA1 in luminal progenitor cells (**Fig. 3e**). We also observed B27-independent progenitor growth in one out of six prophylactic mastectomy samples from *BRCA2* mutation carriers (data not shown) and in three out of 20 reduction mammoplasties, for which minimal family history was available (**Fig. 3d**). The assay may therefore also represent a readout for perturbed luminal progenitor cells in other preneoplastic tissue in addition to that from *BRCA1* mutation carriers.

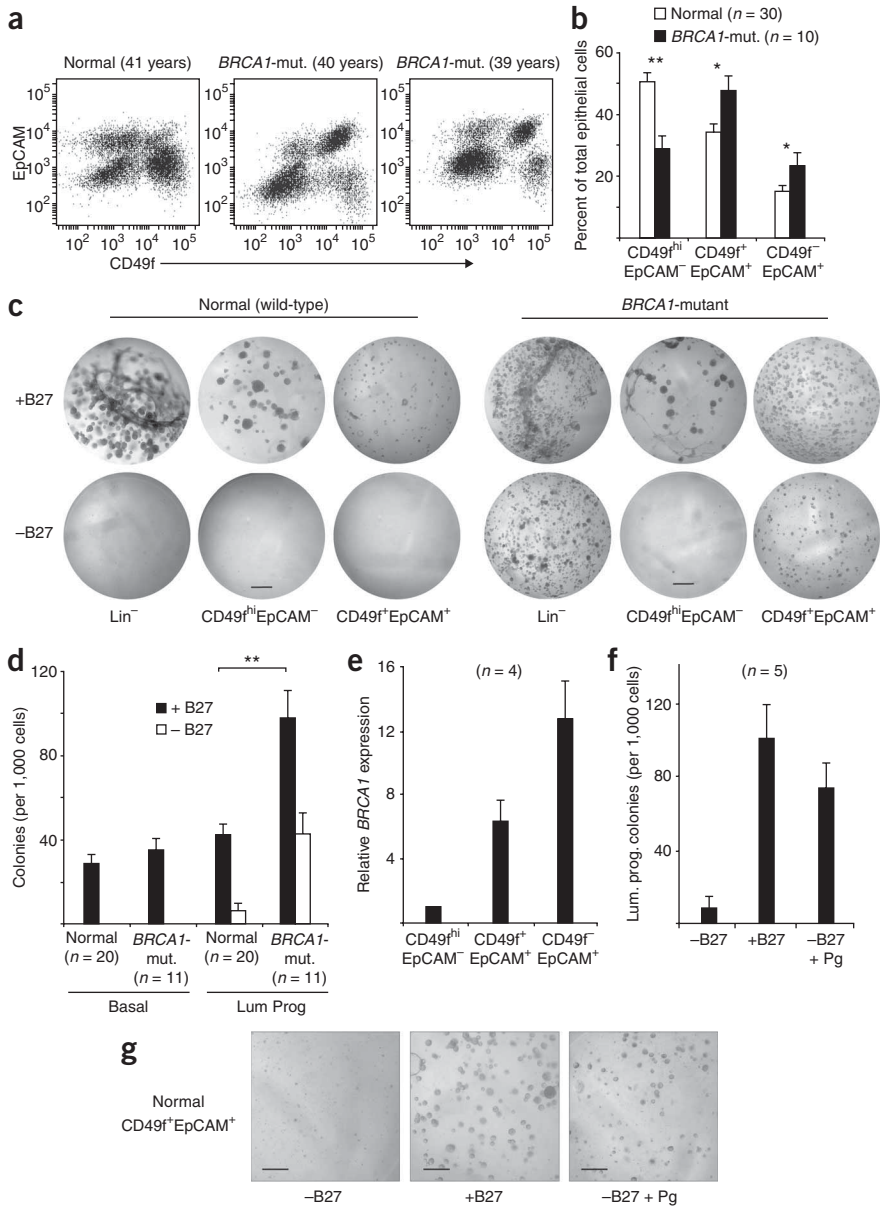

Figure 3 Luminal progenitor cells from *BRCA1* mutation carriers show factor-independent growth *in vitro*. (**a**) Representative FACS dot plots showing the expression of CD49f and EpCAM from age-matched normal and *BRCA1*-mutant breast tissues. An identical number of events are shown in each plot. (**b**) Bar chart depicting the relative proportion of epithelial cell subpopulations (CD49f^{hi}EpCAM^{-}, CD49f^{+}EpCAM^{+} or CD49f $^{-}$EpCAM^{+}) in normal ($n = 30$) and *BRCA1*-mutant ($n = 10$) epithelia. Values are expressed as a percentage of the total number of epithelial cells in the CD49f^{hi}EpCAM^{-}, CD49f^{+}EpCAM^{+} plus CD49f^{-}EpCAM^{+} subpopulations. (**c**) Sorted Lin^{-} cells and CD49f^{hi}EpCAM^{-} or CD49f^{+}EpCAM^{+} subpopulations from normal and *BRCA1*-mutant breast tissue were cultured for 14 d in medium, with (top) or without (bottom) B27 supplement. Data are from 31 independent experiments, each performed in duplicate or triplicate. Scale bars, 0.5 mm. (**d**) Bar chart showing the colony forming ability of basal and luminal progenitor cells from normal ($n = 20$) or *BRCA1*-mutant ($n = 11$) mammary tissue cultured in either the absence or the presence of B27. (**e**) Expression analysis of *BRCA1* in the MaSC-enriched (CD49f^{hi}EpCAM^{-}) and luminal subpopulations (CD49f^{+}EpCAM^{+}, CD49f^{-}EpCAM^{+}) from normal breast tissue by quantitative RT-PCR relative to *18S* rRNA ($n = 4$). (**f**) Bar chart depicting the colony-forming ability of CD49f^{+}EpCAM^{+} cells from normal breast tissue in the absence or presence of B27 supplement, or in the presence of progesterone (Pg), with cultures performed in triplicate ($n = 5$). (**g**) Representative images of cultures from **f**. Scale bars, 250 μm. Data represent means ± s.e.m. $^{*}P < 0.05$, $^{**}P < 0.001$.

To examine whether the altered growth properties observed for human mammary cells reflected an evolutionarily conserved process, we analyzed the luminal progenitor cells[26] from the mammary glands of *Brca1*-deficient mice using a Matrigel assay[21] in which we replaced serum with B27 supplement. Indeed, we found B27 supplement to be necessary for the colony-forming ability of normal luminal progenitor (CD29^{lo}CD24^{+}CD61^{+}) and MaSC-enriched (CD29^{hi}CD24^{+}) cells from wild-type C57BL/6 mammary glands in serum-free medium (**Supplementary Fig. 5**). Similar to their human counterparts, luminal progenitor cells isolated from mouse mammary tumor virus (MMTV)-Cre *Brca1*$^{f/f}$ glands (where conditional deletion results in *Brca1* deficiency) but not wild-type virgin glands showed B27 factor independence ($n = 6$; **Supplementary Fig. 5**).

No apparent ligand-independent progesterone receptor activation

The B27 factor–independent growth of the luminal progenitor subpopulation in preneoplastic mammary tissue from women harboring a *BRCA1* mutation and *Brca1*-deficient mice necessarily implies that these cells have established independence from exogenous factors present in B27 supplement. One of these components, progesterone, was of particular interest, as BRCA1 can regulate progesterone receptor signaling[28], and *Brca1*- and *Tp53*-deficient mice have been shown to overexpress their cognate progesterone receptor[29]. Indeed, addition of progesterone in lieu of B27 supplement to normal luminal progenitor cell cultures led to a substantial rescue of colony-forming cell activity ($n = 5$, **Fig. 3f,g**), whereas other factors, including basic fibroblast growth factor, pituitary extract and transferrin, did not (data not shown).

However, the clonogenic activity of luminal progenitor cells from *BRCA1* mutation carriers was not inhibited by the progesterone receptor antagonists mifepristone (RU486) or onapristone ($n = 2$; data not shown), despite the trend toward higher expression of progesterone receptor in luminal progenitor cells heterozygous for a *BRCA1* mutation (**Supplementary Fig. 6**). These findings suggest that the B27-independent growth of *BRCA1*-mutant luminal progenitor cells is not due to ligand-independent activation of progesterone receptor, as described for mice deficient in both *Brca1* and *Tp53*[29]. Rather, it is likely that an intrinsic defect exists in one or more signaling pathways in *BRCA1*-deficient luminal progenitor cells, allowing them to bypass the requirement for exogenous factors.

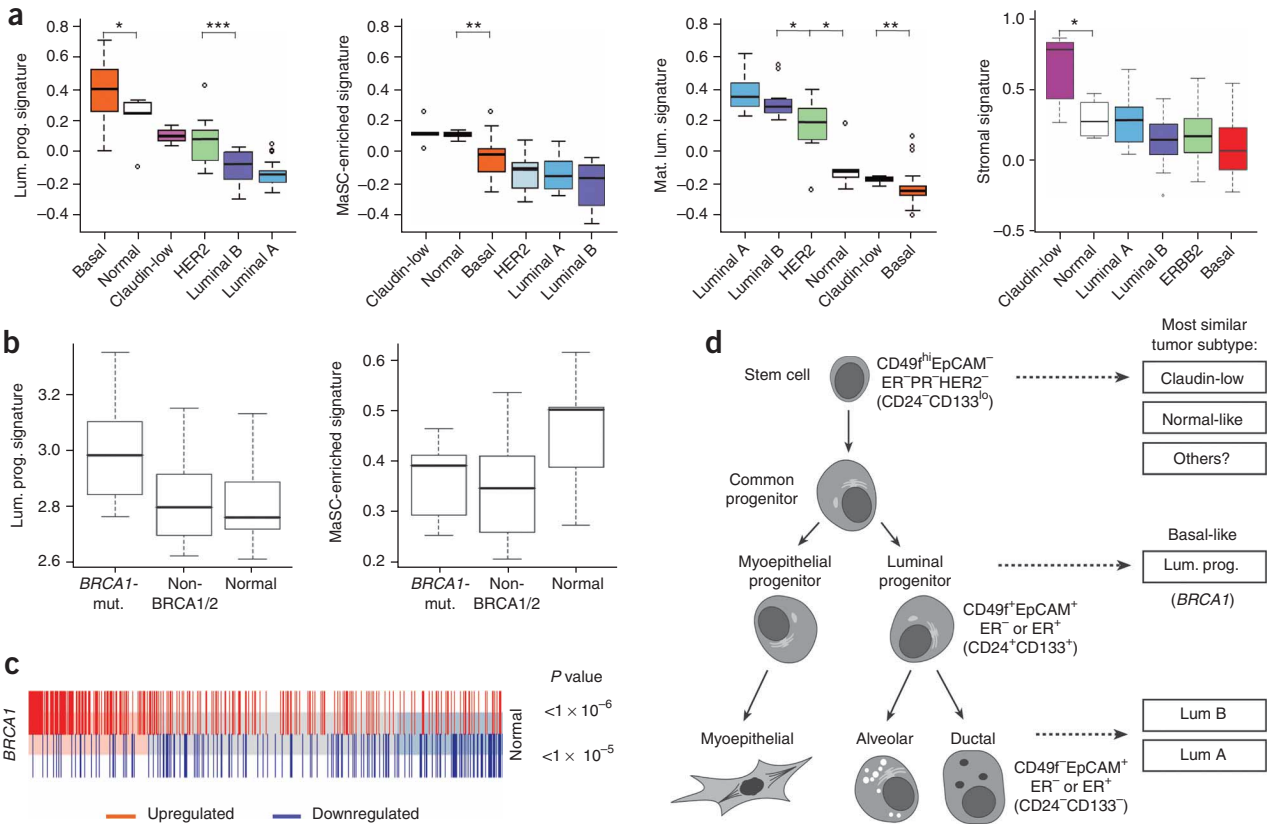

Figure 4 Comparison of gene expression profiles of normal human mammary epithelial and stromal subsets with the major subtypes of breast cancer and with preneoplastic tissue from *BRCA1* mutation carriers. (**a**) Box plots of signature expression scores by tumor subtype for each subset. The luminal progenitor signature scores are highest in the basal subtype of breast cancer. $*P < 0.05$, $**P < 0.01$, $***P = 0.001$. HER2, tyrosine kinase receptor HER2, encoded by *ERBB2*. (**b**) Box plots of signature expression scores by *BRCA1* mutation status for each epithelial subset. (*BRCA1*-mutant, $n = 7$; non-BRCA1/2, $n = 8$; normal, $n = 5$). (**c**) Barcode plot showing the ability of the luminal progenitor signature genes to distinguish between *BRCA1* mutation carriers and other individuals, with corresponding mean-rank gene set test *P* values. Red bars designate upregulated genes in the luminal progenitor signature set ($P < 1 \times 10^{-6}$); blue bars designate downregulated genes ($P < 1 \times 10^{-5}$). (**d**) Schematic model of the human breast epithelial hierarchy and possible relationships with breast cancer subtypes. Subpopulations containing MaSCs, luminal progenitor cells and mature luminal cells are defined by differential expression of CD49f and EpCAM. Gene expression profiling of these subpopulations revealed similarities to specific subtypes of breast cancer, depicted on the right hand side. The luminal progenitor population shares greatest similarity to the basal-like subtype, redefined here as luminal progenitor (Lum. Prog.) subtype. We speculate that the HER2 subtype arises through amplification of the *ERBB2* amplicon in a target cell committed to the luminal lineage. ER, estrogen receptor; PR, progesterone receptor.

Comparison of mammary gene signatures with cancer subtypes

To gain insight into the molecular characteristics of luminal progenitor cells and relationships between normal mammary epithelial cells and breast cancer subtypes, we used microarray profiling to derive gene expression signatures representative of the MaSC-enriched, luminal progenitor, mature luminal epithelial and stromal populations using freshly sorted cells (>90% purity) from normal breast tissue. We found the four mammary cell subpopulations to have distinct gene expression profiles, as shown by hierarchical clustering and multidimension scaling plot analyses (**Supplementary Fig. 7**). We found genes characteristic of each population by subtraction of genes common to the other subsets, including upregulated and downregulated genes (**Supplementary Tables 3** and **4**). We then used these signature gene sets (**Supplementary Tables 5–8**) to interrogate the expression profiles of the six distinct molecular subtypes of human breast cancer described thus far[2]. We used gene set testing and expression signatures, as they represent the best approach for establishing concordances between expression profiles across different cell types and array platforms.

First, we considered the rankings of each population's signature genes in pairwise comparisons between the tumor subtypes. Barcode plots and mean-rank gene set enrichment tests[30] gave a clear ordering of the tumor subtypes for each epithelial cell population. The luminal progenitor signature genes were more associated with basal cancers than the 'normal-like' or claudin-low subtypes (**Supplementary Fig. 8**). The upregulated luminal progenitor signature genes tended to be more highly expressed in basal tumors than in other tumor subtypes ($P = 0.025$ versus normal-like, $P < 1 \times 10^{-6}$ versus other subtypes), whereas the downregulated luminal progenitor signature genes were least expressed in the basal tumors ($P < 1 \times 10^{-4}$ versus claudin-low, $P < 1 \times 10^{-6}$ versus normal-like and other subtypes) (**Supplementary Fig. 8**). Conversely, the MaSC-enriched (basal) signature genes were generally more concordant with claudin-low and 'normal-like' than the basal subtype. Upregulated MaSC-enriched signatures genes were higher in the normal-like and claudin-low subtypes than the basal group ($P < 1 \times 10^{-6}$ and $P = 0.006$, respectively), but the downregulated MaSC-enriched signature genes seemed lowest in the claudin-low subtype, followed by basal, then

'normal-like' subtypes (data not shown). The MaSC-enriched signature is reflective of a heterogeneous population of basal cells, including stem cells, myoepithelial cells and, probably, basal progenitor cells. As anticipated, the mature luminal signature genes were concordant with the luminal A and B subtypes ($P < 1 \times 10^{-6}$) (data not shown).

Second, we computed overall signature expression scores for each cell population in each tumor sample (**Fig. 4a**). The signature scores correlate expression in each tumor sample with that of the cell population, using the signature genes for that population. The same pattern was evident as for the gene set tests. That is, the luminal progenitor signature was highest in the basal subtype of breast cancer ($P = 0.028$ for basal versus normal-like), whereas the MaSC-enriched signature was highest in the claudin-low and normal-like subtypes ($P = 0.003$ for normal-like versus basal) (**Fig. 4a**). Similarly, the mature luminal epithelial signature was most evident in the luminal A and B subtypes (**Fig. 4a**). Notably, the stromal signature score was highest for the claudin-low subtype (**Fig. 4a**).

Correlation between luminal progenitor and *BRCA1* signatures

We next compared the luminal progenitor cell signature with the expression profiles of pathologically normal preneoplastic tissue from *BRCA1* mutation carriers and normal breast tissue, as well as individuals with a strong family history of breast cancer where no *BRCA1* or *BRCA2* mutation has been identified ('non-BRCA1/2', Online Methods). The luminal progenitor expression signature score was highest in prophylactic mastectomy tissue from *BRCA1* mutation carriers compared to normal breast tissue or tissue from non-BRCA1/2 individuals (**Fig. 4b**). Gene set enrichment tests confirmed that the luminal progenitor signature genes were highly ranked in differential expression comparisons between the *BRCA1* mutation–associated tissues and the other two groups (**Fig. 4c** and data not shown). The marked correlation between the molecular signatures of luminal progenitor cells, *BRCA1* mutation–associated breast tissue and basal tumors suggests that the primary cellular manifestation in *BRCA1*-associated and other basal cancers is the luminal progenitor cell rather than the MaSC (**Fig. 4d**).

Higher c-KIT levels in *BRCA1*-mutant breast tissue and tumors

The luminal progenitor gene signature contains a number of highly expressed genes within 'targetable' functional groups including *c-KIT* tyrosine kinase and *CYP24A1*, encoding a vitamin D3 metabolizing enzyme (**Supplementary Table 5**). c-KIT mRNA expression was abundant in luminal progenitors (from normal and *BRCA1* mutation carriers) and heterogeneous amongst luminal cells *in situ*, consistent with it being a marker of progenitor but not mature cells (**Supplementary Fig. 9**). It is notable that c-KIT mRNA expression was twofold higher in preneoplastic *BRCA1* mutation–associated versus non-BRCA1/2 breast tissue ($P = 0.006$). Furthermore, immunostaining of breast tumors from individuals with *BRCA1* germline mutations revealed that 11 of 21 (52.4%) *BRCA1*-associated basal-like tumors expressed c-KIT compared with two of seven (28.6%) non-BRCA1/2–associated basal-like tumors. Even some BRCA1-associated tumors with no basal-like features expressed c-KIT (two of six tumors; 33.3%) (**Supplementary Fig. 9**). Expression of this kinase has previously been reported in sporadic basal-like breast cancer but is infrequently observed in non–basal-like tumors[31,32].

DISCUSSION

Through *in vivo* transplantation studies and *in vitro* cellular assays, we provide evidence for an epithelial cell hierarchy in human mammary tissue that closely parallels that occurring in the mouse mammary gland. Despite some differences in the expression of cell surface markers between species, it is likely that the analogous subpopulations have highly conserved functions. For example, the human MaSC-enriched population lacks expression of the steroid hormone receptors, like the mouse MaSC subset[33]. Furthermore, the luminal progenitor subsets in both human and mouse *BRCA1*-mutant mammary tissue showed B27 factor independence *in vitro*. Notably, estrogen receptor-α was expressed by a substantial fraction of human luminal progenitor cells. Estrogen receptor-α may therefore directly mediate the partial efficacy provided by prophylactic oophorectomy in the prevention of basal breast tumors in *BRCA1* mutation carriers[34–36], compatible with reports suggesting that tamoxifen chemoprophylaxis may be protective[37]. These actively dividing luminal progenitor cells, which are necessary for normal development and homeostasis in cycling adult breast epithelium, presumably have stringent requirements for high-fidelity DNA repair provided by BRCA1.

Even though luminal progenitor cells represent a probable cancer-initiating population in *BRCA1* mutation carriers, the MaSC could still be the cell of origin for *BRCA1*-associated tumors. However, there are fourfold fewer repopulating MaSCs in mouse mammary glands deficient in *Brca1* (M.-L.A.-L., E.L., J.E.V. and M.A.B., unpublished data), further suggesting that the luminal progenitor population is a key cellular target for transformation in *BRCA1* preneoplastic tissue. The perturbed growth properties of these committed luminal progenitor cells are also likely to aberrantly influence luminal cell differentiation. In support of this notion, considerably higher cytokeratin-5/6 expression was noted in the mature luminal subset ($CD49f^{-}EpCAM^{+}$) in *BRCA1* preneoplastic tissue (**Supplementary Fig. 6**). The expression of certain basal markers such as cytokeratin-14 in *BRCA1*-associated tumors is also consistent with luminal progenitor cells having an altered differentiation program.

Here we have shown that the delineation of specific epithelial cell types within the human mammary hierarchy through cell fractionation, together with determination of their molecular profiles, unexpectedly revealed an aberrant luminal progenitor cell population in *BRCA1* mutation carriers, suggesting it may serve as a cellular target for oncogenic events. The remarkable overlap between the luminal progenitor expression signature and that of basal tumors has wider implications for developing prevention and therapeutic targets against basal breast cancer. The shared molecular signatures also suggest that a more accurate name for the basal-like subclass of breast tumors could be the 'luminal progenitor' subtype (**Fig. 4d**). Finally, molecules such as c-KIT may represent new targets for the elimination or modulation of luminal progenitor cells that are the harbinger of *BRCA1*-associated and other basal-like breast tumors.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemedicine/.

Accession codes. Illumina microarray data have been submitted to Gene Expression Omnibus with accession numbers GSE16997 and GSE17072.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

We thank K. Stoev and K. Johnson for excellent animal husbandry, S. Mihajlovic and E. Tsui for expert assistance with histology and F. Battye and his colleagues for expert help in the flow cytometry lab. We thank J. Sambrook, E. McGowan, E. Musgrove and J. Adams for invaluable discussions and R. Reddel (Children's Medical Research Institute) for hTERT-immortalized fibroblasts. We thank

K.U. Wagner (University of Nebraska Medical Center) for MMTV-Cre mice and A. Parlow (National Hormone and Pituitary Program, US National Institute of Diabetes, Digestive and Kidney Diseases) for prolactin. We gratefully acknowledge the invaluable contribution of numerous patients, surgeons, pathologists and tissue bank coordinators, and we thank A. Willems, E. Niedermayr, all kConFab research staff and Family Cancer Clinics and Clinical Follow-Up Study for their contributions to the kConFab resource, as well as the many families who contribute to kConFab. This work was supported by the Victorian Breast Cancer Research Consortium, the Australian National Health and Medical Research Council, the US National Breast Cancer Foundation, the US Department of Defense, the Susan G. Komen Breast Cancer Foundation, the Australian Stem Cell Centre, the Australian Cancer Research Foundation and the Victorian Cancer Biobank. kConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia.

AUTHOR CONTRIBUTIONS

E.L., F.V. and N.C.F. conducted most of the experiments and contributed to the writing of the manuscript. D.W. and G.K.S. performed the bioinformatic analyses and contributed to the writing of the manuscript. B.P, A.H.H. and M.-L.A.-L. performed RNA studies. D.E.G. and T.W. contributed to tissue preparation, immunohistochemistry and cell culture. F.F. helped optimize and performed some of the immunohistochemistry. A.P., H.J.T. and kConFab helped organize the accrual of the human breast tissue material. L.I.H. generated the hTERT-immortalized fibroblasts used for xenotransplantation studies. S.B.F. and M.Y. contributed to c-KIT staining and scoring. J.D.F. and M.A.B. contributed to the *Brca1* experiments in mice. J.E.V. and G.J.L. conceptualized the study, contributed to study design and drafted and finalized the writing of the manuscript.

Published online at http://www.nature.com/naturemedicine/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Turner, N., Tutt, A. & Ashworth, A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer* **4**, 814–819 (2004).
2. Herschkowitz, J.I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).
3. Perou, C.M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
4. Foulkes, W.D. BRCA1 functions as a breast stem cell regulator. *J. Med. Genet.* **41**, 1–5 (2004).
5. Ganesan, S. *et al.* Abnormalities of the inactive X chromosome are a common feature of BRCA1 mutant and sporadic basal-like breast cancer. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 93–97 (2005).
6. Narod, S.A. & Foulkes, W.D. BRCA1 and BRCA2: 1994 and beyond. *Nat. Rev. Cancer* **4**, 665–676 (2004).
7. Venkitaraman, A.R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171–182 (2002).
8. Xu, X. *et al.* Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nat. Genet.* **22**, 37–43 (1999).
9. Bouwman, P. & Jonkers, J. Mouse models for BRCA1 associated tumorigenesis: from fundamental insights to preclinical utility. *Cell Cycle* **7**, 2647–2653 (2008).
10. Furuta, S. *et al.* Depletion of BRCA1 impairs differentiation but enhances proliferation of mammary epithelial cells. *Proc. Natl. Acad. Sci. USA* **102**, 9176–9181 (2005).
11. Kubista, M., Rosner, M., Kubista, E., Bernaschek, G. & Hengstschlager, M. Brca1 regulates *in vitro* differentiation of mammary epithelial cells. *Oncogene* **21**, 4747–4756 (2002).
12. Liu, S. *et al.* BRCA1 regulates human mammary stem/progenitor cell fate. *Proc. Natl. Acad. Sci. USA* **105**, 1680–1685 (2008).
13. Clarke, R.B. *et al.* A putative human breast stem cell population is enriched for steroid receptor-positive cells. *Dev. Biol.* **277**, 443–456 (2005).
14. Dontu, G. *et al.* *In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270 (2003).
15. Gudjonsson, T. *et al.* Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev.* **16**, 693–706 (2002).
16. Raouf, A. *et al.* Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* **3**, 109–118 (2008).
17. Stingl, J., Eaves, C.J., Zandieh, I. & Emerman, J.T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res. Treat.* **67**, 93–109 (2001).
18. Villadsen, R. *et al.* Evidence for a stem cell hierarchy in the adult human breast. *J. Cell Biol.* **177**, 87–101 (2007).
19. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555–567 (2007).
20. Eirew, P. *et al.* A method for quantifying normal human mammary epithelial stem cells with *in vivo* regenerative ability. *Nat. Med.* **14**, 1384–1389 (2008).
21. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
22. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
23. Gusterson, B.A., Ross, D.T., Heath, V.J. & Stein, T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res.* **7**, 143–148 (2005).
24. Nagle, R.B. *et al.* Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. *J. Histochem. Cytochem.* **34**, 869–881 (1986).
25. Kuperwasser, C. *et al.* Reconstruction of functionally normal and malignant human breast tissues in mice. *Proc. Natl. Acad. Sci. USA* **101**, 4966–4971 (2004).
26. Asselin-Labat, M.L. *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat. Cell Biol.* **9**, 201–209 (2007).
27. Romijn, H.J., van Huizen, F. & Wolters, P.S. Towards an improved serum-free, chemically defined medium for long-term culturing of cerebral cortex tissue. *Neurosci. Biobehav. Rev.* **8**, 301–334 (1984).
28. Ma, Y. *et al.* The breast cancer susceptibility gene *BRCA1* regulates progesterone receptor signaling in mammary epithelial cells. *Mol. Endocrinol.* **20**, 14–34 (2006).
29. Poole, A.J. *et al.* Prevention of Brca1-mediated mammary tumorigenesis in mice by a progesterone antagonist. *Science* **314**, 1467–1470 (2006).
30. Michaud, J. *et al.* Integrative analysis of RUNX1 downstream pathways and target genes. *BMC Genomics* **9**, 363 (2008).
31. Nielsen, T.O. *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin. Cancer Res.* **10**, 5367–5374 (2004).
32. Simon, R. *et al.* KIT (CD117)-positive breast cancers are infrequent and lack KIT gene mutations. *Clin. Cancer Res.* **10**, 178–183 (2004).
33. Asselin-Labat, M.L. *et al.* Steroid hormone receptor status of mouse mammary stem cells. *J. Natl. Cancer Inst.* **98**, 1011–1014 (2006).
34. Kauff, N.D. *et al.* Risk-reducing salpingo-oophorectomy for the prevention of BRCA1- and BRCA2-associated breast and gynecologic cancer: a multicenter, prospective study. *J. Clin. Oncol.* **26**, 1331–1337 (2008).
35. Kauff, N.D. *et al.* Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation. *N. Engl. J. Med.* **346**, 1609–1615 (2002).
36. Rebbeck, T.R. *et al.* Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. *N. Engl. J. Med.* **346**, 1616–1622 (2002).
37. Narod, S.A. Modifiers of risk of hereditary breast cancer. *Oncogene* **25**, 5832–5836 (2006).

ONLINE METHODS

Human tissue samples and mice. We obtained normal breast tissue (confirmed by pathology) from reduction mammoplasties and prophylactic mastectomies of known *BRCA1*, *BRCA2* and non-BRCA1/2 mutation carriers from consenting individuals through kConFab[38], the Royal Melbourne Hospital Tissue Bank and the Victorian Cancer Biobank, with approval from the Human Research Ethics Committees of The Walter and Eliza Hall Institute of Medical Research and Melbourne Health. Tissue microarrays were from kConFab samples. Non-BRCA1/2 individuals have a strong family history of breast cancer (kConFab category 1), where no mutation in *BRCA1* or *BRCA2* has been identified in the family by high-sensitivity testing. Normal breast samples refer to reduction mammoplasty specimens, where family history is generally not known. We maintained NOD-SCID *Il2rg*$^{-/-}$ mice[39] (obtained from Jackson Laboratories) in our animal facility according to institutional guidelines. We obtained MMTV-Cre[40] and *Brca1*$^{f/f}$ mice[8] from K.U. Wagner and the US National Cancer Institute, respectively. All mouse experiments were approved by the WEHI Animal Ethics Committee.

Mammary cell preparations. We minced samples from human donors and digested them with 150 U ml^{-1} collagenase (Sigma), 50 U ml^{-1} hyaluronidase (Sigma) and 100 U ml^{-1} DNase (Worthington Biochemical) in DMEM with nutrient mixture F-12 Ham (DME-HAM) supplemented with 5% fetal calf serum (FCS), 5 μg ml^{-1} insulin, 2 mM glutamine, 10 ng ml^{-1} epidermal growth factor and 500 ng ml^{-1} hydrocortisone for 5 to 8 h at 37 °C. We sequentially digested the resulting organoid suspension with 0.25% trypsin and 1 mM EGTA (1 min, 37 °C) and 5 mg ml^{-1} dispase (Roche Diagnostics; 1 min, 37 °C). We obtained a single-cell suspension by filtration through a 40-μm cell strainer (BD-Falcon), and we removed red blood cells by lysis. We prepared mouse mammary cell suspensions as previously described[21].

Staining and sorting. We blocked cells with rat immunoglobulin (Jackson Immunolabs) and antibody to CD16 and CD32 Fcγ II and III receptors (WEHI Monoclonal Antibody Facility) for 10 min before incubation with the appropriate antibodies for 25 min at 4 °C. Where required, we incubated cells with streptavidin for 15 min at 4 °C. We resuspended the cells in 0.5 μg ml^{-1} propidium iodide and sorted them on a FACSAria flow cytometer (Becton Dickinson). We analyzed data with WEASEL software (http://www.wehi.edu.au/faculty/advanced_research_technologies/flow_cytometry/weasel_for_flow_cytometry_data_analysis/).

***In vivo* transplantation.** We injected sorted cells into cleared inguinal mammary fat pads of 3- to 4-week-old NOD-SCID *Il2rg*$^{-/-}$ female mice that had been cleared of endogeneous epithelium as previously described[41] with an adaptation of the 'humanization' method[25]. Briefly, we counted sorted cells and resuspended them in transplantation buffer (50% FCS, 0.04% trypan blue in PBS) containing hTERT-immortalized human mammary stromal fibroblasts (R. Reddel and L.I.H., unpublished data) and then injected them in a 10-μl volume into each mammary fat pad. Typically, we co-injected 500,000 fibroblasts comprising a 50:50 mix of unirradiated to irradiated cells (0.3 Gy). We also implanted estrogen pellets (0.7 mg), prepared by mixing estrogen powder (Sigma) with a silicone elastomer (Nusil Silicone Technology)[42], subcutaneously. We removed the recipient mammary fat pads for evaluation 8–10 weeks after transplantation. We analyzed whole mounts under a dissecting microscope. We excised structures and processed them for H&E or immunohistochemical staining to evaluate morphology or demonstrate human origin, respectively, as described below. For serial transplantation, we digested individual primary transplanted glands from virgin mice with 300 U ml^{-1} collagenase and 100 U ml^{-1} hyaluronidase for 45 min. We passed the cell suspension through a 40-μm strainer and resuspended the resulting pellet in 60–80 μl PBS containing 50% FCS and 0.04% trypan blue before injecting cleared mammary fat pads. We performed whole-mount analyses 5–8 weeks after transplantation.

Histology. For histological sections, we collected portions of the inguinal mammary gland and fixed them overnight in 4% (wt/vol) paraformaldehyde in PBS, pH 7.4 at 4 °C. We embedded mammary glands in paraffin and stained 1.5-μm sections with H&E. For whole-mount analysis, we fixed whole inguinal mammary glands overnight in Carnoy's solution (60% ethanol, 30% chloroform, 10% acetic acid, Merck) before staining with hematoxylin. Before sectioning, we passaged whole-mounted glands through xylene before embedding them in paraffin.

Antibodies, cytospin preparations and immunostaining. Unless otherwise specified, we obtained antibodies for flow cytometry from BD Pharmingen. Antibodies against human antigens were biotin-conjugated antibody to CD24 (Stem Cell Technologies), phycoerythrin (PE)-conjugated antibody to CD31, PE-conjugated antibody to CD45, PE-Cy5–conjugated antibody to CD49f, biotinylated antibody to CD49f (AbCam), allophycocyanin (APC)-conjugated antibody to CD133 (Miltenyi), PE-conjugated antibody to CD140b (eBiosciences), PE-conjugated antibody to CD235a, FITC-conjugated antibody to EpCAM (Stem Cell Technologies) and APC-Cy7–conjugated streptavidin. Antibodies against mouse antigens were PE-conjugated antibody to CD24, FITC-conjugated antibody to CD29 (Millipore), biotin-conjugated antibodies to CD31, CD45 and TERT-119, and APC-conjugated antibody to CD61.

We spun sorted cells onto glass slides with a Cytospin4 centrifuge (Shandon, Thermo). We then fixed the cells in 4% paraformaldehyde, 10% normal buffered formalin or acetone, and we then performed an antigen retrieval step in citrate buffer pH 6.0 using a DAKO pressure cooker (125 °C, 30 s). We incubated the cells sequentially with the primary antibody, biotinylated secondary antibodies (mouse-specific and rabbit-specific), followed by ready-to-use Vectastain avidin and biotinylated horseradish peroxidase macromolecular complex (ABC) reagent (Vector) and 3,3′-diaminobenzidine (DAKO) before counterstaining with hematoxylin.

Unless otherwise specified, we obtained antibodies for immunohistochemical staining from Novocastra. Antibodies to human antigens were: antibody to ALDH-1 (BD Pharmingen), antibody to HER2 (DAKO), antibody to cytokeratin-5/6 (DAKO), antibody to cytokeratin-14, antibody to cytokeratin-8/18, antibody to keratin-19 (Abcam), antibody to EGFR, antibody to estrogen receptor-α, antibody to Muc-1 (Stem Cell Technologies), antibody to nestin (Abcam), antibody to p63 (DAKO), antibody to c-KIT (DAKO), antibody to progesterone receptor A, antibody to milk (Accurate Chemical and Scientific Corp.), antibody to GATA-3 (Santa Cruz) and antibody to vimentin.

***In vitro* cellular assays.** We resuspended cells in 20 μl of endotoxin- and virus-free Matrigel (BD Pharmingen). We placed the Matrigel-cell mixture at the bottom of the well (eight-well Lab-Tek chamber slide, Nunc) and allowed it to set at 37 °C. We filled each well with 400 μl of medium containing 5 μg ml^{-1} insulin, 2 mM glutamine, 10 ng ml^{-1} epidermal growth factor, 500 ng ml^{-1} hydrocortisone and 20 ng ml^{-1} cholera toxin (Sigma). We cultured cells for 14 d in 5% CO_2 and 5% O_2. We supplemented the medium with either B27 (Gibco), mifepristone (RU486, Sigma) or prolactin (a gift from A. Parlow). We photographed, collected and then fixed the gels in 4% paraformaldehyde before processing them for immunostaining.

RNA preparation and quantitative reverse transcription PCR analysis. We isolated total RNA from primary mammary cell subpopulations with Trizol (Invitrogen) or the RNeasy Micro kit (Qiagen) and then treated it with DNase using the DNA-free kit (Ambion). We performed reverse transcription with oligo(dT) primer using Moloney murine leukemia virus reverse transcriptase (Invitrogen) according to the manufacturer's protocol.

We performed quantitative RT-PCR with a Rotorgene RG-6000 (Corbett Research) and SensiMix (dT) DNA kit (Quantace) under the following conditions: 10 min at 95 °C followed by 35 cycles consisting of 15 s at 95 °C, 20 s at 62 °C and 20 s at 72 °C. We determined gene expression with Rotor-Gene software (version 1.7). We used the following primers for quantitative RT-PCR (F, forward; R, reverse): *BRCA1* F: 5′-GAAGAAACCACC AAGGTCCA-3′ and *BRCA1* R: 5′-GTTGATCTGTGGGCATGTTG-3′. *18S* rRNA F: 5′-GTAACCCGTTGAACCCCATT and *18S* rRNA R: 5′-CCATCC AATCGGTAGTAGCG.

Microarray hybridizations. We purified total RNA from sorted cell populations and fresh frozen human breast tissue with the RNeasy Micro kit (Qiagen). For breast tissue, we first homogenized samples with the Polytron-Aggregate (Kinematica). We ascertained RNA quality with the Agilent Bioanalyzer 2100 (Agilent Technologies) using the Agilent RNA 6000 Nanokit (Agilent Technologies) according to the manufacturer's protocol. We labeled either 180 ng (for

total human breast tissue) or up to 500 ng (for sorted cell populations) with the Total Prep RNA amplification kit (Ambion). We prepared labeled complementary RNA (1.5 μg) for hybridization to Illumina HumanWG-6 V3 (R1) BeadChips. After washing, we coupled the chips with Cy3 and scanned them in an Illumina BeadArray Reader. Un-normalized summary probe profiles, with associated probe annotation, were output from BeadStudio.

Statistical analyses. We analyzed microarray data with Bioconductor and limma package software[43,44]. The datasets and the statistical methods used are described in detail in the **Supplementary Methods**.

38. Mann, G.J. *et al.* Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource. *Breast Cancer Res.* **8**, R12 (2006).
39. Shultz, L.D. *et al.* Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2Rγ null mice engrafted with mobilized human hemopoietic stem cells. *J. Immunol.* **174**, 6477–6489 (2005).
40. Wagner, K.U. *et al.* Cre-mediated gene deletion in the mammary gland. *Nucleic Acids Res.* **25**, 4323–4330 (1997).
41. Deome, K.B., Faulkin, L.J. Jr., Bern, H.A. & Blair, P.B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res.* **19**, 515–520 (1959).
42. Laidlaw, I.J. *et al.* The proliferation of normal human breast tissue implanted into athymic nude mice is stimulated by estrogen but not progesterone. *Endocrinology* **136**, 164–171 (1995).
43. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
44. Smyth, G.K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).