

OPEN

A CD146 FACS Protocol Enriches for Luminal Keratin 14/19 Double Positive Human Breast Progenitors

Ólöf Gerður Ísberg^{1,2}, Jiyoung Kim¹, Agla J. Fridriksdóttir¹, Mikkel Morsing^{1,3}, Vera Timmermans-Wielenga⁴, Lone Rønnev-Jessen⁵, Ole W. Petersen¹ & René Villadsen¹ 

Human breast cancer is believed to arise in luminal progenitors within the normal breast. A subset of these are double positive (DP) for basal and luminal keratins and localizes to a putative stem cell zone within ducts. We here present a new protocol based on a combination of CD146 with CD117 and CD326 which provides an up to thirty fold enrichment of the DP cells. We show by expression profiling, colony formation, and morphogenesis that CD146^{high}/CD117^{high}/CD326^{high} DP cells belong to a luminal progenitor compartment. While these DP cells are located quite uniformly in ducts, with age a variant type of DP (vDP) cells, which is mainly CD146-negative, accumulates in lobules. Intriguingly, in specimens with *BRCA1* mutations known to predispose for cancer, higher frequencies of lobular vDP cells are observed. We propose that vDP cells are strong candidates for tracing the cellular origin of breast cancer.

Once fully developed, the breast belongs to a category of tissues and organs with relatively little cellular turnover. The breast epithelium contains two major lineages, and by far the majority of mitotic activity, when present, takes place within the luminal epithelial lineage¹. This may explain why most breast cancers are caricatures of the cellular phenotypes found herein. We have previously investigated two major luminal cellular phenotypes in the human breast gland: Estrogen receptor α -positive (ER α^+) cells and keratin 14-positive (K14⁺) cells of which the latter also express luminal-specific keratins (hereafter referred to as double positive, DP cells)²⁻⁵. Interestingly, similar phenotypes are found within the two most frequent subtypes of human breast cancer – the luminal subtype and basal-like subtype, respectively⁶. While we have devised a protocol for the prospective isolation and culture of the ER α^+ , studying pure populations of DP cells has so far had to rely heavily on cloning of cultured cells with no or little prior enrichment^{4,5}. However, assays based on collecting ducts and lobules under the microscope have strongly suggested that DP cells are indeed progenitors originating primarily from within a ductal stem cell zone⁴. Others have described DP cells as primarily a lobular characteristic^{7,8}. In fact, there is increasing evidence for the existence of both ductal and lobular progenitor cells with different marker expression and differentiation potential⁸⁻¹¹. The exact characterization of luminal progenitors has been complicated by the fact that these populations seem to be highly variable among individuals¹⁰. For instance, both the number and the distribution of progenitor cells appear to change with donor age or number of generations in culture^{9,12}. In culture, we have previously enriched partially for DP cells based on CD117 and others have confirmed the progenitor properties of such populations^{12,13}. However, CD117 is widely expressed in the luminal compartment of the adult breast^{2,14}. Here, we enrich for DP cells by inclusion of CD146 in a FACS protocol and further show that the appearance of a CD146^{neg}, variant DP (vDP) cell correlates with higher age and predisposition to cancer.

¹Department of Cellular and Molecular Medicine and Novo Nordisk Foundation Center for Stem Cell Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ²Biomedical Center, University of Iceland, Reykjavik, Iceland. ³Division of Translational Cancer Research, Department of Laboratory Medicine, Lund University, Lund, Sweden. ⁴Department of Pathology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. ⁵Section for Cell Biology and Physiology, Department of Biology, Faculty of Science, University of Copenhagen, Copenhagen, Denmark. Ólöf Gerður Ísberg and Jiyoung Kim contributed equally. Correspondence and requests for materials should be addressed to R.V. (email: r.villadsen@sund.ku.dk)

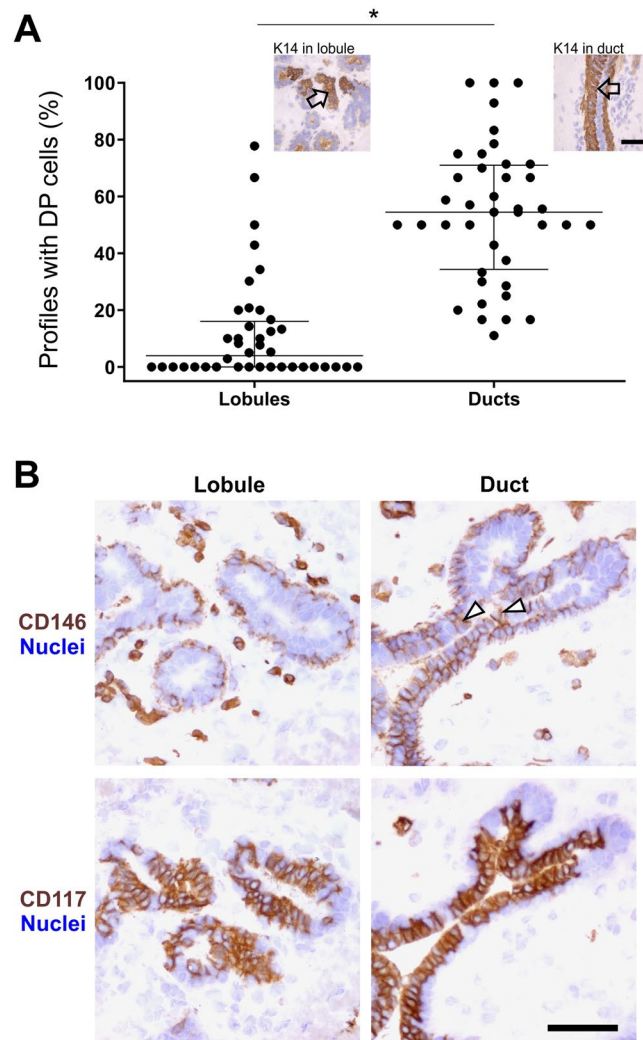


Figure 1. DP and CD146⁺ luminal cells accumulate in ducts of the human normal breast gland. **(A)** Dot plot comparing the proportion of structures containing DP cells shows that on average DP cells are more frequent in ducts. Image inserts show examples of DP cells marked by arrows in a lobule (left) and a duct (right) after immunoperoxidase stain. * $p < 0.0001$ tested by Mann-Whitney test ($n = 40$ biopsies, listed in Supplementary Table S1). Bar, 50 μm . **(B)** Representative immunoperoxidase stainings in serial sections demonstrating that CD146⁺ luminal cells are preferably present in ductal structures (upper panel), whereas CD117⁺ cells are generally present in both lobules and ducts (lower panel). Arrowheads mark examples of CD146⁺ cells in the luminal compartment. Bar, 50 μm .

Results

DP and CD146-positive cells accumulate in ducts of the human normal breast gland. We and others have previously demonstrated the presence of a stem cell zone rich in DP- or Aldehyde dehydrogenase-positive (ALDH⁺) cells in the ducts of the human breast^{4,15}. Here, based on immunohistochemistry and quantification of stained ductal and lobular profiles, in a sample of 40 human normal breast biopsies, we confirm a topographical accumulation of DP cells in ducts (Fig. 1A, Supplementary Fig. S1, Supplementary Tables S1 and S2). By applying candidates from our antibody repository¹⁶ in search for surface markers other than CD117 eligible for prospective isolation of DP cells we came across CD146, also referred to as melanoma cell adhesion molecule (MCAM). While CD146 in general is more widespread in stromal- and myoepithelial cells, in the luminal compartment, unlike CD117, CD146 is restricted mainly to the ducts (Fig. 1B, Supplementary Tables S3 and S4). Therefore, CD146 seemed promising as a more specific marker of DP cells than CD117.

A CD146^{high}/CD117^{high}/CD326^{high} FACS protocol enriches for DP cells. With the aim of enriching for DP cells we included CD146 in a previously described protocol used by us and others^{2,4,17,18} for isolation of CD326^{high} luminal cells and further stratified the cells according to CD117 expression (Fig. 2A). Thus, three subpopulations were sorted as CD146^{high}/CD117^{high}/CD326^{high}, CD146^{low}/CD117^{high}/CD326^{high}, and CD146^{low}/CD117^{low}/CD326^{high}, in the following referred to as CD146^{high}/CD117^{high}, CD146^{low}/CD117^{high} and CD146^{low}/CD117^{low}, respectively. The luminal identity of these cells was confirmed by the absence of α -smooth muscle actin

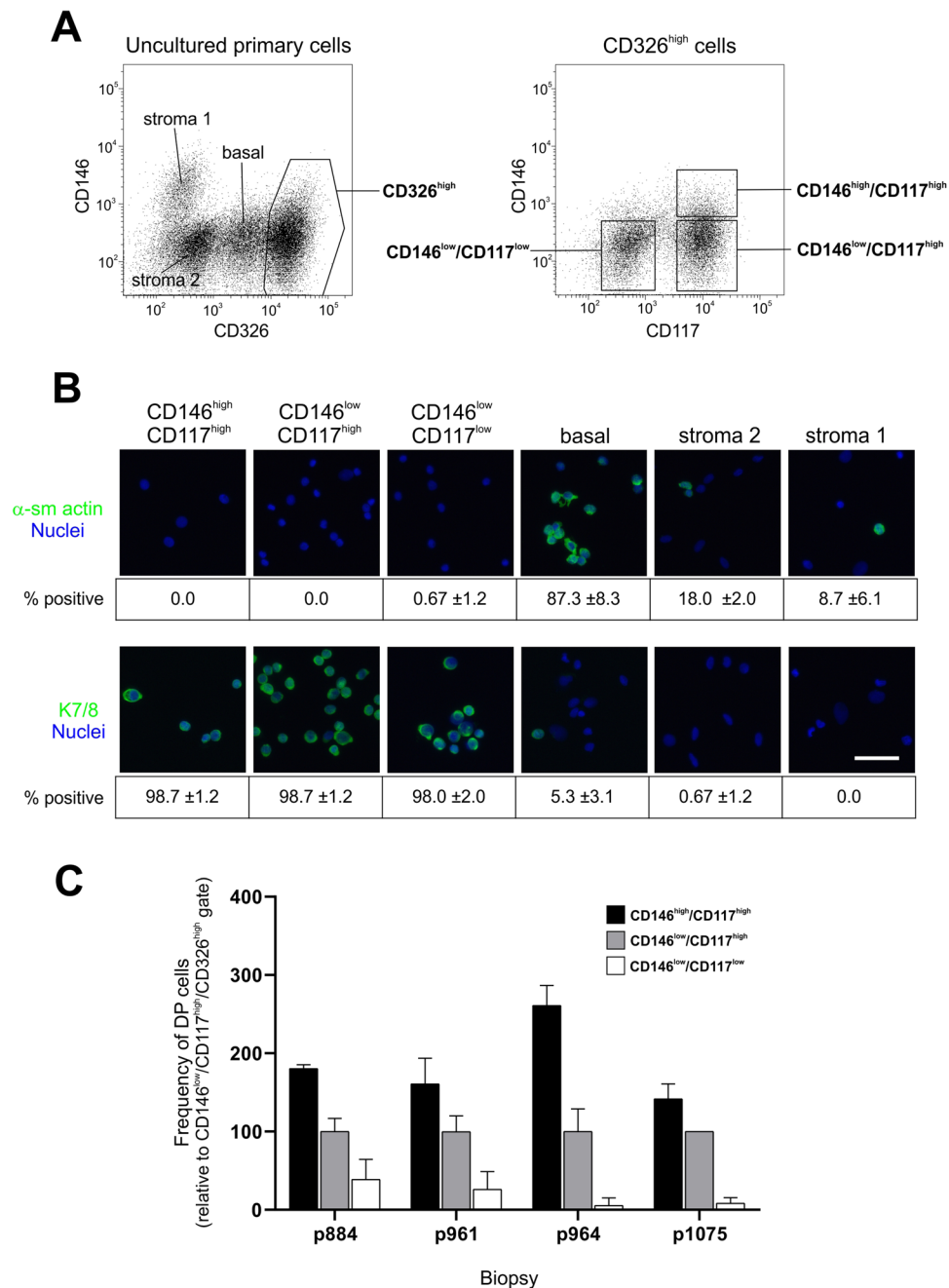


Figure 2. A CD146^{high}/CD117^{high}/CD326^{high} FACS protocol enriches for DP cells. (A) FACS diagrams showing uncultured primary cells gated into luminal CD326^{high} cells (left diagram, right frame) from which CD146^{high}/CD117^{high}, CD146^{low}/CD117^{high} and CD146^{low}/CD117^{low} populations were selected (right diagram). (B) Smears of sorted CD146^{high}/CD117^{high}, CD146^{low}/CD117^{high}, CD146^{low}/CD117^{low}, basal and stromal cells immunostained for α -sm actin and K7/8. Bar, 50 μ m. The low frequencies of α -sm actin-positive cells in the luminal subpopulations are indicative of few, if any, contaminating basal cells. (C) Bar graph of the relative number of DP cells in FACS sorted populations showing a higher frequency of K14⁺/K19⁺ DP cells from the CD146^{high}/CD117^{high} gate (black bars) compared to CD146^{low}/CD117^{high} (grey bars) and CD146^{low}/CD117^{low} cells (white bars) ($p < 0.05$ tested by ANOVA with Tukey HSD test, $n = 4$ biopsies). The frequency of CD146^{low}/CD117^{high} DP cells was set at an arbitrary level of 100.

(α -sm actin) and the presence of luminal keratin 7/8 in stainings of cellular smears (Fig. 2B). When comparing sorted luminal subpopulations with respect to the presence of DP cells in terms of K19/K14 double positivity the CD146^{high}/CD117^{high} gate consistently came out as the one with the highest frequency of DP cells (Fig. 2C). The CD146^{high}/CD117^{high} gate typically comprised 10% of the total CD326^{high} luminal population (ranging from 5 to

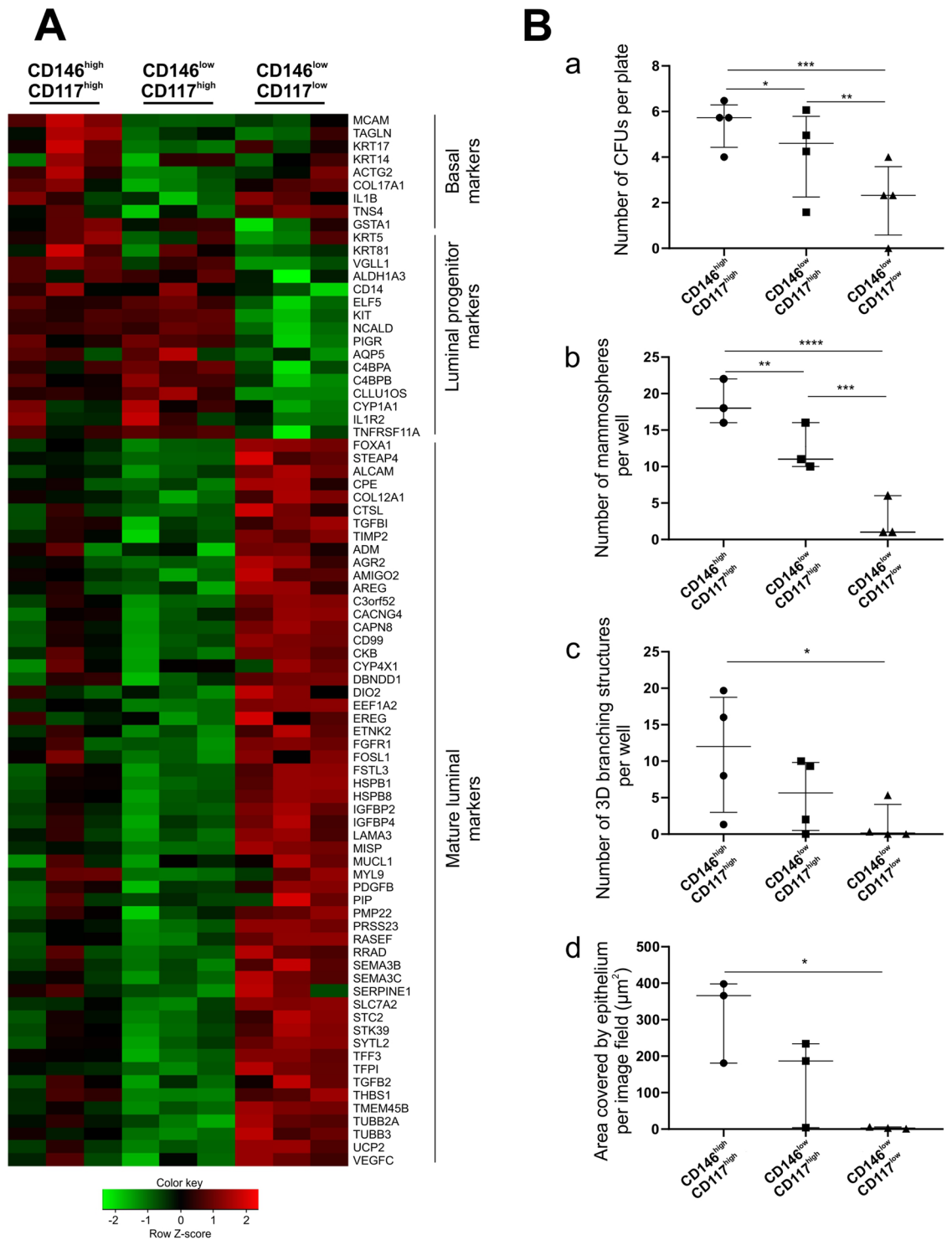


Figure 3. Characterization of CD146^{high}/CD117^{high} cells. **(A)** Heatmap of selected DEGs in three luminal populations isolated by FACS. Genes are sorted according to lineage into basal, luminal progenitor and mature luminal markers. The RNA-Seq data demonstrate that the CD146^{high}/CD117^{high} population differs from CD146^{low}/CD117^{high} cells by a higher expression of basal markers, while CD146^{low}/CD117^{low} cells have increased expression of mature luminal markers. Log₂-scaled expression values are presented for each gene, and the color key indicate the Z-score (n = 3 biopsies). **(B)** Functional progenitor assays demonstrate that CD146^{high}/CD117^{high} cells generally exhibit the highest level of activity. **(a)** In a colony forming unit (CFU) assay significantly more colonies were generated from CD146^{high}/CD117^{high} cells compared to the other two populations (n = 4 biopsies). **(b)** In a mammosphere assay CD146^{high}/CD117^{high} cells generated significantly more spheres than the other two populations (n = 3 biopsies). **(c)** In a 3D morphogenesis assay only CD146^{high}/CD117^{high} cells developed significantly more branching structures than CD146^{low}/CD117^{low} cells (n = 4 biopsies). **(d)** In a branching morphogenesis assay on fibroblast feeders only CD146^{high}/CD117^{high} cells formed

significantly more branching structures than CD146^{low}/CD117^{low} cells (n = 3 biopsies). *p < 0.05, **p < 0.005, ***p < 0.0005, and ****p < 0.00005, tested by ANCOVA with Tukey's HSD test. Each scatter dot plot is lined at median with interquartile range.

15%, data not shown). In summary, these results show that isolating a CD326^{high} luminal CD146^{high}/CD117^{high} population enriches for DP cells.

Characterization of CD146^{high}/CD117^{high} cells. To compare CD146^{high}/CD117^{high} cells with CD146^{low}/CD117^{high} and CD146^{low}/CD117^{low} cells RNA sequencing (RNA-Seq) analysis was performed (Supplementary Tables S5 and S6) and differentially expressed genes (DEGs) among the three subpopulations were established by the NOISeq method (Supplementary Table S7). As revealed in a heatmap of selected DEGs, the CD146^{high}/CD117^{high} cells expressed a number of basal/myoepithelial markers in addition to typical markers of progenitor cells (Fig. 3A). In this respect, it is worthwhile to emphasize that this expression concerns RNA only and not detectable protein. Thus, we have no reason to believe that it reflects contamination with myoepithelial cells per se. Rather, it may relate to what others have found in mice of some transient hybrid multipotent progenitors co-expressing basal and luminal markers¹⁹. Both CD146^{high}/CD117^{high} as well as CD146^{low}/CD117^{high} cells expressed relatively high levels of several luminal progenitor markers, including *KIT* and *TNFRSF11A/RANK* that have been correlated to progenitors accumulating in preneoplastic breast tissue of *BRCA1* mutation carriers^{20,21}. By contrast, the CD146^{low}/CD117^{low} cells remained mature luminal-like with higher levels of, for instance, *ALCAM* and *ESR1/ER α* (Fig. 3A). Application of gene ontology (GO) analysis as well as computerized pathway enrichment analysis using Kyoto encyclopedia of genes and genomes (KEGG) database further revealed cytoskeletal components and pathways of focal adhesions as most significantly different between CD146^{high}/CD117^{high} and CD146^{low}/CD117^{high} populations (Supplementary Table S8 and Supplementary Fig. S2).

To functionally assess the progenitor potential of the three sorted luminal subpopulations we analyzed their performance in the following assays: a colony forming unit (CFU) assay, a mammosphere assay and a three dimensional (3D) morphogenesis assay, which have all been implemented previously by us and others to analyze progenitor activity of breast epithelial cells^{4,5,22,23}. In addition, we included a two dimensional (2D) feeder based morphogenesis assay which allows for analysis of branching structures upon epithelial-stromal interaction²⁴. Based on these assays, the CD146^{high}/CD117^{high} cells demonstrated the highest progenitor potential (Figs 3B and S3).

In order to further characterize the luminal CD146^{high}/CD117^{high} profile *in situ*, we performed multicolor immunohistochemistry that demonstrates correlation between CD146 and a number of putative progenitor markers in ductal structures (Fig. 4). As expected CD146-positive cells could be found to coexpress CD117 or K14. Furthermore, markers we previously described as being present in ductal stem cell zones, including SSEA4 as well as keratins 5, 15 and 17⁴, were co-expressed with CD146 in cells of the luminal compartment. By contrast, CD146 was generally not present in cells that were positive for mature luminal marker Ks20.8².

Based on these findings, we conclude that inclusion of CD146 contributes a significant improvement over that of CD117/CD326 or CD326 only in attempts to prospectively isolate human breast DP cells with a progenitor profile.

Variant DP cells accumulate in lobules with age and in tissue from BRCA1 mutation carriers.

We have previously reported that in culture the number of DP cells increases with the number of population doublings¹². This prompted us to test whether the frequency of DP cells also *in situ* increases with age. Indeed, based on multicolor imaging of smears directly from biopsies we found an increase in relative frequency of DP cells with age (n = 20 samples, Fig. 5A and Supplementary Table S9). Keeping in mind that DP cells are already present in relatively high numbers in the ducts, it was not surprising that the age-related increase in DP numbers manifested itself in the lobules when comparing “young” (here defined as <29 years old with 2.9% lobules containing DP cells) and “old” (>29 years with 20.5% lobules containing DP cells) women (Fig. 5B). To analyze whether DP cells in lobules differ from DP cells in ducts we investigated a number of biopsies by immunofluorescent staining. As it turned out lobular DP cells were mostly CD146^{neg}, and thus referred to here as variant DP (vDP cells) (Fig. 5C and Supplementary Table S10). This led us to speculate on a possible pathophysiological role of vDP cells in breast cancer evolution which is after all an age-related disease. To get a preliminary impression of this we examined a sample of breast tissue specimens from women with known mutations in the *BRCA1* gene and another sample of basal-like breast cancers with the majority of the neoplastic cells being DP. As the normal-derived samples from *BRCA1* mutation carriers were completely anonymously donated, we could not make an exact age-matching of this material to that from presumed non-carriers. However, there is no reason to believe that the *BRCA1*-affected women were particularly old when undergoing mastectomy of the breast²¹. Irrespective of age, the tissue samples from *BRCA1* mutation carriers were characterized by having significantly more DP cells (40.5% lobules containing DP cells) (Fig. 5B). Furthermore, these were more active in terms of cell cycling (Fig. 6 and Supplementary Table S11). Both lobular DP cells from *BRCA1* mutation carriers and cancer associated DP were generally CD146^{neg} and as such similar to age related lobular vDP cells (Supplementary Tables S10 and S12).

Collectively, we conclude that CD146^{high}/CD117^{high}/CD326^{high} progenitors represent what could be referred to as the default luminal progenitor in the human breast and that additional vDP cells arise with age and under pathological conditions, suggesting that vDP cells represent precursor cells and their descendants in tissues predisposed to cancer development as well as in overt cancer.

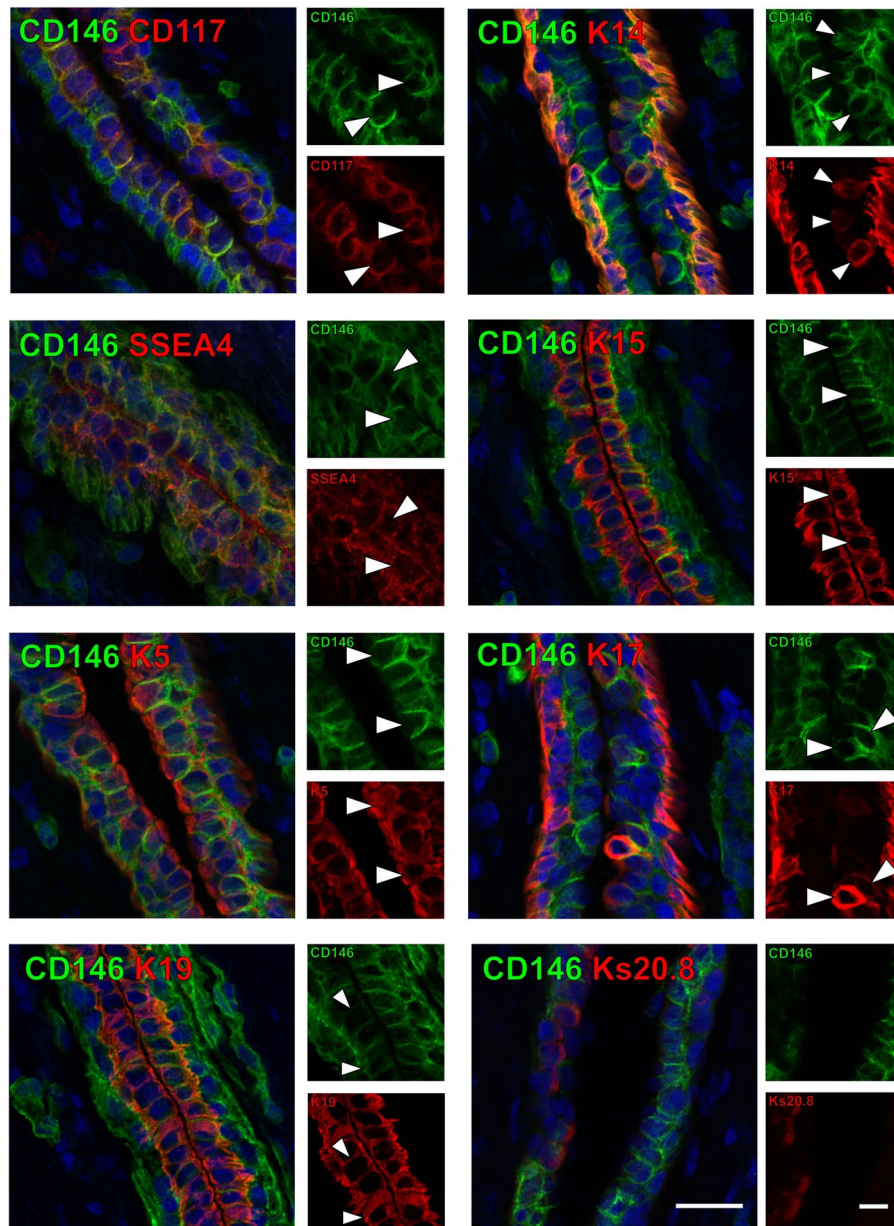


Figure 4. CD146^{high} luminal cells express with a number of progenitor markers. Immunofluorescent double stainings of CD146 (green) along with putative progenitor markers CD117, K14, SSEA4, K15, K5, K17 as well as K19 and Ks20.8 (all in red) in normal breast tissue sections. Nuclei are stained with DAPI (blue). Image subsets are shown in single color channels, and arrowheads mark examples of co-stained cells. While CD146^{high} cells co-express progenitor markers and K19, CD146^{high} and Ks20.8⁺ cells are generally confined to separate luminal compartments. Bar, 50 μ m. Image subsets, bar, 20 μ m.

Discussion

In this study we present a novel protocol for enrichment of luminal progenitors harboring DP cells and provide evidence for their progenitor status. The preferential location of these progenitors within ductal zones is in line with our previous observation based on culturing cells from microcollected breast gland structures⁴. Apart from demonstrating an increased level of activity in culture-based progenitor assays, the CD146^{high}/CD117^{high}/CD326^{high} cells have a unique expression profile showing an increased level of basal/myoepithelial markers. This profile is reminiscent of the transient hybrid population that has been described during early development of the mouse mammary gland¹⁹, suggesting that the CD146^{high}/CD117^{high}/CD326^{high} compartment harbors an early luminal progenitor close to the apex of hierarchical differentiation tree. In support of this, others have recently published a study including single cell RNA profiling to evaluate the heterogeneity of ALDH⁺ luminal progenitors, in which the cluster with the highest levels of canonical markers of breast stemness also expressed high levels of CD146¹⁰. Both CD146^{high}/CD117^{high} and CD146^{low}/CD117^{high} cells express higher levels of several luminal progenitor markers compared to CD146^{low}/CD117^{low} luminal cells, including TNFRSF11A/RANK. As with

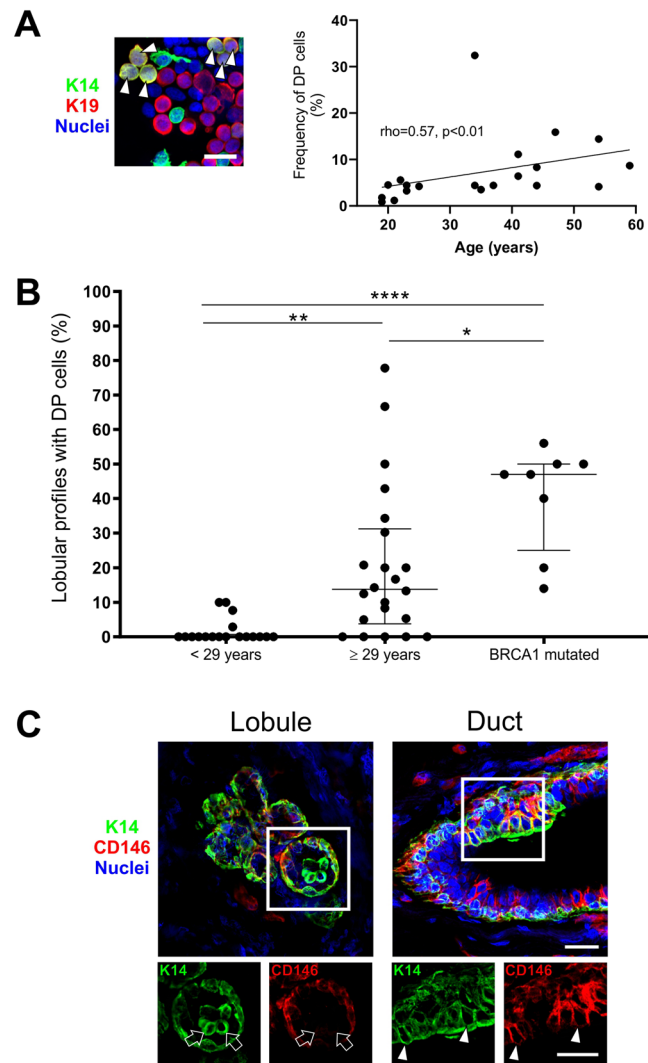


Figure 5. Variant DP cells accumulate in lobules with age and in tissue from *BRCA1* mutation carriers. (A) Immunofluorescent staining of crude smears with K14 (green), K19 (red) and nuclei (blue) (left image). Arrowheads mark DP cells. Bar, 20 μ m. A positive correlation was found between age and the frequency of DP cells (right), analyzed by Spearman rank test ($\rho = 0.57$, $p < 0.01$). (B) Dot plots of the proportion of lobular structures with DP cells in women < 29 years (average age: 19.7 years, median: 19), ≥ 29 years (average age: 44.4 years, median: 43) and *BRCA1* mutation carriers (information about donor age not available). The age-selected data for lobules are derived from Supplementary Table S1, which are also included in Fig. 1A. The proportion of lobules containing DP cells is significantly higher in the older age group (20.5%) and in women with *BRCA1* mutations (40.5%) as compared to the younger age group (2.9%). * $p < 0.05$, ** $p < 0.005$, and **** $p < 0.000005$ using ANOVA with Tukey's HSD test. Each scatter dot plot is lined at median with interquartile range (< 29 years: $n = 18$, ≥ 29 years: $n = 22$, *BRCA1* mutated: $n = 8$). (C) Immunofluorescent staining demonstrating a segregation of DP and CD146⁺ cells in a lobule (left image) compared to a duct (right image) with CD146-expressing DP cells. K14 is shown in green, CD146 in red and nuclei in blue. Image subsets are shown in single color channels. Arrows indicate DP cells that are CD146^{neg} while arrowheads point to CD146⁺ DP cells. Bars, 25 μ m.

KIT/CD117 increased expression of RANK has been implicated with aberrant accumulation in preneoplastic breast epithelium of *BRCA1* mutation carriers, which in turn suggests that such cells could be a target for breast carcinogenesis^{20,21}.

The age-related increase of DP cells reported here concurs with an earlier study in which culture based assays demonstrate age-dependent expansion of DP cells¹². That DP cells tend to accumulate with age, and are more frequently present in lobular structures in women ≥ 29 years of age may be of significance with regard to carcinogenesis since the majority of breast cancers arises within terminal ductal lobular units²⁵. Furthermore, the presence of CD146^{neg} or CD146^{low} v DP cells in lobules implies that this population could be a different entity than its ductal counterpart expressing CD146. Recent data provided by our laboratory suggest that ducts and lobules do in fact contain basal stem cells or progenitors with different properties which may impact on the progeny at the

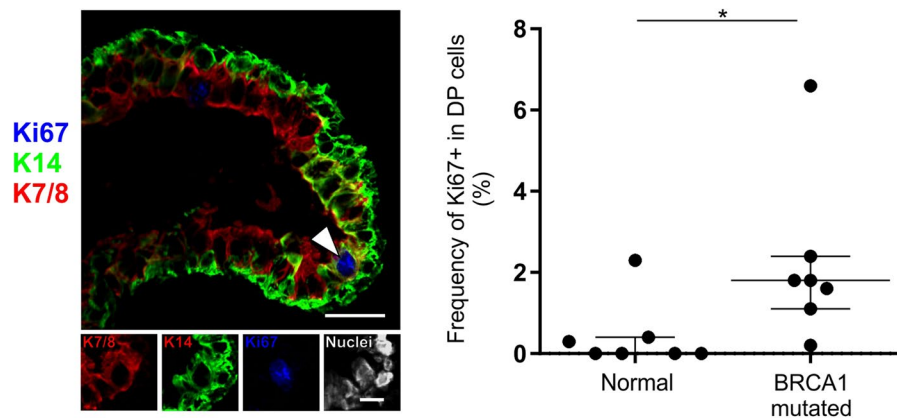


Figure 6. DP cells are more frequently cycling in tissue from BRCA1 mutation carriers. (A) Normal tissue from a woman with known BRCA1 mutation immunostained for K14 (green), cell cycle marker Ki67 (blue) and luminal keratin marker CAM5.2 (red). Arrowhead marks a Ki67⁺ DP cell. Bar, 25 μ m. Lower panel image subsets are shown in single color channels, including DAPI nuclear stain (white). Bar: 10 μ m. (B) Dot plot comparing the proportion of Ki67⁺ cells in the DP population between women with BRCA1 mutations and women with no known mutations. * $p < 0.017$ (normal: $n = 7$; BRCA1 mutated: $n = 7$) by Mann Whitney test.

respective anatomical regions⁹. Intriguingly, the vDP-like phenotype observed here in selected basal-like tumors is not incompatible with their origin in lobules.

BRCA1 mutation carriers have a profound increased lifetime breast cancer risk²⁶. Lim *et al.* reported that the breast glands of BRCA1 mutation carriers have increased numbers of luminal progenitors, suggesting that this population contains the cell of origin for subsequent tumor formation²⁰. While here we cannot correlate the data of the BRCA sample with age, indeed, irrespective of age, the frequency of DP cells is higher among BRCA1 mutation carriers. DP cells of BRCA1 mutation carriers are also more frequently cycling compared to non-carriers as demonstrated by Ki67 staining, which supports that these women may host luminal progenitors at an aberrant frequency. An increased proliferative state could also imply an increased risk of accumulating transforming mutations in these cells. Furthermore, just as is generally the case for sporadic basal-like tumors, breast carcinomas derived from BRCA1 mutation carriers are more proliferative when compared to ER α ⁺ tumors^{27,28}.

In conclusion, we suggest that the luminal progenitor compartment containing DP cells is key for understanding the cellular origin of breast cancer.

Materials and Methods

Human breast tissue samples. Normal breast biopsies were acquired from healthy women undergoing reduction mammoplasty for cosmetic reasons at Capio CFR Hospitaler, Lyngby or Hellerup, Denmark. All donors were informed before surgery and agreed by written consent to donate tissue. The specimen donor's personal information was confidential and protected except the date and year of birth. In some cases only the year of birth was available. Material from some of the biopsies have been used previously in several studies^{2,9,16}; however, picking the sample material was done randomly and all relevant tissue processing were done specifically for this study which has not been previously published. Normal breast tissue specimens from BRCA1 mutation carriers as well as breast carcinomas were obtained completely anonymously from patients at the State University Hospital, Copenhagen. The storage and use of human material has been reviewed and approved by the Regional Scientific Ethical Committees (H-2-2011-052 and H-3-2010-095) and the Danish Data Protection Agency (2011-41-6722) and has been handled according to established guidelines in subsequent experiments. In total, normal-derived samples from 80 women were included of which 10 were known BRCA1 mutation carriers. Of the biopsies that were used for flow cytometry four were immunostained, three biopsies were submitted to RNA expression profiling and nine biopsies were included in cell culture assays. 23 ER α -negative breast carcinomas were selected from a sample material of 150 invasive breast carcinomas.

Immunostaining. *Immunosmears:* The 20 biopsies included for staining of crude smears were picked randomly among material used for other routine analyses in the lab. Isolation of breast epithelial organoids was performed as previously described²⁹. Primary epithelial organoids were dissociated into single cells with 0.25% trypsin and 1 mmol/L EDTA. Cells were suspended in PBS with 10% goat serum, dripped onto glass slides with a pipette and air dried before fixation in either ice cold methanol for 5 minutes or in 3.7% formaldehyde (Merck, Darmstadt, Germany) followed by permeabilization with Triton X-100 detergent (Sigma-Aldrich). Smears were incubated with primary antibodies for at least 2 hours before incubation with isotype-specific secondary fluorescent antibodies for 30 minutes (Supplementary Table S13). Double positive (DP) cells were detected by co-expression of K14/K19. Additionally, crude smears from some biopsies were also analyzed by co-expression of K14/K7/8 (using CAM5.2 that detects keratins 7 and 8³⁰). K7 is more widely expressed in the luminal compartment compared to K19⁹. Cells sorted by FACS were spun down, dripped onto glass slides and fixed as described above. Quantification of DP cells among sorted cells was done by quantifying 3×100 cells for each population ($n = 4$ biopsies). Analysis of the presence of basal cells and luminal cells in sorted cells was performed with

staining for α -sm actin and CAM5.2, respectively. For each population 3×100 cells were quantified. Nuclei were stained with DAPI (Life Technologies, Thermo Fisher Scientific). **Immunohistochemistry:** Snap frozen tissue specimens were cut at $6 \mu\text{m}$ on a cryostat (Microm HM560, Axlab, Denmark) and were subsequently fixed either in methanol or formaldehyde as described for immunosmears. For peroxidase stainings, Ultravision ONE Detection System was utilized according to the manufacturer's instructions (Labvision, Thermo Fisher Scientific). Nuclei were stained with hematoxylin. For immunofluorescence the specimens were incubated for 2 hours with primary antibody followed by incubation with secondary isotype-specific Alexa Fluor antibodies, as described for stainings of smears. Nuclei were stained with DAPI. Antibodies and dilutions are listed in Supplementary Table S13.

Microscopy. Routine analysis of brightfield and fluorescent samples were performed on a Leica DM550B microscope. Peroxidase images were acquired using this microscope equipped with a DFC550 digital camera (Leica). Fluorescent images presented in the figures were acquired using a confocal microscope (Zeiss LSM 700) located at the Core Facility for Integrated Microscopy, Faculty of Health and Medical Sciences, University of Copenhagen. Phase contrast images were acquired using a Leica DM IL equipped with a MC170 HD digital camera. For image acquisition at low magnification a Leica MDG41 equipped with a MC170 HD digital camera was used.

Quantification and statistical analysis. Quantification of the frequency of structures containing luminal K14^+ cells were done under the microscope based on sections stained by immunoperoxidase. A lobule or duct with one or more positive cells present was scored as positive. Evaluation of the distribution of CD146- and CD117-expressing luminal cells were performed on serial sections stained by immunoperoxidase. Quantification of the frequency of K14/K19 and K14/K7/8 DP cells on smears were performed either directly under the microscope utilizing filters that visualize Alexa Fluor 488, Alexa Fluor 568 or both simultaneously or on images that were acquired and digitally stored. Quantification of the frequency of K14/K7/8 DP cells in ducts and lobules (Table S2) were performed on digitally stored images. In both cases areas of interest was located based on nuclear DAPI stain to avoid any bias in selection. A TDLU consists of the most distal part of the ductal-lobular system and include the collection of lobular ductules/acini branching from the intralobular terminal duct, as well as the connecting extralobular terminal duct³¹ (Fig. S1). For evaluation and quantification of DP cells in lobules, intralobular ducts were regarded as a ductal segment to distinguish between ductules/acini and the corresponding duct when visually distinguishable – either by histology or by the presence of distinct $\text{K14}^+/\text{K7/8}^-$ basal cells. Quantification of the proportion of DP cells expressing Ki67 was done using a confocal microscope by first assessing the number DP cells in a structure before evaluating Ki67 expression by image acquisition (Supplementary Table S11). Generally, several sections from a biopsy sample were evaluated. All statistical analyses were performed using the computing program R (Version 3.5.1) and graphs were drawn either in R or the graphic pad Prism (version 8.0) and mentioned in figure- and table legends where statistics was used. In Denmark the average age of primiparous women is 29 years (Danmarks Statistik 2017, www.statistikbanken.dk/FOD11), which was used to divide women into two age-groups (Fig. 5B).

FACS. Primary breast organoids were dissociated into single cells as described for immunosmears. Viable cell suspensions were filtered through a $100 \mu\text{m}$ filter and resuspended in HEPES buffer supplemented with 0.5% BSA (bovine fraction V; Sigma-Aldrich) and 2 mM EDTA, and incubated for 30 min at 4°C with antibodies CD146 (1:20, clone P1H12, Alexa Fluor (AF) 647 conjugate, BD Biosciences), CD117 (1:20, clone 104D2, PE conjugate, BD Biosciences) and CD326 (1:20, clone 9C4, AF488 conjugate, BioLegend) (See Supplementary Table S13). Controls were incubated without antibodies and in selected experiments as Fluorescence Minus One (FMO) controls omitting CD146 antibody. Following incubation the cells were washed twice in HEPES/BSA/EDTA buffer and filtered through a 10 or $20 \mu\text{m}$ filter cup (Filcons). For dead cell discrimination, $1 \mu\text{g}/\text{mL}$ propidium iodide (Invitrogen) was added prior to analysis in a flow cytometer (FACSARIA II or FACSFusion, BD Biosciences). When sorting into 96 well plates the instrument's automated cell deposition unit (ACDU) and single cell sorting parameters were utilized.

RNA-Sequencing. Total RNA from $\text{CD146}^{\text{high}}/\text{CD117}^{\text{high}}$, $\text{CD146}^{\text{low}}/\text{CD117}^{\text{high}}$ and $\text{CD146}^{\text{low}}/\text{CD117}^{\text{low}}$ luminal cells sorted from three different biopsies was isolated using $500 \mu\text{L}$ TRIzol[®] (Invitrogen) and a spin-column method (Zymo research). Library construction, sequencing process and the data analysis were conducted by Beijing Genomics Institute, as described previously⁹. In short, cDNA library of the enriched mRNA was fragmented into 200 bps and sequencing adaptors were ligated and amplified before quality control with a Agilent 2100 bioanalyzer. The libraries were sequenced on a HiSeq2000 sequencer (Illumina), 50 bp high output to average depth of 10 million reads. The summary of RNASeq data is included in Supplementary Table S5 and a detailed list of genes in Supplementary Table S6. Readout alignment was performed by SOAP (version 2.21) and gene expression was quantified by RPKM algorithm. Differently expressed genes among groups were identified by the NOISeq method using a cutoff value as fold difference higher than 2 between two groups with probability higher than 0.8³². The selection of basal-, luminal progenitor- and mature luminal markers shown in the heatmap in Fig. 3A was based on previously published expression data^{2,20,33}. GO and KEGG databases were used for gene ontology enrichment analysis and pathway enrichment analysis, respectively.

Cell culture. All cell cultures were incubated in a humidified incubation chamber in 5% CO_2 at 37°C .

CFU assay: Each of the $\text{CD146}^{\text{high}}/\text{CD117}^{\text{high}}$, $\text{CD146}^{\text{low}}/\text{CD117}^{\text{high}}$ and $\text{CD146}^{\text{low}}/\text{CD117}^{\text{low}}$ subpopulations were sorted directly into a 96 well plate precoated with collagen I (Corning) as 10 cells per well in the presence of $100 \mu\text{L}$ FAD2 medium (DMEM, high glucose, no calcium (Life Technologies) 3:1 v/v with 2 mM glutamine, $0.5 \mu\text{g}/\text{mL}$ hydrocortisone (Sigma-Aldrich), $5 \mu\text{g}/\text{mL}$ insulin (Sigma-Aldrich), $10 \text{ ng}/\text{mL}$ cholera toxin (Sigma-Aldrich), $10 \text{ ng}/$

mL EGF (Peprotech), 1.8×10^{-4} M adenine, $10 \mu\text{M}$ Y-27632 (Y0503, *Sigma-Aldrich*) and 5% FCS (*Sigma-Aldrich*) as described in². Cultures were monitored for 14 days before fixation and hematoxylin staining. Proliferating clonal islets were determined under a microscope as colonies consisting of more than 25 cells. The sample material was derived from four different biopsies. Mammosphere assay: Sorted cells were counted and plated in triplicates in *ultra-low* attachment 24-well plates (Costar) with 5,000 cells/well in mammary epithelial growth medium (MEGM, MEBM (Lonza) supplemented with 2% B27 (Invitrogen), 20 ng/mL EGF (Peprotech), 20 ng/mL bFGF (Peprotech), and 4 $\mu\text{g/mL}$ heparin (Sigma)). Cultures were monitored for 14 days before final evaluation was performed by counting mammospheres $\geq 80 \mu\text{m}$ in diameter. The sample material was derived from three different biopsies. 3D morphogenesis assay: Sorted cells were counted and 20,000 cells per well were embedded in triplicates in 300 μL Matrigel (BD Biosciences) on top of a primary epithelial feeder overlaid by 250 μL Matrigel in 24 well plates. The cells were cultured in F12 medium (Ham's F12 (Invitrogen) supplemented with 2 mM glutamine, 5 $\mu\text{g/mL}$ insulin, 1 $\mu\text{g/mL}$ hydrocortisone, 0.1 $\mu\text{g/mL}$ cholera toxin, 10 ng/mL EGF and 5% FCS). After 21 days branching structures were quantified under a phase contrast microscope. The sample material was derived from four different biopsies. Branching morphogenesis assay with epithelial-stromal interaction: Sorted cells were seeded in triplicates of 5,000 cells on confluent human CD105^{high}/CD26^{low} fibroblast feeders in T12.5 culture flasks. This *co-culture* condition was previously established with modified breastoid base medium without HEPES (DMEM/F-12, 1:1), 1 $\mu\text{g/mL}$ hydrocortisone (*Sigma-Aldrich*), 9 $\mu\text{g/mL}$ insulin (*Sigma-Aldrich*), 5 $\mu\text{g/mL}$ transferrin (*Sigma-Aldrich*), 5.2 ng/mL Na-Selenite (BD Industries), 100 μM ethanolamine (*Sigma-Aldrich*), 20 ng/mL bFGF, 5 nM amphiregulin (R&D Systems), with the addition of 10 μM Y-27632 (Axon Medchem), 1.8×10^{-4} M adenine (Sigma Aldrich) and the serum replacement B27 (20 $\mu\text{L/mL}$, Life Technologies)²⁴. After 14 days cultures were fixed and stained for K19 by immunoperoxidase. After image acquisition stained epithelial structures were analyzed and quantified (ImageJ version 1.49 t) as described previously²⁴. Briefly, epithelial structures were segmented followed by analysis of the number of structures and the total area using the plugin Analyse Particles. The sample material was derived from three different biopsies.

References

- Clarke, R. B., Howell, A., Potten, C. S. & Anderson, E. Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res.* **57**, 4987–91 (1997).
- Fridriksdottir, A. J. *et al.* Propagation of oestrogen receptor-positive and oestrogen-responsive normal human breast cells in culture. *Nat. Commun.* **6**, 8786 (2015).
- Hopkinson, B. M. *et al.* Establishment of a normal-derived estrogen receptor-positive cell line comparable to the prevailing human breast cancer subtype. *Oncotarget* **8** (2017).
- Villadsen, R. *et al.* Evidence for a stem cell hierarchy in the adult human breast. *J. Cell Biol.* **177**, 87–101 (2007).
- 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).
- Alshareeda, A. T. *et al.* Characteristics of basal cytokeratin expression in breast cancer. *Breast Cancer Res. Treat.* **139**, 23–37 (2013).
- Santagata, S. *et al.* Taxonomy of breast cancer based on normal cell phenotype predicts outcome. *J Clin Invest* **124**, 858–870 (2014).
- Honeth, G. *et al.* Models of Breast Morphogenesis Based on Localization of Stem Cells in the Developing Mammary Lobule. *Stem Cell Reports* **4**, 699–711 (2015).
- Fridriksdottir, A. J. *et al.* Proof of region-specific multipotent progenitors in human breast epithelia. *Proc Natl Acad Sci USA* **114**, E10102–E10111 (2017).
- Colacino, J. A. *et al.* Heterogeneity of Human Breast Stem and Progenitor Cells as Revealed by Transcriptional Profiling. *Stem Cell Reports* **10**, 1596–1609 (2018).
- Arendt, L. M. *et al.* Anatomical localization of progenitor cells in human breast tissue reveals enrichment of uncommitted cells within immature lobules. *Breast Cancer Res.* **16**, 453 (2014).
- Garbe, J. C. *et al.* Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia. *Cancer Res.* **72**, 3687–3701 (2012).
- Britschgi, A. *et al.* The Hippo kinases LATS1 and 2 control human breast cell fate via crosstalk with ER α . *Nature* **541**, 541–545 (2017).
- Kim, J. & Villadsen, R. Expression of Luminal Progenitor Marker CD117 in the Human Breast Gland. *J. Histochem. Cytochem.* **66**, 879–888 (2018).
- Honeth, G. *et al.* Aldehyde dehydrogenase and estrogen receptor define a hierarchy of cellular differentiation in the normal human mammary epithelium. *Breast Cancer Res.* **16**, R52 (2014).
- Balk-Møller, E. *et al.* A marker of endocrine receptor-positive cells, CEACAM6, is shared by two major classes of breast cancer: Luminal and HER2-enriched. *Am. J. Pathol.* **184**, 1198–1208 (2014).
- 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).
- Stingl, J., Eaves, C. J., Kuusk, U. & Emsman, J. T. Phenotypic and functional characterization *in vitro* of a multipotent epithelial cell present in the normal adult human breast. *Differentiation.* **63**, 201–213 (1998).
- Wuidart, A. *et al.* Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat. Cell Biol.* **20**, 666–676 (2018).
- Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).
- Nolan, E. *et al.* RANK ligand as a potential target for breast cancer prevention in BRCA1-mutation carriers. *Nat. Med.* **22**, 933–939 (2016).
- Clayton, H., Titley, I. & Vivanco, M. Growth and differentiation of progenitor/stem cells derived from the human mammary gland. *Exp. Cell Res.* **297**, 444–460 (2004).
- Dontu, G. *et al.* *In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270 (2003).
- Morsing, M. *et al.* Evidence of two distinct functionally specialized fibroblast lineages in breast stroma. *Breast Cancer Res.* **18**, 108 (2016).
- Tabár, L. *et al.* A Proposal to Unify the Classification of Breast and Prostate Cancers Based on the Anatomic Site of Cancer Origin and on Long-term Patient Outcome. *Breast Cancer Basic Clin. Res.* **8**, 15–38 (2014).
- Kuchenbaecker, K. B. *et al.* Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* **317**, 2402 (2017).
- Reis-Filho, J. S. & Tutt, A. N. J. Triple negative tumours: a critical review. *Histopathology* **52**, 108–118 (2007).

28. Hassanein, M. *et al.* Prediction of BRCA1 germ-line mutation status in patients with breast cancer using histoprognosis grade, MS110, Lys27H3, vimentin, and Ki67. *Pathobiology* **80**, 219–27 (2013).
29. Rønnev-Jessen, L. & Petersen, O. W. Induction of alpha-smooth muscle actin by transforming growth factor-beta 1 in quiescent human breast gland fibroblasts. Implications for myofibroblast generation in breast neoplasia. *Lab. Invest.* **68**, 696–707 (1993).
30. Hsu, J.-D., Yao, C.-C., Han, L.-W. & Han, C.-P. CAM5.2 Is Not Identical to Cytokeratins 8 and 18. *Am. J. Clin. Pathol.* **133**, 514–514 (2010).
31. Rønnev-Jessen, L., Petersen, O. W. & Bissell, M. J. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol. Rev.* **76**, 69–125 (1996).
32. Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**, 2213–23 (2011).
33. Lim, E. *et al.* Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21 (2010).

Acknowledgements

We thank Lena Kristensen, Tove Marianne Lund and Anita Sharma Friismose for expert technical assistance. Benedikte Thuesen and Trine Foged Henriksen, Capió CFR Hospitaler are acknowledged for providing breast biopsy material. The Core Facility for Integrated Microscopy (University of Copenhagen) is acknowledged for confocal microscope accessibility. This work was supported by Novo Nordisk Fonden and Danish Research Council grant 10-092798 (to DanStem), Toyota-Fonden Denmark and Anita og Tage Therkelsens Fond (to R.V.), Familien Erichsens Mindefond and Vera og Carl Johan Michaelsens Legat (to J.K.), Harboefonden, Else og Mogens Wedell-wedellborgs Fond and Danish Cancer Society Grant R146-A9257 (to L.R.-J.).

Author Contributions

R.V. and O.W.P. conceived the project. R.V., O.W.P., O.G.I. and J.K. designed the study. O.G.I., J.K., R.V., A.J.F. and M.M. performed the experiments. O.W.P., R.V., J.K., L.R.-J., O.G.I., A.J.F., V.T.-W. and M.M. analyzed the data and wrote the manuscript. J.K. performed the statistical calculations.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-50903-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019