

RESEARCH ARTICLE

Open Access

# Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland

Mona Shehata<sup>1\*</sup>, Andrew Teschendorff<sup>2</sup>, Gemma Sharp<sup>1</sup>, Nikola Novcic<sup>1</sup>, I Alasdair Russell<sup>1</sup>, Stefanie Avril<sup>1,3</sup>, Michael Prater<sup>1</sup>, Peter Eirew<sup>4</sup>, Carlos Caldas<sup>5,6,7,8</sup>, Christine J Watson<sup>9</sup> and John Stingl<sup>1\*</sup>

## Abstract

**Introduction:** The organisation of the mammary epithelial hierarchy is poorly understood. Our hypothesis is that the luminal cell compartment is more complex than initially described, and that an understanding of the developmental relationships within this lineage will help in understanding the cellular context in which breast tumours occur.

**Methods:** We used fluorescence-activated cell sorting along with *in vitro* and *in vivo* functional assays to examine the growth and differentiation properties of distinct subsets of human and mouse mammary epithelial cells. We also examined how loss of steroid hormones influenced these populations *in vivo*. Gene expression profiles were also obtained for all the purified cell populations and correlated to those obtained from breast tumours.

**Results:** The luminal cell compartment of the mouse mammary gland can be resolved into nonclonogenic oestrogen receptor-positive (ER<sup>+</sup>) luminal cells, ER<sup>+</sup> luminal progenitors and oestrogen receptor-negative (ER<sup>-</sup>) luminal progenitors. The ER<sup>+</sup> luminal progenitors are unique in regard to cell survival, as they are relatively insensitive to loss of oestrogen and progesterone when compared with the other types of mammary epithelial cells. Analysis of normal human breast tissue reveals a similar hierarchical organisation composed of nonclonogenic luminal cells, and relatively differentiated (EpCAM<sup>+</sup>CD49f<sup>+</sup>ALDH<sup>-</sup>) and undifferentiated (EpCAM<sup>+</sup>CD49f<sup>+</sup>ALDH<sup>+</sup>) luminal progenitors. In addition, approximately one-quarter of human breast samples examined contained an additional population that had a distinct luminal progenitor phenotype, characterised by low expression of ERBB3 and low proliferative potential. Parent-progeny relationship experiments demonstrated that all luminal progenitor populations in both species are highly plastic and, at low frequencies, can generate progeny representing all mammary cell types. The ER<sup>-</sup> luminal progenitors in the mouse and the ALDH<sup>+</sup> luminal progenitors in the human appear to be analogous populations since they both have gene signatures that are associated with alveolar differentiation and resemble those obtained from basal-like breast tumours.

**Conclusion:** The luminal cell compartment in the mammary epithelium is more heterogeneous than initially perceived since progenitors of varying levels of luminal cell differentiation and proliferative capacities can be identified. An understanding of these cells will be essential for understanding the origins and the cellular context of human breast tumours.

## Introduction

Human breast tumours are very heterogeneous, with approximately five molecular subtypes recognised; these molecular subtypes are categorised as Luminal A, Luminal B, claudin<sup>low</sup>, basal-like and Her2 [1-3]. Currently unknown is whether breast tumours have a common

cell of origin or whether different types of tumours originate from different types of cells, or whether it is a combination of these two processes [4]. Support for the second hypothesis comes from studies in which different populations of human breast epithelial cells were selected from *in vitro* cultures or were purified using flow cytometry and reverse-engineered into tumours of distinct subtypes using lentiviral vectors [5,6]. Similar results have been observed in a mouse model where loss of *Brcal* in different epithelial cell compartments

\* Correspondence: [Mona.Shehata@cancer.org.uk](mailto:Mona.Shehata@cancer.org.uk); [john.stingl@cancer.org.uk](mailto:john.stingl@cancer.org.uk)  
<sup>1</sup>Mammary Stem Cell Laboratory, Cancer Research UK, Cambridge Research Institute, Li Ka-Shing Centre, Robinson Way, Cambridge CB2 0RE, UK  
Full list of author information is available at the end of the article

resulted in tumours with different histologies [7]. An understanding of the properties of the normal mammary epithelial cell hierarchy will thus be important in understanding the cellular context in which human breast tumours occur. Similarly, an understanding of this hierarchy may also give insight into the properties of cancer stem cells and the behaviour of tumours during the emergence of therapeutic resistance.

The mammary epithelium has traditionally been described as a bilineage epithelium composed of luminal cells and basally-positioned myoepithelial cells that are collectively organised in a series of ducts that drain lobuloalveolar structures during lactation. Previous studies have demonstrated that mammary stem cells have features characteristic of basal cells, whereas the bulk of the progenitor cells display predominantly luminal features and have luminal-restricted development potential [8,9], although a recent report has demonstrated that a separate stem/progenitor cell maintains each lineage during adult tissue homeostasis [10]. The luminal cell compartment is heterogeneous since only a subset of these cells expresses oestrogen receptor (ER) [11]. Most of these ER<sup>+</sup> cells are perceived to be relatively mature cells since they are rarely observed to be cycling in adult mammary tissue [12,13]. However, rare proliferating ER<sup>+</sup> cells can be detected in the mouse mammary gland, suggesting the existence of an ER<sup>+</sup> progenitor cell [14]. More recently, fluorescence-activated cell sorting has been used to prospectively isolate luminal progenitors (LPs) from the mouse mammary gland based on differential expression of c-Kit and CD14 or c-Kit and Sca1 [15,16]. This latter study was able to identify a subpopulation of cells with a Sca1<sup>+</sup>c-Kit<sup>+</sup> phenotype that was enriched for ER<sup>+</sup> LP cells [15]. Similarly, an ER<sup>-</sup> LP cell that has a CD24<sup>high</sup>Sca1<sup>-</sup>c-Kit<sup>+</sup> phenotype can also be identified [15,17]. These cells appear to function as alveolar progenitor cells and are characterised by high expression of the transcription factor Elf5 [18].

To further characterise the cells that make up the luminal cell hierarchy in both the human and mouse mammary glands, we used a combination of fluorescence-activated cell sorting, *in vitro* and *in vivo* functional assays and gene expression profiling strategies. Our results describe the prospective isolation and functional characterisation of discrete ER<sup>+</sup> and ER<sup>-</sup> LP populations that are present in both species. Our results also demonstrate that both of these populations are developmentally plastic and display multilineage potential, and that the ER<sup>+</sup> LPs, at least in the mouse, have a relative survival advantage in a low oestrogen/progesterone environment. We also identify a novel breast cell type in the human mammary epithelium that is characterised by low expression of ERBB3.

## Materials and methods

### Dissociation of human and mouse mammary tissue

All primary human material was derived from 11 reduction mammoplasties at Addenbrooke's Hospital, Cambridge, UK, under full informed consent and in accordance with the National Research Ethics Service, Cambridgeshire 2 Research Ethics Committee approval (08/H0308/178) as part of the Adult Breast Stem Cell Study. All tissue donors had no previous history of cancer and were premenopausal (ages 18 to 46). Mammary tissue was dissociated to single cell suspensions as previously described [19].

The number 3 and/or number 4 mammary glands were dissected from 10-week-old to 14-week-old virgin or 20-day pregnant female C57BL6/J, C57BL6/J.CBA-Tg (ACTbEGFP) and FVB mice and were dissociated in DMEM/F12 (with 2.5 mM L-glutamine and 15 mM HEPES; Gibco, Paisley, Renfrewshire, UK) supplemented with 1 mg/ml collagenase (Roche, Burgess Hill, West Sussex, UK), 100 U/ml hyaluronidase (Sigma, Poole, Dorset, UK) and 50 µg/ml gentamicin (Gibco) for 14 to 16 hours at 37°C. The mammary glands were then processed to single cells as previously described [8]. In some experiments, 8-week-old C57BL/6J mice were ovariectomised or sham-operated 3 weeks prior to collection of mammary tissue.

### Flow cytometry

Single cell suspensions of human mammary cells were treated to detect the enzyme activity of aldehyde dehydrogenase (ALDH) using the Aldefluor Kit (StemCell Technologies, Grenoble, Rhône-Alpes, France) as per the manufacturer's instructions. The cells were then preblocked with 10% normal rat serum (Sigma) and incubated with the following primary antibodies (Table S1 in Additional file 1): CD31-PE/Cy7, CD45-PE/Cy7, epithelial cell adhesion molecule (EpCAM)-PE, CD49f-Alexa Fluor (AF) 647 or CD49f-Pacific Blue, ERBB3-biotin, CD44-AF647, MUC1-AF647, and CD24-AF647. Hank's balanced salt solution supplemented with 2% FBS (Gibco) was used as the diluent for all preblock, antibody incubation and washing steps.

Mouse mammary cells were preblocked with 10% normal rat serum and then incubated with the following primary antibodies (Table S1 in Additional file 1): CD31-biotin, CD45-biotin, Ter119-biotin, BP-1-biotin, EpCAM-AF647, CD49f-AF488, or CD49f-Pacific Blue, CD49b-PE and Sca1-PE/Cy7. CD45, Ter119, CD31 and BP-1 were used to deplete contaminating haematopoietic cells, endothelial cells and a proportion of stromal cells, respectively (collectively termed Lin<sup>+</sup> cells). Where required, single cell suspensions of mouse mammary cells were treated to detect the enzyme activity of

ALDH using the Aldefluor Kit as per the manufacturer's instructions, and then cells were stained as above.

Biotin-conjugated antibodies were detected with streptavidin-APC-Cy7 (BioLegend, Bar Hill, Cambridgeshire, UK). Cells were then filtered through a 30 µm cell strainer and incubated with 4',6-diamidino-2-phenylindole (Invitrogen, Paisley, Renfrewshire, UK) or propidium iodide (Sigma). Human cells were sorted using an Influx (Becton Dickinson, Oxford, Oxfordshire, UK) and mouse cells were analysed using an LSRII (Becton Dickinson) and they were sorted on a FACSAria I (Becton Dickinson) or a MoFlo (Beckman Coulter, High Wycombe, Buckinghamshire, UK). The gating cascade is shown in Additional file 2. A cell recovery count was performed after each sort. Single-stained control cells were used to perform compensation manually. Gates were set in reference to negative controls stained with isotype antibodies conjugated to individual fluorochromes or to fluorescence-minus-one controls (omitting one reagent at a time). The ALDH<sup>+</sup> gate was set in reference to control populations incubated with the ALDH inhibitor DEAB in addition to Aldefluor. Flow cytometry data were analysed using FlowJo™ software (Tree Star, Inc., Ashland, OR, USA).

#### **Renal grafting and mouse mammary repopulating unit assays**

All animal work was approved by Cambridge Research Institute Animal Ethics Committee and the Home Office. Renal capsule experiments were carried out on 10-week-old female NOD/SCID IL2Rγ<sup>-/-</sup> (NSG) mice as previously described [19] with the modification that in some experiments the collagen gels were supplemented with 20% growth factor-reduced Matrigel (BD Biosciences, Oxford, Oxfordshire, UK). A sialastic pellet containing 2 mg 17β-oestradiol and 4 mg progesterone was implanted subcutaneously in recipient mice when human cells were being transplanted [20]. In some experiments, the hormone pellets were surgically excised 5 weeks post surgery and the mice mated. To recover renal gels, recipient mice were killed and the retrieved gels were fixed in 4% paraformaldehyde for 1 hour before being processed into paraffin. On occasion, gels were dissociated for 4 to 5 hours at 37°C in Mouse EpiCult-B™ media (StemCell Technologies) supplemented with 5% FBS, 600 U/ml collagenase and 200 U/ml hyaluronidase. After digestion, cells were washed in Hank's balanced salt solution supplemented with 2% FBS, trypsinised for 5 minutes with gentle pipetting and injected into the cleared mammary fat pads of NSG mice as described below.

For the mouse mammary repopulating unit (MRU) assays, donor cells were suspended in 65% Hank's balanced salt solution supplemented with 2% FBS

additionally supplemented with 25% growth factor-reduced Matrigel and 10% trypan blue solution (0.4%; Sigma), such that a 10 µl injection volume contained the desired cell dose. The endogenous mammary epithelium in the inguinal glands of 3-week-old female C57BL6/J or NSG mice was cleared and cells were injected into cleared fat pads as previously described [21]. The mice were mated 3 weeks after surgery and the number 4 glands were removed during pregnancy and fixed in Carnoy's fixative and stained with carmine alum. An outgrowth was scored positive if it contained both lobular and ductal elements. MRU frequencies were calculated using the Extreme Limiting Dilution Analysis tool [22]). In some experiments, mice were kept in a virgin state for a total of 10 weeks and the number 4 glands removed for analysis by flow cytometry or histology.

#### ***In vitro* colony-forming assays**

Flow-sorted human mammary cells were seeded into 60 mm plates with  $2.5 \times 10^5$  irradiated NIH-3T3 feeder cells. The cultures were maintained in Human EpiCult-B (StemCell Technologies) supplemented with 5% FBS (StemCell Technologies) and 50 µg/ml gentamicin for 24 to 48 hours and then the media changed to serum-free conditions and maintained for an additional 10 to 12 days. Flow-sorted mouse cells were cultured in Mouse EpiCult-B and 50 µg/ml gentamicin in the presence of irradiated feeders for 5 to 7 days. At the end of the assays, the colonies were fixed with acetone:methanol (1:1), stained with Giemsa (Fisher Scientific, Cramlington, Northumberland, UK) and enumerated under a microscope.

In some experiments, the sorted cells were seeded within growth factor-reduced Matrigel and cultured in the presence of Human or Mouse EpiCult-B and irradiated feeders for 14 to 21 days. In some experiments, the culture media were changed after 7 days into differentiation media (DMEM/F12 with Glutamax (Gibco) supplemented with 10% FCS, 1 µM dexamethasone, 5 µg/ml insulin and 5 µg/ml prolactin) to induce lactogenic differentiation of the mammary cells, and the cultures were maintained for a further 7 to 14 days. At the end of the assay, the gels were then fixed in 4% paraformaldehyde and embedded in paraffin for sectioning and immunostaining.

#### **Immunofluorescence and immunohistochemistry**

Sorted cells were allowed to adhere to µ-Slide eight-well chamber poly-lysine-coated slides (Ibidi, Uddingston Glasgow, UK) for 15 minutes before fixation in 4% paraformaldehyde. Cells were blocked in 10% normal goat serum (Sigma) for 1 hour and stained with primary antibodies (Table S1 in Additional file 1) overnight at 4°C. Goat anti-mouse or anti-rabbit antibody conjugated to

either AF488 or AF555 (Invitrogen) was used to detect primary antibodies. IgG antibodies at the same concentration as the primaries were used as isotype controls. Slides were stained with 4',6-diamidino-2-phenylindole to visualise the nuclei. Paraffin-embedded renal gels and Matrigel cultures were sectioned at 4  $\mu$ m, deparaffinised and boiled in pH 6.0 citrate buffer. The sections were stained as above. Where required, a Mouse on Mouse (Vector Labs, Peterborough, Cambridgeshire, UK) pre-blocking kit was used as per the manufacturer's instructions.

#### RNA preparation, quantitative RT-PCR analysis

Freshly sorted cells were pelleted and the supernatant removed. RNA was extracted using the PicoPure™ RNA extraction kit (Applied Biosystems, Paisley, Renfrewshire, UK) as per the manufacturer's instructions, and samples were treated with DNase using the RNase-free DNase Set (Qiagen, Crawley, West Sussex, UK). For mouse quantitative RT-PCR analysis, RNA from sorted cells of five independent experiments was collected. For human quantitative RT-PCR analysis, RNA from sorted cells of six different human breast specimens was collected. cDNA was generated using 100 ng RNA and random hexamers in a 20  $\mu$ l reaction using SuperScript III (Invitrogen) according to the manufacturer's instructions. cDNA was diluted 1/10 and 1  $\mu$ l was used in a 10  $\mu$ l volume reaction with 2 $\times$  SYBR Green Fast PCR Master Mix (Applied Biosciences) and 1  $\mu$ l of 5  $\mu$ M forward and reverse primers (Table S2 in Additional file 1) and H<sub>2</sub>O. The real-time PCR reactions for each sample were performed in triplicate with an ABI 7900 Real Time PCR system under the following conditions: 95°C for 20 seconds followed by 40 cycles of 95°C for 1 second and 60°C for 20 seconds, followed by a dissociation run to obtain melt profiles of the amplicons. A no-template control (no cDNA) was used as a control for all primers, also performed in triplicate. Results were analysed with the delta-delta method normalised to two housekeeping genes (Ppia and Rpl13a or UBC and TBP for mouse and human samples, respectively) and compared with a comparator sample (nonclonogenic luminal (NCL) cells).

#### Microarrays

Total RNA was purified from freshly sorted cell populations using the PicoPure™ RNA extraction kit. Up to 250 ng RNA was labelled according to the standard Illumina gene expression array protocols with the Ambion TotalPrep 96 kit (4397949; Ambion, Paisley, Renfrewshire, UK). Biotinylated complementary RNA was quality controlled using Agilent Bioanalyser and quantified by spectrophotometry (Nanodrop, Ringmer, East Sussex, UK), and 750 ng cRNA was hybridised to Illumina Mouse6 or HumanHT12v4 BeadChips and washed,

stained and scanned according to the standard protocol (WGGX DirectHyb Assay Guide 11286331 RevA; Illumina, Saffron Walden, Essex, UK). Arrays were scanned on an Illumina BeadArray scanner, and data were processed using the Bioconductor beadarray package [23]. (Further information can be found in Additional file 3.) All data files can be accessed via the Gene Expression Omnibus [GEO:GSE35399].

#### Correlation of normal cell subpopulations with breast cancer datasets

##### Centroid construction

Centroids of gene expression for each cell subpopulation were built from the union set of top differentially expressed genes between each pair of cell subtypes. To identify differentially expressed genes we first filtered genes according to variability and then used the limma R package to rank them according to differential expression using *B* statistics. The False Discovery Rate was estimated using the *q*-value R package. To avoid skewing the number of centroid genes to specific cell types, we selected the top 250 upregulated and top 250 downregulated genes in each cell-type comparison. All of these passed False Discovery Rate corrected  $P < 0.05$ . Since the three LP subpopulations were more similar to each other than to any of the other cell types, the corresponding centroids were constructed by the union set of the top 100 upregulated and downregulated genes for each of the three pairwise comparisons. A centroid for the whole LP population (including the three LP subpopulations) was also constructed. Finally, for each cell type the centroid was constructed by averaging the expression of each gene across the samples belonging to that cell type.

##### Correlation scores

Having constructed the cell-type-specific centroids, we next assessed their correlations to breast cancer profiles. First, each breast tumour profile was classified into one of the five intrinsic subtypes using the SSP predictor of Hu and colleagues [24] or the claudin<sup>low</sup> subtype assigned by Herschkowitz and colleagues [25]. Second, each tumour gene expression profile was correlated to each of the centroids using a linear regression. Centroid profiles as well as tumour profiles were scaled to unit variance in order to ensure that regression coefficients, which reflect Pearson correlations, are comparable. Since each tumour can be thought of as a potential mixture of transformed cells from the different normal cell subpopulations, we also modelled each tumour profile as an explicit mixture of the normal cell centroids using a multivariate regression framework. In this framework, each regression coefficient represents a partial correlation and reflects the strength of association between the tumour profile and a given cell-type centroid after

taking into account the contributions from the other cell types.

#### **Decision tree classifier**

We used a nearest centroid decision tree classifier to assign to each breast cancer a cell-type subpopulation according to how similar their tumour profile is to each of the cell centroids. To achieve this assignment we used a decision tree. First, each tumour was assigned to either the stromal, basal, luminal or LP centroid using the nearest centroid rule on the Pearson correlation scores. If a sample classified according to the LP type, we then assigned it to a further LP subtype using the nearest centroid rule against the correlation scores computed from the individual LP subtype-specific centroids.

#### **Statistical analysis**

Data presented are the mean of multiple independent experiments and the standard error of the mean. One-way analysis of variance was used to test multiple groups followed by Tukey's post test to test significant differences between pairs of results. Comparisons between just two groups were analysed by Student's *t* test. Significance was set at \* =  $P < 0.05$ , \*\* =  $P < 0.01$  or \*\*\* =  $P < 0.0001$ .

## **Results**

### **Two distinct luminal progenitor cell types exist within the mouse mammary epithelium**

To test the hypothesis that the LP population is a heterogeneous population, we dissociated mammary glands from 10-week-old virgin C57BL6/J females and analysed the liberated cells using flow cytometry to detect EpCAM, and CD49f ( $\alpha_6$ -integrin; Figure 1A and Additional file 2A). We used EpCAM rather than the previously described CD24 [8,9] since the use of EpCAM permits greater resolution of the luminal and basal cell subpopulations. We also used CD49b ( $\alpha_2$ -integrin) instead of CD61 ( $\beta_3$ -integrin) to identify LPs. We observed that CD49b<sup>+</sup> was a more selective marker of LPs than the previously reported CD61 [26] since up to 47% of progenitors are of CD61<sup>-</sup> phenotype (Additional file 4A,B).

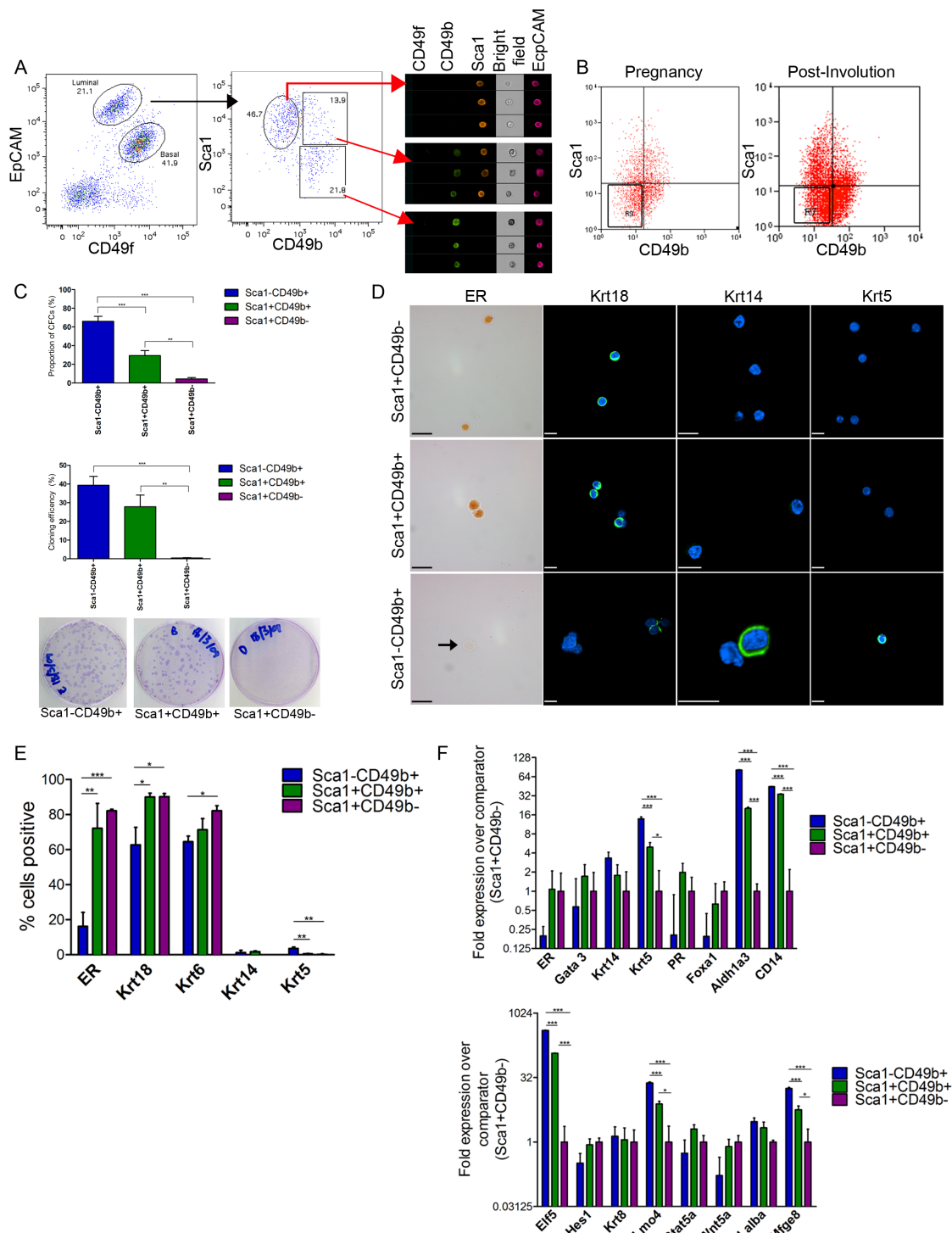
As shown in Figure 1A, the luminal cell compartment can be subdivided into three distinct subpopulations on the basis of expression of CD49b and Sca1. We observed a fourth population (Sca1<sup>-</sup>CD49b<sup>-</sup>) during pregnancy and remnants of this population are maintained throughout the involution and post-involution stages (region gates R9 and R7 in Figure 1B; see Additional file 4C). Flow sorting the three subpopulations from the virgin gland and seeding them into two-dimensional colony-forming cell (CFC) assays reveals that progenitor activity is restricted to the Sca1<sup>-</sup>CD49b<sup>+</sup> and Sca1<sup>+</sup>CD49b<sup>+</sup> subpopulations (Figure 1C). Cloning

efficiency was observed to be approximately 25% and 40% for Sca1<sup>+</sup>CD49b<sup>+</sup> and Sca1<sup>-</sup>CD49b<sup>+</sup> progenitors, respectively. However, we suspect that these cell populations may be pure progenitor cells, since flow-associated toxicity is calculated to be as high as 75% (Additional file 4D).

Previously, gene expression profiling of sorted mouse mammary epithelial cells identified CD14, a co-receptor for bacterial lipopolysaccharide, as being highly enriched in the luminal cell population [8]. When we examined the distribution of CD14 expression among the luminal population, we observe that approximately 86% of LP cells express this protein, whereas CD14<sup>-</sup> luminal cells are relatively deficient in CFCs (Additional file 4E,F,G). Other studies have reported that c-Kit expression identifies both ER<sup>-</sup> and ER<sup>+</sup> LPs [15] and progenitors that are primed to generate progeny that can synthesise milk proteins [16]. When examining c-Kit expression in luminal cells, we observed variation between mouse strains, with c-Kit expression localised to a minority of Sca1<sup>+</sup> luminal cells, and only in FVB mice (Additional file 4H). However, we were unable to observe any significant expression of c-Kit among the luminal cells isolated from C57BL6/J mice, even when using the same c-kit antibody clone (2B8) that was previously used [15]. The c-kit expression levels in FVB mice in our study were quite low when compared with the other studies; one explanation for this discrepancy may be the use of different tissue dissociation protocols.

Immunostaining of sorted populations reveals that both the Sca1<sup>+</sup>CD49b<sup>-</sup> and Sca1<sup>+</sup>CD49b<sup>+</sup> cells express high levels of luminal differentiation markers such as ER and keratin (Krt)18 compared with the Sca1<sup>-</sup>CD49b<sup>+</sup> cells (Figure 1D,E). These results demonstrate that there are two functionally distinct types of ER cells in the mammary gland; the vast majority are relatively mature with little proliferation capacity, but a small population representing ~9% of all luminal cells are ER<sup>+</sup> progenitors. The ER<sup>+</sup> progenitors (Sca1<sup>+</sup>CD49b<sup>+</sup>) express higher transcript levels of luminal differentiation transcripts such as ER, FoxA1 and Gata3 and lower levels of Krt5 and Krt14 when compared with Sca1<sup>-</sup>CD49b<sup>+</sup> cells (Figure 1F). Immunostaining of Sca1<sup>-</sup>CD49b<sup>+</sup> (ER<sup>-</sup>) cells demonstrates that these cells express lower levels of Krt18 and low but detectable levels of the basal cell-specific Krt5 (Figure 1D,E). These cells express no to low levels of ER, which is in agreement with a previously published report [17]. This intermediate level of expression for both luminal and basal cell markers suggests that these cells are a progenitor cell intermediate between the basal stem cells and the more differentiated ER<sup>+</sup> LPs.

The ER<sup>-</sup> progenitor subpopulation also has relatively higher levels of milk protein transcripts in the virgin



**Figure 1 Mouse luminal compartment contains distinct subpopulations.** (A) Distribution of epithelial cell adhesion molecule (EpCAM) and CD49f among Lin<sup>-</sup> cells resolves the luminal and basal cell subpopulations (left). Expression of Sca1 and CD49b among the luminal cells resolves three subpopulations (middle), whose morphology can be visualised using the Image Stream™ analyser (Merck Millipore, Watford, Hertfordshire, UK) (right). (B) Expression of Sca1 and CD49b in pregnant (left; R9 gate) and 6 weeks post involution (right; R7 gate) mammary glands indicating the emergence of a fourth luminal cell population. (C) Bar chart showing the distribution of colony-forming cells (CFCs; top) and cloning efficiencies (middle) of the three luminal populations. Images of CFCs indicating only CD49b<sup>+</sup> cells can form colonies (bottom). (D) Immunocytochemical and immunofluorescence analysis of mouse luminal subpopulations. Cells were sorted, fixed onto slides and stained to detect oestrogen receptor (ER), keratin (Krt)18, Krt14 and Krt5. (E) Bar chart showing the percentage positive cells for each population. (F) Gene expression analysis of Sca1<sup>-</sup>CD49b<sup>+</sup> and Sca1<sup>+</sup>CD49b<sup>+</sup> populations relative to the comparator (Sca1<sup>+</sup>CD49b<sup>-</sup>) for luminal and basal transcripts. Error bars for all indicate the standard error of the mean for six independent experiments. \**P* < 0.05, \*\*\**P* < 0.01, \*\*\*\**P* < 0.0001. Scale bars = 10 μm.

state compared with the ER<sup>+</sup> progenitor cells, including  $\alpha$ -lactalbumin (Lalba) and milk fat globule-epidermal growth factor 8 (Mfg-e8), an observation consistent with previous reports [17]. These cells also express high levels of Elf5 and Lmo4, both of which have been involved with specifying alveolar cell fate (Figure 1F) [18,27]. These results suggest that ER<sup>-</sup> cells probably represent alveolar progenitors and are primed for milk production.

ALDH is an enzyme family previously reported to identify human mammary stem cells [28]. When we examined ALDH among the mammary cell populations, we observed that all of the ER<sup>-</sup> and a subset of the ER<sup>+</sup> LPs show high levels of enzyme activity, whereas the basal cells (EpCAM<sup>lo</sup>CD49<sup>hi</sup>) and NCL (Sca1<sup>+</sup>CD49b<sup>-</sup>) ER<sup>+</sup> cells show low or absent activity (Additional file 4I).

#### **Luminal cell population in the mouse mammary gland is relatively deficient in mammary repopulating units**

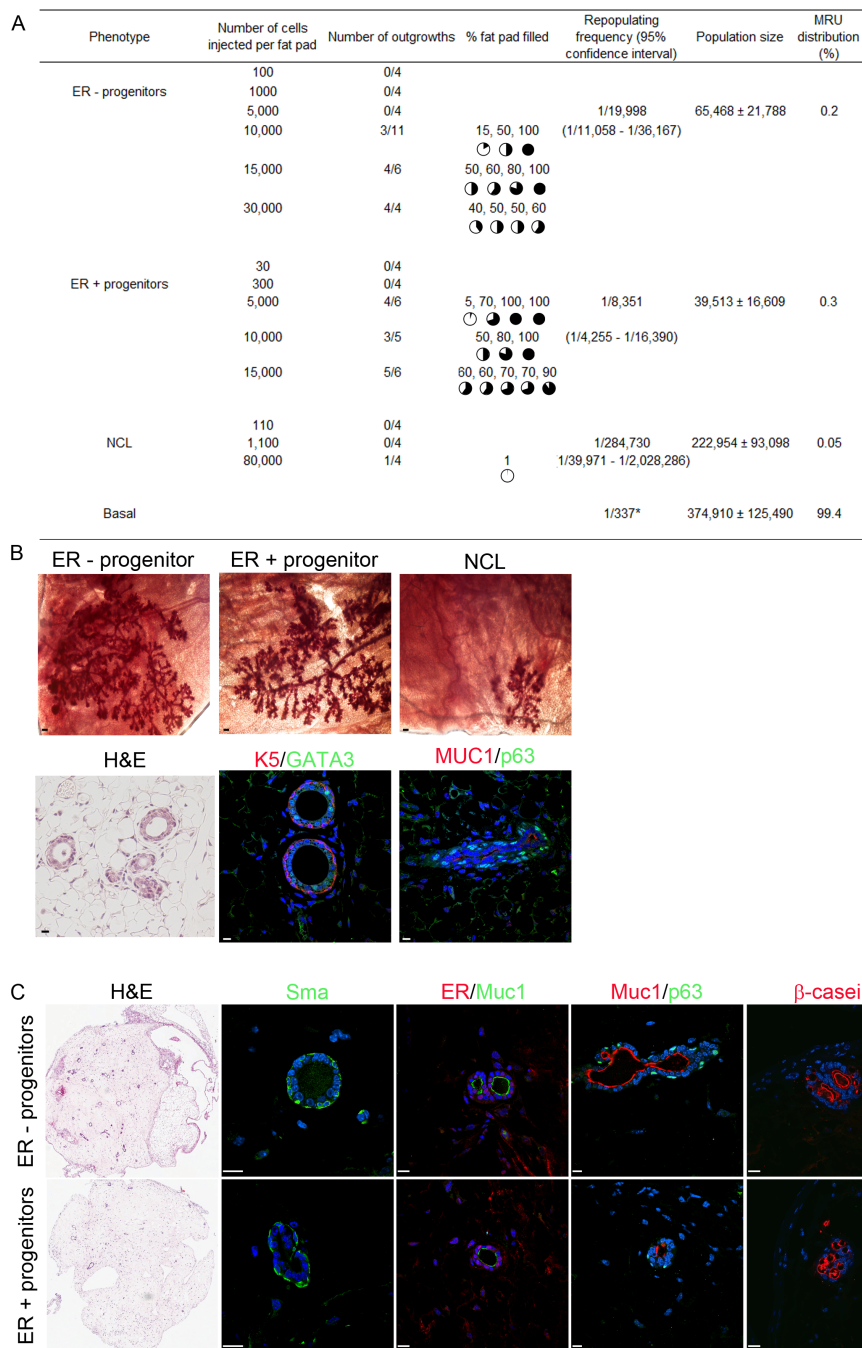
Previous studies have demonstrated that mammary stem cells are localised within the basal cell compartment of the mouse mammary epithelium since MRU-enriched populations have a basal cell-signature [8,9]. However, a recent report has challenged the notion that basal cells are the most potent stem cells since a subpopulation of MRUs expressing high levels of the luminal cell differentiation marker CD24 can be also be detected [29]. To investigate this further, we double sorted the three luminal cell subpopulations to ensure purity and minimise contamination from other cell types (Additional file 5) and transplanted them into 25% Matrigel at limiting dilutions into cleared mammary fat pads of recipient mice. As shown in Figure 2A, MRUs are present within the luminal population, albeit at exceedingly low frequencies. No robust MRUs could be detected in the NCL cells, although occasional small ductal-lobular structures could be detected at very low frequencies (Figure 2A,B). Outgrowths derived from these LPs are morphologically normal and contain all of the subpopulations as those derived from basal cells, although a skewing in the distribution of luminal and basal cell populations towards the basal cells is observed when analysed by flow cytometry (Figure 2B; see Additional file 6A). Secondary transplantations reveal that four of the six primary outgrowths contained >5 MRUs, indicating that some of the luminal MRUs are potent and have extensive self-renewal capacity and can generate normal glands (Additional file 6B,C). However, when the distribution of MRUs among all of the mammary cell populations is calculated, we observe that approximately 99% of all MRUs present in a mouse mammary gland are localised within the basal cell population (Figure 2A).

Van Keymeulen and colleagues previously demonstrated that flow-sorted luminal cells can contribute to the

luminal epithelium upon transplantation into cleared mammary fat pads, but only when co-transplanted with an approximate equivalent number of basal cells [10]. To determine which luminal cell population has this *in vivo* engrafting potential, double-sorted GFP<sup>+</sup> luminal populations (ER<sup>-</sup> progenitors, ER<sup>+</sup> progenitors and NCL cells) were mixed with equal numbers of wild-type total mammary epithelial cells such that the final ratio of GFP<sup>+</sup> marked luminal cells to wildtype basal cells was approximately 2:1 (Additional file 7). These cell mixtures were then transplanted into cleared mammary fat pads of NSG mice. Outgrowths containing GFP<sup>+</sup> cells could be obtained for all mammary luminal cell populations, although, like the transplants described in Figure 2A, the frequency of this event was rare since only 1 in 30,000 to 340,000 sorted GFP<sup>+</sup> luminal cells could engraft (Additional file 7). Unlike the previous report by Van Keymeulen and colleagues, the outgrowths generated in these experiments were not lineage-restricted since the engraftments contained both basal and luminal cells that expressed GFP (Additional file 8). Interestingly, when GFP<sup>+</sup> cells engrafted, no engraftment by the co-injected wildtype basal and luminal cells was observed, even when nonlimiting numbers of basal cells were transplanted (Additional file 7). Further work is required to reconcile these two studies.

#### **Mouse mammary epithelial cells are developmentally plastic**

In an attempt to establish parent-progeny relationships between the different populations described in Figure 1A, we sorted the different luminal cell populations and seeded them into *in vitro* and *in vivo* assays. When ER<sup>-</sup> and ER<sup>+</sup> progenitor cells are seeded into three-dimensional Matrigel cultures and maintained for 3 weeks, we observed that the ER<sup>-</sup> progenitors generated translucent alveolar-like structures that contained eosinophilic material in the lumen, whereas the ER<sup>+</sup> progenitors generated alveolar-like outgrowths that were optically dense without any deposits (Additional file 6D). These results demonstrate that these two types of progenitors are functionally distinct. However, when these colonies were examined for expression of lineage markers such as Muc1 and p63, we observed that both progenitors can generate colonies that contain both luminal and basal cell lineages (Additional file 6D). When double-sorted ER<sup>-</sup> and ER<sup>+</sup> LPs were seeded within collagen/Matrigel gels or in 100% collagen gels and transplanted under the renal capsule of NSG mice, both populations generated outgrowths that contained both luminal (Muc1<sup>+</sup>ER<sup>+</sup>) and myoepithelial (Sma<sup>+</sup>p63<sup>+</sup>) cells in virgin mice and  $\beta$ -casein<sup>+</sup> cells in pregnant mice (Figure 2C; see Additional file 9A). These outgrowths in virgin mice also contained MRUs since cells dissociated from these renal



**Figure 2 Mouse luminal progenitors have multilineage differentiation potential but a rare mammary repopulating unit potential. (A)**

Double-sorted oestrogen receptor-negative (ER<sup>-</sup>) progenitors, oestrogen receptor-positive (ER<sup>+</sup>) progenitors and nonclonogenic luminal (NCL) ER<sup>+</sup> cells were injected at the indicated numbers into cleared fat pads and the mammary repopulating unit (MRU) frequency and distribution in each subpopulation was determined. \*Basal MRU frequency is from Prater MD, Petit V, Russell IA, Girardi R, Menon S, Schulte R, Deugnier M-A, Glukhova MA and Stingl J (manuscript submitted). **(B)** Representative whole mounts of outgrowths derived from the different luminal cell populations (upper panels). Representative primary outgrowths of H & E and immunostained sections to detect luminal (GATA3 and MUC1) and basal (K5 and p63) cells (lower panels). Scale bar = 100 μm and 10 μm, respectively. **(C)** Phenotype of colonies generated when ER<sup>-</sup> and ER<sup>+</sup> progenitors are cultured in Matrigel. Scale bar = 10 μm.



grafts could engraft multiple cleared mammary fat pads (Additional file 9B).

As a further check for high-fidelity sorting, donor cells (GFP<sup>-</sup> or GFP<sup>+</sup>) were mixed with genetically tagged (GFP<sup>+</sup> or GFP<sup>-</sup>) cells (Additional file 9C) and the genotype of the resultant outgrowths was checked by immunohistochemistry for GFP (Additional file 9D). The outgrowths generated were positive for both luminal and basal markers and expressed the appropriate genotype (Additional file 9D,E). These results demonstrate that ER<sup>+</sup> LPs have the potential to dedifferentiate to MRUs.

#### **ER<sup>+</sup> luminal progenitors are relatively insensitive to loss of oestrogen and progesterone**

Both luminal and basal cells, and particularly mammary stem cells, were previously reported to be susceptible to withdrawal of oestrogen and progesterone [30]. To investigate the effects of oestrogen and progesterone withdrawal on the different subtypes of mammary luminal cells, 8-week-old C57BL6/J were ovariectomised and the change in the number of different mammary cell types was determined 3 weeks later.

The NCL cells were acutely sensitive to loss of oestrogen since the size of these populations decreased by 72%, with the proportion of these cells to total epithelial cells decreasing in ovariectomised mice when compared with sham-operated mice (Figure 3A,B,C). The basal, ER<sup>-</sup> progenitor and ER<sup>+</sup> progenitor subpopulations were all mildly sensitive to the withdrawal of oestrogen, decreasing by 46%, 47% and 37%, respectively (Figure 3A,B); no statistically significant effect was observed with any of the luminal populations. When the CFC content from ovariectomised and sham-operated mice was analysed, the ER<sup>-</sup> progenitor population in ovariectomised mice contained 31% fewer progenitors than control mice, and the colonies that were generated were smaller (Figure 3D,E). The number and size of the resultant colonies derived from ER<sup>+</sup> progenitors isolated from ovariectomised mice were marginally smaller when compared with controls, although these differences were not statistically significant. No effect of ovariectomy was observed on any of the non-epithelial (EpCAM<sup>-</sup>CD49f<sup>-</sup> and EpCAM<sup>-</sup>CD49f<sup>+</sup>) cell populations.

These results demonstrate that all populations of mammary epithelial cells are sensitive to loss of steroid hormones, but the ER<sup>+</sup> LPs are only mildly affected and have a survival advantage when compared with the other epithelial cell populations.

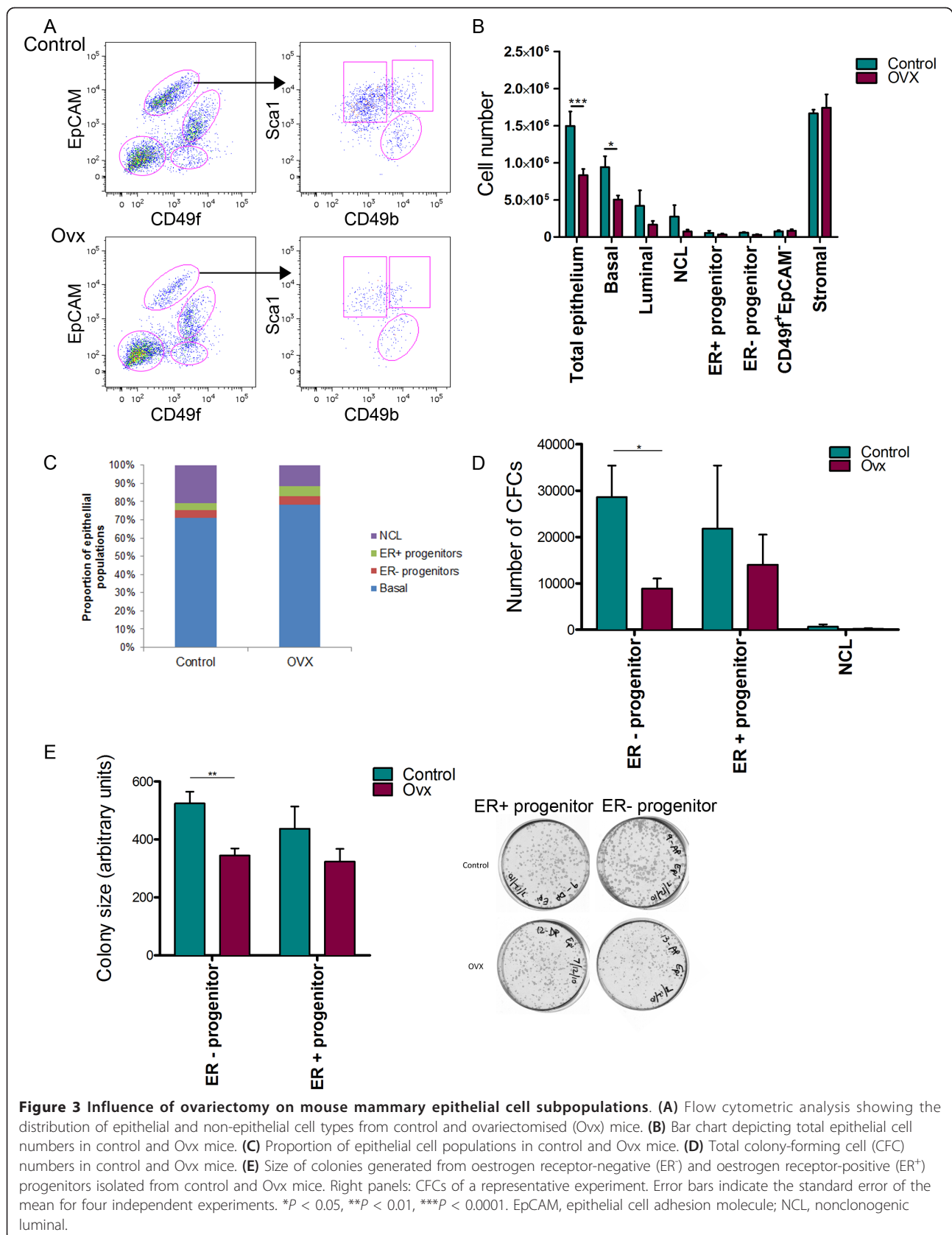
#### **Luminal progenitor compartment in human mammary gland is composed of three distinct cell types**

Previous studies have demonstrated that the luminal compartment in the human mammary epithelium can

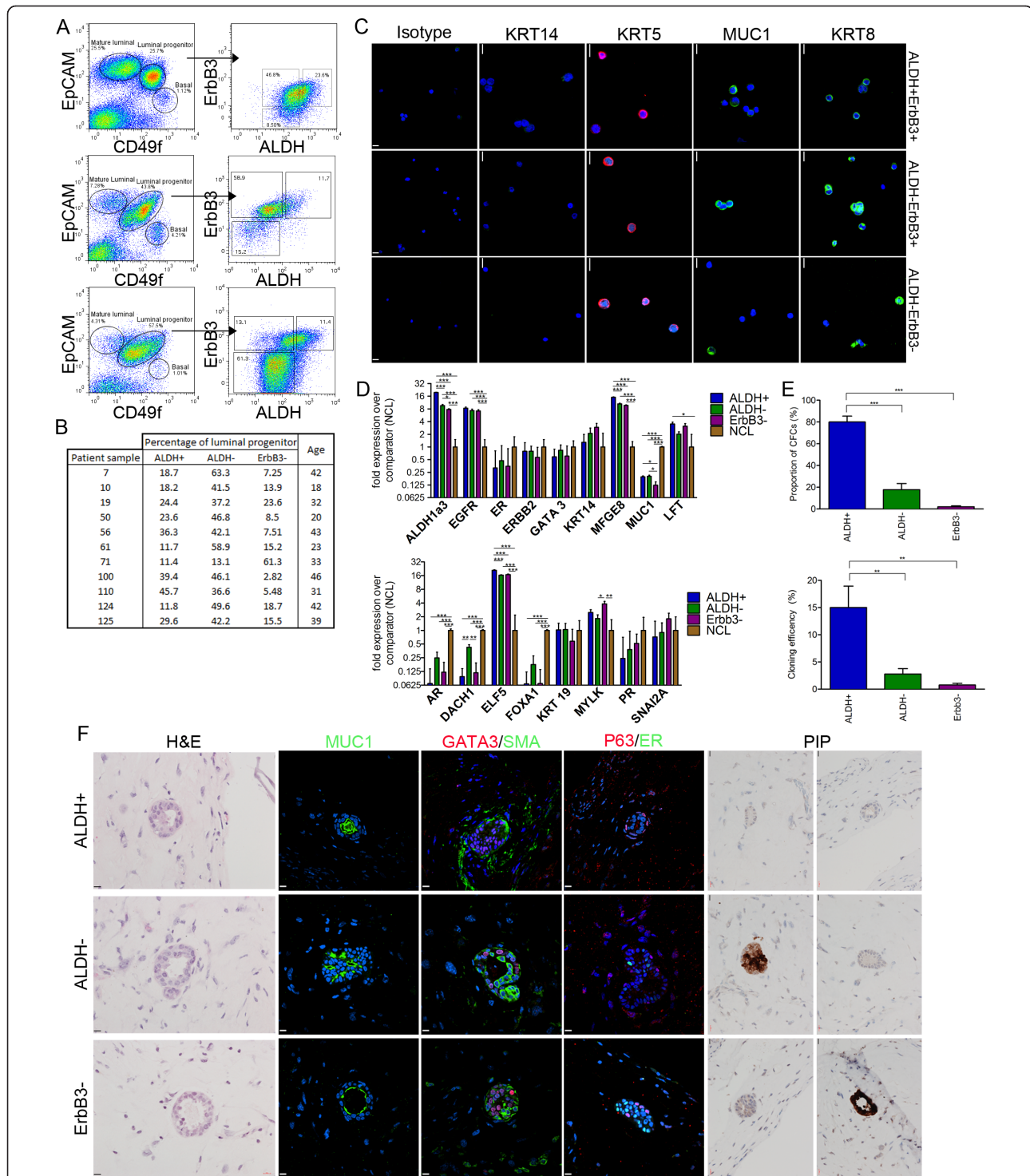
be divided into a luminal-restricted progenitor population (EpCAM<sup>+</sup>CD49f<sup>+</sup>) and mature NCL (EpCAM<sup>+</sup>CD49f<sup>-</sup>) cells that express high levels of ER [31,32]. To test the hypothesis that this LP cell compartment is heterogeneous, as in the mouse, we screened the expression of a variety of markers in 11 reduction mammoplasty samples using flow cytometry. We observed that the differential expression of ALDH and ERBB3 was able to resolve the LP population into not two but three subpopulations of cells: ALDH<sup>+</sup>ERBB3<sup>+</sup> (ALDH<sup>+</sup>), ALDH<sup>-</sup>ERBB3<sup>+</sup> (ALDH<sup>-</sup>) and ALDH<sup>-</sup>ERBB3<sup>-</sup> (ERBB3<sup>-</sup>) (Figure 4A; see Additional file 2B,C - the threshold between ALDH<sup>-</sup> and ALDH<sup>+</sup> is set with reference to a control population incubated in the presence of ALDH inhibitor DEAB, whilst ERBB3 gating was determined using FMO controls). The proportion of these subtypes of cells is very variable between different donors, especially the ERBB3<sup>-</sup> subpopulation, which can range in frequency from 2 to 65% of the total LP population (Figure 4B). However, only one-quarter of all patients have a distinctly identifiable ERBB3<sup>-</sup> population (Figure 4A, middle and lower panels). This population appears to have no correlations with age, although other clinical parameters such as parity, menstrual stage and oral contraceptive use are not known.

All three subpopulations express the luminal-specific KRT8; they also express KRT5 but not KRT14 (Figure 4C). These latter two keratins have historically been considered specific for basal cells, although a recent report has challenged this notion [33]. Gene expression analysis by quantitative RT-PCR confirms that ALDH expression is highest in the ALDH<sup>+</sup> population and lowest in the NCL cells (Figure 4D). When we examine the distribution of the luminal differentiation markers MUC1, AR and FOXA1 among the luminal cell populations, we consistently observe that the ERBB3<sup>-</sup> population displays the lowest levels of luminal differentiation, followed by ALDH<sup>+</sup>, with ALDH<sup>-</sup> and NCL cells displaying the highest levels of luminal differentiation (Figure 4D) - with the exception being that ALDH<sup>+</sup> cells express the highest levels of MUC1 protein (Additional file 10A). Transcripts for ER and GATA3 are not significantly different in the NCL and ALDH<sup>-</sup> subpopulations when compared with the ERBB3<sup>-</sup> and ALDH<sup>+</sup> subpopulations. The ERBB3<sup>-</sup> LPs express the highest levels of the basal-specific genes KRT14, myosin light chain kinase and snail homolog 2a and lower levels of the luminal marker KRT19 when compared with the other LP populations. Interestingly, the ALDH<sup>+</sup> subpopulation expresses the highest levels of transcripts for ELF5, MFG-E8 and LFT (lactoferrin), thereby suggesting that these cells are primed for milk production (Figure 4D).

It has previously been reported that human breast cancer stem cells often have an EpCAM<sup>+</sup>CD44<sup>+</sup>CD24<sup>-/</sup>



**Figure 3 Influence of ovariectomy on mouse mammary epithelial cell subpopulations. (A)** Flow cytometric analysis showing the distribution of epithelial and non-epithelial cell types from control and ovariectomised (Ovx) mice. **(B)** Bar chart depicting total epithelial cell numbers in control and Ovx mice. **(C)** Proportion of epithelial cell populations in control and Ovx mice. **(D)** Total colony-forming cell (CFC) numbers in control and Ovx mice. **(E)** Size of colonies generated from oestrogen receptor-negative (ER<sup>-</sup>) and oestrogen receptor-positive (ER<sup>+</sup>) progenitors isolated from control and Ovx mice. Right panels: CFCs of a representative experiment. Error bars indicate the standard error of the mean for four independent experiments. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.0001. EpCAM, epithelial cell adhesion molecule; NCL, nonclonogenic luminal.



**Figure 4 Three distinct luminal progenitors exist in the human mammary gland. (A)** Left: distribution of CD49f and epithelial cell adhesion molecule (EpCAM) in the Lin<sup>-</sup> population from three patients. Right: distribution of aldehyde dehydrogenase (ALDH) and ERBB3 in the luminal progenitor subset. **(B)** Summary of distribution of luminal progenitor populations and the age of the patients. **(C)** Expression of keratin (KRT)8, MUC1, KRT5 and KRT14 among the three luminal progenitor populations. **(D)** Gene expression analysis of ALDH<sup>+</sup>, ALDH<sup>-</sup> and ERBB3<sup>-</sup> subpopulations relative to the comparator populations (nonclonogenic luminal (NCL)) for oestrogen receptor (ER), KRT14, MFG-E8, epidermal growth factor receptor (EGFR), GATA3, ERBB2, MUC1, ALDH1A3, LFT, AR, DACH1, ELF5, FOXA1, KRT19, myosin light chain kinase (MYLK), PR and snail homolog 2a (SNAI2A). **(E)** Frequency and distribution of colony-forming cells (CFCs) among the three luminal progenitor populations. **(F)** H & E and immunostained sections of xenograft gels derived from ALDH<sup>+</sup>, ALDH<sup>-</sup> and ERBB3<sup>-</sup> progenitors. Shown is the expression of MUC1, GATA3, SMA, p63, ER and prolactin-induced protein (PIP) among outgrowths. All error bars indicate the standard error of the mean from at least five independent samples. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.0001. Scale bars = 10 μm.

<sup>low</sup> phenotype [34]. To understand the normal cellular context of this signature, we used flow sorting to determine the distribution of CD24 and CD44 among the different human mammary epithelial cell populations described in Figure 4A. Our results demonstrate that none of the luminal cell populations have this phenotype and that this phenotype is restricted only to the basal cell population (Additional file 10B,C).

To interrogate the proliferative capacities of the three different LP populations, purified cells from each population were seeded into CFC assays. Results demonstrate that the ALDH<sup>+</sup> subpopulation had the highest cloning efficiencies and contained the highest proportion of CFCs (Figure 4E). Surprisingly the ERBB3<sup>-</sup> population had a very low cloning efficiency and contained an almost undetectable number of progenitor cells, which was unexpected since these cells have a relatively undifferentiated phenotype.

These results demonstrate that the ALDH<sup>+</sup> subpopulation in the human mammary gland is analogous to the ER<sup>-</sup> population in the mouse because both populations contain the highest proportion of progenitors and express high levels of ALDH1a3 and alveolar-associated genes (Figures 1C,F and 4D,E; see Additional file 4F).

#### **Normal human mammary epithelial cells are developmentally plastic**

To further characterise the growth and differentiation potential of the three LPs, we sorted these cells and seeded them into collagen gels that were then transplanted under the renal capsule of female NSG mice. All three subpopulations have the ability to generate hollow acinar multilayered structures (Figure 4F), albeit with vastly different efficiencies since the ERBB3<sup>-</sup> subpopulation generated very few outgrowths. All three LPs gave rise to engraftments that contained both luminal (MUC1<sup>+</sup>GATA3<sup>+</sup>) and basal (p63<sup>+</sup>SMA<sup>+</sup>) cells (Figure 4F). Some engraftments from all populations generated single-layered structures that only contained luminal cells (Additional file 11A). Both ALDH<sup>-</sup> and ERBB3<sup>-</sup> progenitors could generate both ER<sup>+</sup> and ER<sup>-</sup> cells (Figure 4F). The ALDH<sup>+</sup> cells, despite being able to generate GATA3<sup>+</sup> cells, were unable to generate ER<sup>+</sup> cells during the initial 5-week assay. However, ER<sup>+</sup> progeny could be detected when the assay was extended for an additional 3-week period, thereby suggesting that ALDH<sup>+</sup> cells are a primitive progenitor cell that needs additional time to generate all cell lineages (Additional file 11B). A similar pattern of expression was observed when the grafts were examined for expression of prolactin-induced protein, a protein whose expression occurs in the majority of ER<sup>+</sup> breast tumours [35]. Both ALDH<sup>-</sup> and ERBB3<sup>-</sup> progenitors could generate prolactin-induced protein positive

progeny, whereas ALDH<sup>+</sup> cells were unable to do so during the 5-week assay (Figure 4F).

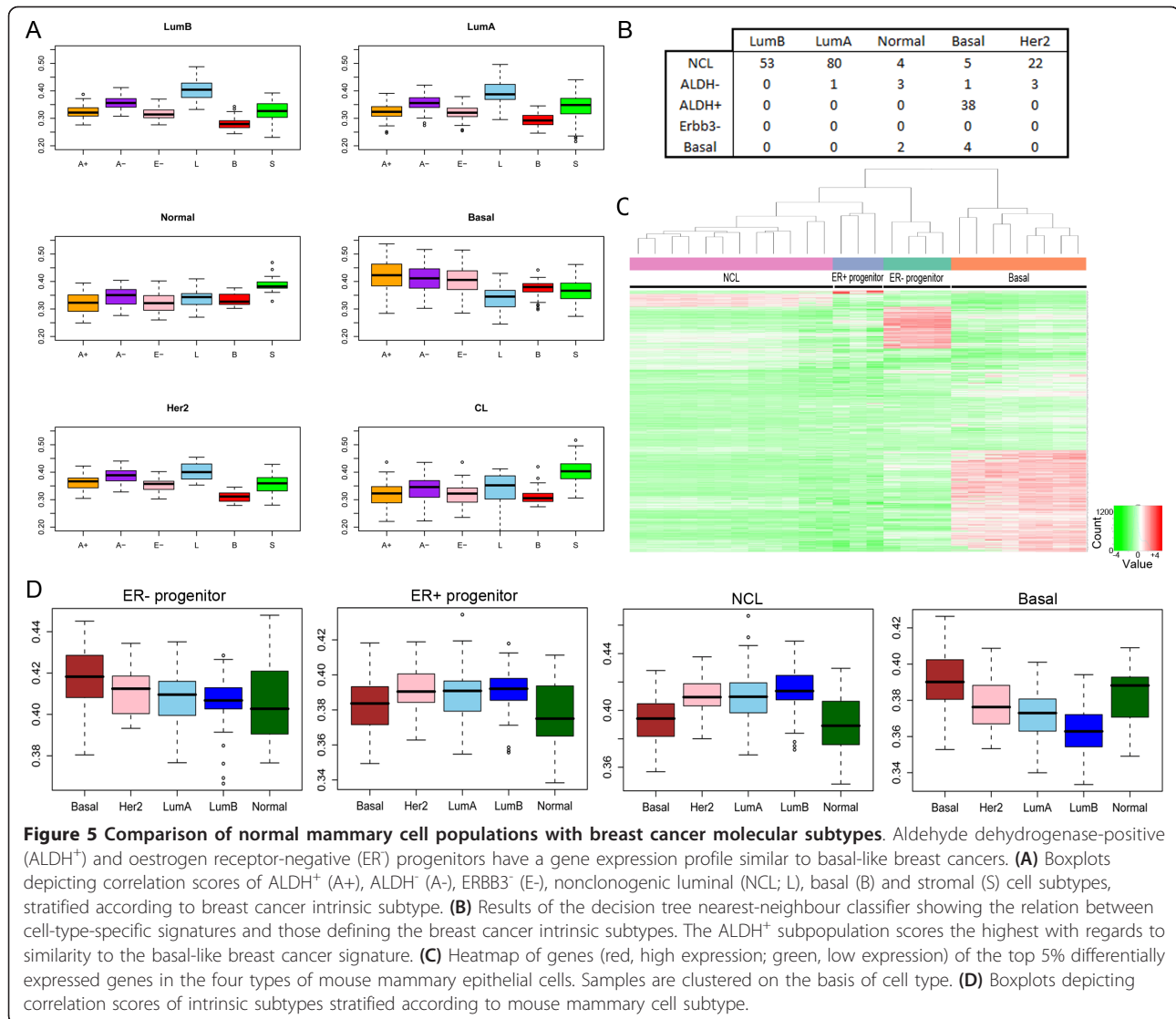
#### **ALDH<sup>+</sup> luminal progenitors have a gene signature similar to that obtained from basal-like breast cancers**

The LP population has previously been shown to have a gene expression signature resembling that of basal-like breast tumours, while the NCL cells resemble Luminal A/B tumours [32]. We hypothesised that subdividing the LP population would identify a closer relationship between the different types of mammary epithelial cells and the different breast cancer subtypes.

To test this hypothesis we sorted six different freshly isolated mammary cell populations (NCL, ALDH<sup>-</sup>, ALDH<sup>+</sup>, ERBB3<sup>-</sup>, basal and stromal) isolated from up to 11 mammaplasty samples and obtained gene expression profiles of these cells (Additional files 12 and 13). As expected, all three LP populations had gene profiles more similar to basal-like breast cancers than the other sorted breast cell populations (Figure 5A; see Additional file 14). Although the gene signature from the ALDH<sup>-</sup> population most strongly correlates with basal-like cancers, it also has some correlations with the Luminal A and B signatures (Figure 5A). When we created a decision tree for the different luminal subpopulations, however, we observed that the gene signature of the ALDH<sup>+</sup> population, and not the ALDH<sup>-</sup> or ERBB3<sup>-</sup> subpopulations, had the highest correlation with those obtained from basal-like breast tumours (Figure 5B). Consistent with what was previously reported, the gene expression signature of the NCL cells resembled those obtained from Luminal A/B breast cancer subtypes and the stromal cells resembled the claudin<sup>low</sup> subtype (Figure 5A) [32]. To obtain a broader picture of the molecular characteristics of the different mouse mammary epithelial cell subpopulations, gene expression profiles of these cells were obtained. Results showed that the gene signatures of each luminal cell population are unique and distinct from basal cells (Figure 5C; see Additional file 15). Similar observations are seen when the microarray expression profiles of the purified mouse mammary cell populations are compared with those obtained from human breast tumours since the ER<sup>-</sup> LPs have a gene expression profile that most resembles that of human basal-like breast tumours and the NCL ER<sup>+</sup> cells resembling Luminal A/B tumours (Figure 5D).

#### **Discussion**

The results presented here demonstrate that the luminal cell compartment in both the human and mouse mammary glands is much more heterogeneous than initially perceived since progenitors of varying levels of luminal cell differentiation can be identified and prospectively isolated. In the mouse, these populations resolve as



separable ER<sup>+</sup> and ER<sup>-</sup> subpopulations, whereas in the human the ALDH<sup>+</sup> and ALDH<sup>-</sup> subpopulations appear to comprise a larger contiguous population. The cell types of the different species appear to be homologous to one another; for example, the ER<sup>-</sup> LPs in the mouse are equivalent to the ALDH<sup>+</sup> cells in the human, and likewise for the ER<sup>+</sup> luminal mouse progenitors and the ALDH<sup>-</sup> luminal human progenitor cells because both populations collectively express higher levels of luminal cell differentiation markers than the ER<sup>-</sup>/ALDH<sup>+</sup> subpopulations. The ER<sup>+</sup> cells in the mouse are probably ductal-restricted progenitors since they express higher levels of ER and FoxA1, transcription factors that have been demonstrated to be essential for ductal but not lobular morphogenesis during mammary gland development [36]. A similar distribution of FOXA1 is also observed in human ALDH<sup>-</sup> cells, thereby suggesting that these

cells function as ductal progenitors in the human mammary gland. Likewise, the ER<sup>-</sup> progenitor cells identified in the mouse mammary gland appear to be alveolar progenitors since they express high levels of Elf5 and Lmo4, transcription factors that specify alveolar cell fate [18,27], as well as milk components including Lalba and Mfg-e8 [37]. In the homologous human population, ALDH<sup>+</sup> cells express high levels of ELF5 and the milk proteins MFGE8 and LFT, which supports the concept that these cells represent a pool of progenitors that are primed to generate alveoli during pregnancy.

Our results demonstrate that there are two types of ER<sup>+</sup> cells in the mammary epithelium; most have little or no proliferative potential and thus are interpreted as being relatively mature, but a small population of ER<sup>+</sup> progenitors can be identified and prospectively isolated. Whether these ER<sup>+</sup> progenitors give rise to the mature

ER<sup>+</sup> luminal cells is not known; *in vivo* lineage tracing experiments using an ER<sup>+</sup> progenitor-specific promoter will need to be performed to establish the developmental relationships between these two cell types.

Results presented herein demonstrate that ER<sup>+</sup> progenitor cells, at least in the normal mouse mammary gland, appear to have a selective advantage over the other mammary epithelial cell populations in adapting to a low-oestrogen environment. ER expression levels in individual mammary epithelial cells are inversely correlated to circulating oestrogen levels [38,39], and thus we hypothesise that the low-oestrogen environment promotes high levels of ER expression and skews the LP pool from an ER<sup>-</sup> state to an ER<sup>+</sup> state. Whether there is a similar preferential survival of ER<sup>+</sup> LP cells in the human breast after menopause is not known, although histological studies comparing premenopausal and postmenopausal women report an increase in the frequency of total ER<sup>+</sup> cells and proliferating ER<sup>+</sup> cells within the postmenopausal mammary gland [40,41]. Garbe and colleagues recently reported that there is an enrichment of LP cells within the mammary epithelium with advancing age. However, these LPs were skewed to a more basal phenotype, which is at odds with our results describing an enrichment of a more luminal type of progenitor [42].

Both oestrogen and progesterone can influence the proliferation of ER<sup>-</sup>/PR<sup>-</sup> cells in the mammary epithelium via paracrine factors such as amphiregulin, Wnt-4 and RANK ligand [30,43-47]. Progesterone can directly promote the proliferation of PR<sup>+</sup> cells via upregulation of cyclin D<sub>1</sub> and cyclin D<sub>2</sub> [47]. Evidence also suggests that oestrogen can directly promote the proliferation of ER<sup>+</sup> cells in the normal mammary gland since recruitment of ER<sup>+</sup> cells into the cell cycle is maximal when circulating oestrogen levels are highest [48]. These ER<sup>+</sup> progenitors are of interest because they represent a potential target cell for malignant transformation. ER<sup>+</sup> cells are typically distributed within the normal mammary epithelium as single cells, but in atypical ductal hyperplasia and ductal carcinoma *in situ* the ER<sup>+</sup> cells tend to be clustered as contiguous islands, suggesting clonal expansion of a mutated ER<sup>+</sup> precursor cell. The frequency of proliferating ER<sup>+</sup> cells in these islands of cells is positively correlated with breast cancer risk, again suggesting that these ER<sup>+</sup> cells are precursors for invasive breast lesions [41].

The hypothesis that postmenopausal breast cancer originates in undifferentiated (ALDH<sup>-</sup>) progenitors does not fit with the fact that these tumours are predominantly of the Luminal A/B subtypes because the ALDH<sup>-</sup> progenitors described herein have highest correlation with basal-like cancers. One possible explanation for

this discrepancy is that tumours may originate in an ALDH<sup>-</sup> progenitor population, but these cells then differentiate to produce mature ER<sup>+</sup> progeny that have a Luminal A/B signature. Another possible explanation is that all of the gene signatures obtained for this study were obtained from premenopausal women (mean age 33.6 years), and that a gene signature of an ALDH<sup>-</sup> cell in the postmenopausal state could be different (for example, more Luminal A/B-like) than those obtained from premenopausal women. More work in determining the role of these cells in breast cancer progression is clearly required.

ERBB3 is a member of the epidermal growth factor receptor family and often forms heterodimers with ERBB2 [49]. ERBB3 is overexpressed in approximately 22% of breast cancer cases [50], and 25% of cases are reported as being ERBB3-negative. The ERBB3<sup>-</sup> LPs in the human mammary gland are somewhat unusual since they have an intermediate phenotype between luminal and basal cells, but appear to be deficient in growth potential. Currently unknown is why only one-quarter of patient samples contain this subpopulation and why, when present, the size of this population is so variable. There is no apparent correlation with the age of the tissue donor in the small sample set analysed here. At the time of collection of these tissue samples no information was available regarding parity history, menstrual cycle status and oral contraception use. Further studies with a much larger well-annotated sample set will be required to gain an insight into the nature of these cells.

Although the function of these ERBB3<sup>-</sup> cells is currently not known, we hypothesise that these cells may be involved with alveologenesis. In the mouse, ErbB3 is required, via phosphoinositide 3-kinase signalling, for the development of the ducts during mammary gland development, but is not essential for the formation of lobules during pregnancy [51,52]. Balko and colleagues reported that ERBB3 expression in the mammary epithelium is highest in the luminal cell populations and lowest in the basal cells, and that loss of ErbB3 in the luminal cell compartment in mice results in an increase in apoptosis of these cells and an expansion of the basal cell population via paracrine signalling [53]. Our results regarding the distribution of ERBB3 among human mammary luminal cells agree with those described by Balko and colleagues, although the presence of a subpopulation of cells with a LP phenotype that are ERBB3<sup>-</sup> appears to be a novel observation.

We used functional assays to establish the growth and differentiation potential of the different types of human and mouse LP cells, and in agreement with the findings of Keller and colleagues [5] we observed that all types of freshly isolated LPs in both species display multilineage

potential when assayed using reconstitution assays, albeit with different frequencies. Other groups have reported that luminal cells in both human and mouse mammary glands can generate multiple lineages of progeny when assayed *in vitro* [17,54-56]. However, these latter observations are not consistent with the results presented here and by Eirew and colleagues [31] since we observe that the majority of cells with *in vivo* engrafting potential are localised within the basal cell compartment. Part of this discrepancy could be due to different groups testing for stemness using different assays. However, the observation that both luminal and basal cell populations exhibit stem cell properties is consistent with a recent report by Van Keymeulen and colleagues. They demonstrate by *in vivo* lineage tracing that both the luminal and basal cell compartments in the mouse mammary gland are maintained by their own stem cell populations during normal tissue homeostasis [10]. This study highlights important caveats in interpreting the results of reconstitution assays, as the differentiation repertoire of cell populations may be perturbed when taken out of a normal tissue environment and purified from other mammary cell types. Reconstitution assays can also mask stem cell potential if cells are transplanted in the absence of appropriate helper cells [10]; this can have obvious limitations when trying to identify putative cancer stem cell populations.

The transplantation process possibly allows LPs to display an expanded differentiation repertoire. Additional work is thus needed to discriminate the growth and differentiation potential that may occur outside normal tissue homeostasis from lineage differentiation that occurs in normal homeostasis. Identifying promoters that are specific for each of the different luminal cell populations will be essential so that the identity of the luminal stem cells and the developmental relationships between the different luminal cells can be established. Such promoters will also be essential for designing transgenic mouse mammary tumour models so that the cell of origin for different molecular types of breast tumours can be established. Current candidates for cell-specific promoters include CD14 for the entire LP population and Elf5 for the ER<sup>-</sup> LP population. Promoters that are specific for the nonclonogenic ER<sup>+</sup> cells and ER<sup>+</sup> LPs have yet to be identified.

ALDH has previously been reported to be a marker of breast stem cells [28]. However, we and others have observed that ALDH is expressed primarily in the LP compartment in both humans and mice [57,58]. The expression pattern of ALDH within the LP compartment as opposed to the basal stem cell containing compartment clarifies a discrepancy regarding the influence of the loss of BRCA1 on human epithelial cell

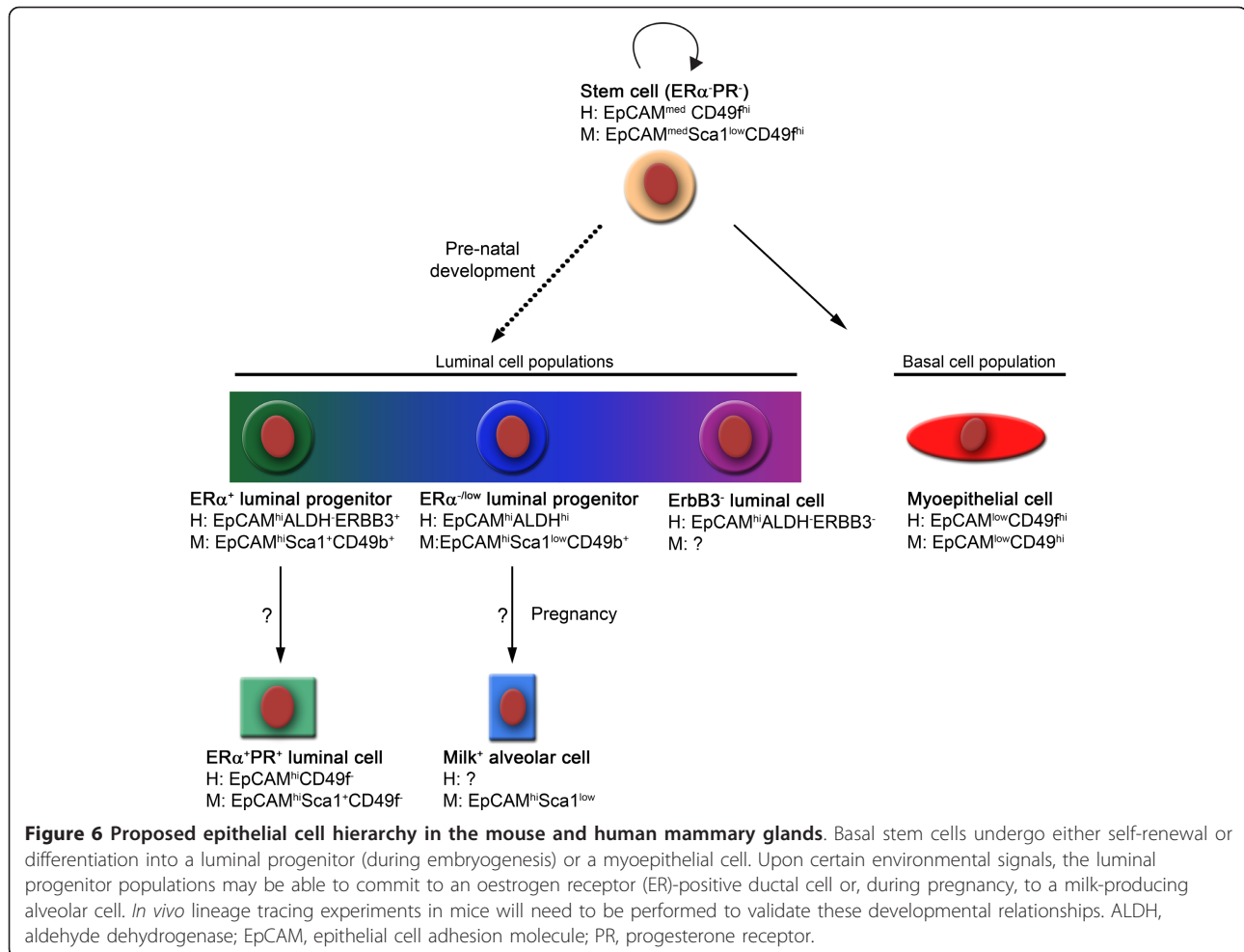
differentiation [59]. Women who have lost an allele of BRCA1 have smaller basal cell populations and expanded LP cell populations [32], whereas forced downregulation of BRCA1 results in the expansion of the ALDH<sup>+</sup> subpopulation, a population initially interpreted as being the stem cell population and distinct from the LP cells [60]. In hindsight, it is now clear that knockdown of BRCA1 results in expansion of the ALDH<sup>+</sup> LP population. In the human epithelium, the ALDH<sup>-</sup> and ALDH<sup>+</sup> subpopulations are clearly part of a larger contiguous population, thereby indicating that these two subpopulations are developmentally tightly linked. We observed that an increase in luminal cell differentiation exists across this population as ALDH expression is lost, and thus one could envision that even low levels of developmental plasticity of an ALDH<sup>-</sup> cell to an ALDH<sup>+</sup> cell could result in loss of luminal cell differentiation.

The results presented here suggest that ALDH1a3 is one of the ALDH isoforms that is being detected by the Aldefluor substrate. Different ALDH isoforms have been shown to be important in different types of cancers, but ALDH1a3 is emerging as a potential cancer stem cell marker in breast cancer [28,61]. ALDH expression, as determined by immunohistochemical staining of tissue sections, has been linked to several breast cancer parameters including ER negativity, high histological grade and general association with basal-like breast cancers [62,63]. Our results using purified subpopulations of human breast epithelial cells are in agreement with these conclusions.

In the context of normal development, therefore, a model of the mammary epithelial cell hierarchy is presented in Figure 6. The basal stem cells undergo either self-renewal or differentiation into a LP (during embryogenesis) or a myoepithelial cell. Environmental signals control commitment of the LP populations to further differentiate to an ER<sup>+</sup> ductal cell or, during pregnancy, to a milk-producing alveolar cell. *In vivo* lineage tracing experiments in mice will need to be performed to validate these developmental relationships.

## Conclusion

The results presented in this manuscript demonstrate that there is more heterogeneity present in the luminal mammary epithelium of both humans and mice than initially perceived. LP cells with gene expression patterns consistent with ductal and alveolar progenitors can be identified and prospectively isolated from both species, with these latter cells having gene expression profiles that strongly resemble those obtained from basal-like breast tumours. The LP compartment in the human mammary epithelium also contains an additional



cell population that is characterised by lack of ERBB3 expression and low proliferative potential. The ER<sup>+</sup> LPs in the mouse are unique in that they are relatively insensitive to loss of oestrogen and progesterone when compared with the other mammary cell populations; this may have implications for the incidence of ER<sup>+</sup> and ER<sup>-</sup> breast cancer in premenopausal women versus post-menopausal women.

## Additional material

**Additional file 1: Table S1 presenting antibodies used for immunostaining; primers used for amplification of p53 isoforms and actin by RT-PCR (nested PCRs), and antibodies used to stain the different cell populations.** Table S2 presenting SYBR primers used for quantitative RT-PCR analysis: mouse and human specific primers used for this study.

**Additional file 2: Figure presenting the gating cascade.** (A) Gating strategy for flow cytometric analysis and sorting for mouse mammary epithelial cells. Cells were gated on forward (FSC) and side (SSC) scatter to remove debris. Then FSC-W/A and SSC-W/A were selected respectively to obtain single cells. 4',6-diamidino-2-phenylindole (DAPI)-positive and lineage-positive cells were excluded. (B) Gating approach for flow cytometric analysis and sorting for human mammary epithelial cells. Cells

were gated on FSC and SSC to remove debris. Then FSC-width and SSC gating were set to exclude doublets. DAPI-positive and lineage-positive cells were excluded. (C) Representative ALDH FACS profiles of total viable cell populations. Cells incubated with ALDH substrate (right) or ALDH and the specific inhibitor DEAB, (left). DEAB controls were used to set the gating strategy to define the ALDH<sup>+</sup> population.

**Additional file 3: MIAME checklist detailing the microarray experimental information.**

**Additional file 4: Figure showing phenotypic characterisation of mouse epithelial cell subpopulations.** (A) Distribution of CD61 among luminal cells. (B) Distribution of CD61 among luminal CFCs showing mean  $\pm$  standard error of the mean. (C) Proportion of the different luminal subpopulations in virgin and post-involution mammary cells. (D) Effects of antibody staining and flow sorting on colony forming efficiencies. (E) Distribution of CD14 among luminal cells. (F) Distribution of CD14 and Sca1 among luminal (CD24<sup>hi</sup>) epithelial cells. (G) Distribution of CD14 and CD24 among luminal CFCs showing mean  $\pm$  standard error of the mean. (H) Distribution of c-Kit among luminal cells in C57BL6/J (upper panel) and FVB mice (lower panel). (I) Flow cytometric analysis showing the distribution of ALDH among subtypes of mouse mammary epithelial cells.

**Additional file 5: Figure showing the gating strategy for double-sorting mouse luminal cell subpopulations.**

**Additional file 6: (A) Flow cytometric analysis to determine the distribution of the different epithelial cell populations generated from transplanted ER<sup>+</sup> progenitors, ER<sup>-</sup> progenitors and basal cells.**



(B) Number of secondary outgrowth derived from primary transplants of ER<sup>-</sup> and ER<sup>+</sup> progenitors. (C) H & E and immunostained sections of secondary outgrowths. Top panel: H & E section of the entire fat pad and a zoomed image of the black square. Lower panel: sections from outgrowths stained to detect Krt14 and Gata3. (E) Morphology of colonies generated when ER<sup>-</sup> and ER<sup>+</sup> progenitors are cultured in Matrigel. Scale bars = 10 μm.

**Additional file 7: Table presenting the MRU frequency and distribution in each GFP<sup>+</sup> subpopulation determined for double-sorted ER<sup>-</sup> progenitors, ER<sup>+</sup> progenitors and NCL cells co-injected with wildtype mammary epithelial cells at the indicated numbers into cleared fat pads.**

**Additional file 8: Figure showing luminal cells have multilineage potential.** Flow analysis to determine the genotype and distribution of the different epithelial cell populations generated from co-transplanting wildtype mammary cells and (A) GFP<sup>+</sup> ER<sup>-</sup> progenitors, (B) GFP<sup>+</sup> ER<sup>+</sup> progenitors and (C) GFP<sup>+</sup> NCL cells. Far right: GFP<sup>+</sup> outgrowths of the initial cell population. Scale bars = 100 μm.

**Additional file 9: Figure showing both ER<sup>-</sup> and ER<sup>+</sup> luminal progenitors can have multilineage potential.** (A) Expression of ER, p63, Krt5 and Muc1 among renal graft outgrowths generated in 100% collagen gels. (B) Representative whole mounts of outgrowths generated from ER<sup>-</sup> and ER<sup>+</sup> luminal progenitors initially propagated as subrenal transplants then transplanted into cleared mammary fat pads. Subrenal grafts were dissociated into single cells and all cells injected into the cleared fat pad of secondary recipient mice. Scale bar = 100 μm. Right: bar chart showing the percentage of the fat pad filled by outgrowths 6 to 8 weeks post transplantation (*n* = 4). (C) Schematic illustration of GFP<sup>+/+</sup> purity check. (D) Representative sections of renal grafts derived from different progenitor types immunostained to detect GFP. (E) Immunofluorescence staining of renal grafts derived from GFP<sup>+</sup> donor cells. Sections stained with antibodies to detect Krt5, ER, p63, Muc1, Sma and Gata3.

**Additional file 10: Figure showing distribution of (A) MUC1, (B) CD24 and (C) CD44 among human mammary epithelial cell subtypes.**

**Additional file 11: Figure showing immunohistochemistry of xenograft gels derived from ALDH<sup>+</sup>, ALDH<sup>-</sup> and ERBB3<sup>-</sup> progenitors for (A) p63 expression and (B) ER expression in engraftments >8 weeks.** Some outgrowths generated from all populations do not contain basal cells. Scale bars = 10 μm.

**Additional file 12: Dataset for microarray centroids of purified human mammary cell populations.**

**Additional file 13: Dataset for microarray centroids of purified human luminal progenitor cells.**

**Additional file 14: Figure showing boxplots depicting correlation scores of ALDH<sup>+</sup> (A+), ALDH<sup>-</sup> (A-), ERBB3<sup>-</sup> (E-), NCL (L), basal (B) and stromal (S) cell subtypes, stratified according to breast cancer intrinsic subtype from another two cancer datasets: (A) Fridlyand and colleagues [64], and (B) Schmidt and colleagues [65].**

**Additional file 15: Dataset for microarray centroids of purified mouse mammary cell populations.**

#### Abbreviations

ALDH: aldehyde dehydrogenase; AF: Alexa Fluor; CFC: colony-forming cell; DMEM: Dulbecco modified Eagle's medium; EpCAM: epithelial cell adhesion molecule; ER: oestrogen receptor; FBS: foetal bovine serum; GFP: green fluorescent protein; H & E: haematoxylin and eosin; Krt: keratin; LP: luminal progenitor; MRU: mammary repopulating unit; NCL: nonclonogenic luminal; NSG: NOD/SCID IL2Rγ<sup>-/-</sup>; PCR: polymerase chain reaction; RT: reverse transcriptase.

#### Acknowledgements

The authors thank R Girardi (Cambridge Research Institute) for useful discussions and S-J Dawson (Cambridge Research Institute) for critical

reviewing of the manuscript. They thank the core facilities at the Cambridge Research Institute for enabling experiments. In particular, M Leeson and R Schulte for assistance with cell sorting; the Biological Resources Unit for animal husbandry; B Wilson and J Miller for sectioning and immunohistochemistry; M Osborne for conducting the microarrays; and R Russell, S Menon and S MacArthur for microarray analysis and discussions. The authors thank L Jones for facilitating and coordinating access to human tissue. They would like to acknowledge the support of The University of Cambridge, The Breast Cancer Campaign, Cancer Research UK, Hutchison Whampoa Limited and NIHR Cambridge Biomedical Research Centre.

#### Author details

<sup>1</sup>Mammary Stem Cell Laboratory, Cancer Research UK, Cambridge Research Institute, Li Ka-Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. <sup>2</sup>Statistical Cancer Genomics, UCL Cancer Institute, University College London, 72 Huntley Street, London WC1E 6BT, UK. <sup>3</sup>Present address: Department of Pathology, Technical University Munich, 21 Arcis Street, Munich 80333, Germany. <sup>4</sup>Department of Molecular Oncology, British Columbia Cancer Research Centre, 675 West 10<sup>th</sup> Avenue, Vancouver V5Z 1L3, Canada. <sup>5</sup>Department of Oncology, University of Cambridge, Worts Causeway, Cambridge CB1 9RN, UK. <sup>6</sup>Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS, Foundation Trust and NIHR Cambridge Biomedical Research Centre, Box 97 Hills Road, Cambridge CB2 2QQ, UK. <sup>7</sup>Cambridge Experimental Cancer Medicine Centre, Cancer Research UK, Cambridge Research Institute, Li Ka-Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. <sup>8</sup>Breast Cancer Functional Genomics Laboratory, Cancer Research UK, Cambridge Research Institute, Li Ka-Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. <sup>9</sup>Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK.

#### Authors' contributions

MS was responsible for conception and design, collection and assembly of data, data analysis and interpretation, and manuscript writing. AT was responsible for data analysis and interpretation. GS, NN, and MP were responsible for collection and assembly of data. IAR was responsible for provision of study material and data interpretation. SA was responsible for conception and design and for data interpretation. PE and CC were responsible for the provision of study material. CJW was responsible for conception and design and for financial support. JS was responsible for conception and design, collection and assembly of data, data analysis and interpretation, and manuscript writing. All authors read and approved the final manuscript.

#### Competing interests

JS is a paid consultant for StemCell Technologies Inc. All remaining authors declare that they have no competing interests.

Received: 29 February 2012 Revised: 25 September 2012  
Accepted: 15 October 2012 Published: 22 October 2012

#### References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
2. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: **Breast cancer classification and prognosis based on gene expression profiles from a population-based study.** *Proc Natl Acad Sci USA* 2003, **100**:10393-10398.
3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**:10869-10874.
4. Stingl J, Caldas C: **Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis.** *Nat Rev Cancer* 2007, **7**:791-799.
5. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, Smith AE, Prat A, Perou CM, Gilmore H, Schnitt S, Naber SP, Garlick JA, Kuperwasser C: **Defining the cellular precursors to human breast cancer.** *Proc Natl Acad Sci USA* 2012, **109**:2772-2777.

6. Ince TA, Richardson AL, Bell GW, Saitoh M, Godar S, Karnoub AE, Iglehart JD, Weinberg RA: **Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes.** *Cancer Cell* 2007, **12**:160-170.
7. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, Mackay A, Grigoriadis A, Tutt A, Ashworth A, Reis-Filho JS, Smalley MJ: **BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells.** *Cell Stem Cell* 2010, **7**:403-417.
8. Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li H, Eaves CJ: **Purification and unique properties of mammary epithelial stem cells.** *Nature* 2006, **439**:993-997.
9. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, Wu L, Lindeman GJ, Visvader JE: **Generation of a functional mammary gland from a single stem cell.** *Nature* 2006, **439**:84-88.
10. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C: **Distinct stem cells contribute to mammary gland development and maintenance.** *Nature* 2011, **479**:189-193.
11. Petersen OW, Hoyer PE, van Deurs B: **Frequency and distribution of estrogen receptor-positive cells in normal, nonlactating human breast tissue.** *Cancer Res* 1987, **47**:5748-5751.
12. Clarke RB, Howell A, Potten CS, Anderson E: **Dissociation between steroid receptor expression and cell proliferation in the human breast.** *Cancer Res* 1997, **57**:4987-4991.
13. Russo J, Ao X, Grill C, Russo IH: **Pattern of distribution of cells positive for estrogen receptor alpha and progesterone receptor in relation to proliferating cells in the mammary gland.** *Breast Cancer Res Treat* 1999, **53**:217-227.
14. Booth BW, Smith GH: **Estrogen receptor-alpha and progesterone receptor are expressed in label-retaining mammary epithelial cells that divide asymmetrically and retain their template DNA strands.** *Breast Cancer Res* 2006, **8**:R49.
15. Regan JL, Kendrick H, Magnay FA, Vafaizadeh V, Groner B, Smalley MJ: **c-Kit is required for growth and survival of the cells of origin of Brca1-mutation-associated breast cancer.** *Oncogene* 2011, **31**:869-83.
16. Asselin-Labat ML, Sutherland KD, Vaillant F, Gyorki DE, Wu D, Holroyd S, Breslin K, Ward T, Shi W, Bath ML, Deb S, Fox SB, Smyth GK, Lindeman GJ, Visvader JE: **Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor caspase-14.** *Mol Cell Biol* 2011, **31**:4609-4622.
17. Sleeman KE, Kendrick H, Robertson D, Isacke CM, Ashworth A, Smalley MJ: **Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland.** *J Cell Biol* 2007, **176**:19-26.
18. Oakes SR, Naylor MJ, Asselin-Labat ML, Blazek KD, Gardiner-Garden M, Hilton HN, Kazlauskas M, Pritchard MA, Chodosh LA, Pfeffer PL, Lindeman GJ, Visvader JE, Ormandy CJ: **The Ets transcription factor Elf5 specifies mammary alveolar cell fate.** *Genes Dev* 2008, **22**:581-586.
19. Eirew P, Stingl J, Eaves CJ: **Quantitation of human mammary epithelial stem cells with in vivo regenerative properties using a subrenal capsule xenotransplantation assay.** *Nat Protoc* 2010, **5**:1945-1956.
20. Laidlaw IJ, Clarke RB, Howell A, Owen AW, Potten CS, Anderson E: **The proliferation of normal human breast tissue implanted into athymic nude mice is stimulated by estrogen but not progesterone.** *Endocrinology* 1995, **136**:164-171.
21. Young LJ: *The Cleared Mammary Fat Pad and the Transplantation of Mammary Gland Morphological Structures and Cells* Kluwer Academic/Plenum Press New York, USA; 2000.
22. Hu Y, Smyth GK: **ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays.** *J Immunol Methods* 2009, **347**:70-78.
23. Dunning MJ, Smith ML, Ritchie ME, Tavare S: **Beadarray: R classes and methods for illumina bead-based data.** *Bioinformatics* 2007, **23**:2183-2184.
24. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, Nobel A, Parker J, Ewend MG, Sawyer LR, Wu J, Liu Y, Nanda R, Tretiakova M, Ruiz Orrico A, Dreher D, Palazzo JP, Perreard L, Nelson E, Mone M, Hansen H, Mullins M, Quackenbush JF, Ellis MJ, Olopade OI, Bernard PS, et al: **The molecular portraits of breast tumors are conserved across microarray platforms.** *BMC Genomics* 2006, **7**:96.
25. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khrantsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM: **Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors.** *Genome Biol* 2007, **8**:R76.
26. Asselin-Labat ML, Shackleton M, Stingl J, Vaillant F, Forrest NC, Eaves CJ, Visvader JE, Lindeman GJ: **Steroid hormone receptor status of mouse mammary stem cells.** *J Natl Cancer Inst* 2006, **98**:1011-1014.
27. Sum EY, Shackleton M, Hahm K, Thomas RM, O'Reilly LA, Wagner KU, Lindeman GJ, Visvader JE: **Loss of the LIM domain protein Lmo4 in the mammary gland during pregnancy impedes lobuloalveolar development.** *Oncogene* 2005, **24**:4820-4828.
28. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, Jacquemier J, Viens P, Kleer CG, Liu S, Schott A, Hayes D, Birnbaum D, Wicha MS, Dontu G: **ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome.** *Cell Stem Cell* 2007, **1**:555-567.
29. Jeselsohn R, Brown NE, Arendt L, Klebba I, Hu MG, Kuperwasser C, Hinds PW: **Cyclin D1 kinase activity is required for the self-renewal of mammary stem and progenitor cells that are targets of MMTV-ErbB2 tumorigenesis.** *Cancer Cell* 2010, **17**:65-76.
30. Asselin-Labat ML, Vaillant F, Sheridan JM, Pal B, Wu D, Simpson ER, Yasuda H, Smyth GK, Martin TJ, Lindeman GJ, Visvader JE: **Control of mammary stem cell function by steroid hormone signalling.** *Nature* 2010, **465**:798-802.
31. Eirew P, Stingl J, Raouf A, Turashvili G, Aparicio S, Emsman JT, Eaves CJ: **A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability.** *Nat Med* 2008, **14**:1384-1389.
32. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, Thorne HJ, Fox SB, Yan M, French JD, Brown MA, Smyth GK, Visvader JE, Lindeman GJ: **Aberant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.** *Nat Med* 2009, **15**:907-913.
33. Gusterson B: **Do 'basal-like' breast cancers really exist?** *Nat Rev Cancer* 2009, **9**:128-134.
34. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF: **Prospective identification of tumorigenic breast cancer cells.** *Proc Natl Acad Sci USA* 2003, **100**:3983-3988.
35. Dumont M, Dauvois S, Simard J, Garcia T, Schachter B, Labrie F: **Antagonism between estrogens and androgens on GCDFFP-15 gene expression in ZR-75-1 cells and correlation between GCDFFP-15 and estrogen as well as progesterone receptor expression in human breast cancer.** *J Steroid Biochem* 1989, **34**:397-402.
36. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, Godwin AK, Korach KS, Visvader JE, Kaestner KH, Abdul-Karim FW, Montano MM, Keri RA: **FOXA1 is an essential determinant of ERα expression and mammary ductal morphogenesis.** *Development* 2010, **137**:2045-2054.
37. Maningat PD, Sen P, Rijnkels M, Sunehag AL, Hadsell DL, Bray M, Haymond MW: **Gene expression in the human mammary epithelium during lactation: the milk fat globule transcriptome.** *Physiol Genomics* 2009, **37**:12-22.
38. Parmar H, Young P, Emsman JT, Neve RM, Dairkee S, Cunha GR: **A novel method for growing human breast epithelium in vivo using mouse and human mammary fibroblasts.** *Endocrinology* 2002, **143**:4886-4896.
39. Shyamala G, Chou YC, Louie SG, Guzman RC, Smith GH, Nandi S: **Cellular expression of estrogen and progesterone receptors in mammary glands: regulation by hormones, development and aging.** *J Steroid Biochem Mol Biol* 2002, **80**:137-148.
40. Lawson JS, Field AS, Champion S, Tran D, Ishikura H, Trichopoulos D: **Low oestrogen receptor alpha expression in normal breast tissue underlies low breast cancer incidence in Japan.** *Lancet* 1999, **354**:1787-1788.
41. Shoker BS, Jarvis C, Clarke RB, Anderson E, Hewlett J, Davies MP, Sibson DR, Sloane JP: **Estrogen receptor-positive proliferating cells in the normal and precancerous breast.** *Am J Pathol* 1999, **155**:1811-1815.
42. Garbe JC, Pepin F, Pelissier FA, Sputova K, Fridriksdottir AJ, Guo DE, Villadsen R, Park M, Petersen OW, Borowsky AD, Stampfer MR, Labarge MA: **Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia.** *Cancer Res* 2012, **72**:3687-3701.
43. Brisken C, Heineman A, Chavarria T, Elenbaas B, Tan J, Dey SK, McMahon JA, McMahon AP, Weinberg RA: **Essential function of Wnt-4 in mammary**

- gland development downstream of progesterone signaling. *Genes Dev* 2000, **14**:650-654.
44. Ciarloni L, Mallepell S, Briskin C: **Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development.** *Proc Natl Acad Sci USA* 2007, **104**:5455-5460.
45. Mulac-Jericevic B, Lydon JP, DeMayo FJ, Conneely OM: **Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform.** *Proc Natl Acad Sci USA* 2003, **100**:9744-9749.
46. Joshi PA, Jackson HW, Beristain AG, Di Grappa MA, Mote PA, Clarke CL, Stingl J, Waterhouse PD, Khokha R: **Progesterone induces adult mammary stem cell expansion.** *Nature* 2010, **465**:803-807.
47. Beleut M, Rajaram RD, Caikovski M, Ayyanan A, Germano D, Choi Y, Schneider P, Briskin C: **Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland.** *Proc Natl Acad Sci USA* 2010, **107**:2989-2994.
48. Zeps N, Bentel JM, Papadimitriou JM, D'Antuono MF, Dawkins HJ: **Estrogen receptor-negative epithelial cells in mouse mammary gland development and growth.** *Differentiation* 1998, **62**:221-226.
49. Yarden Y, Sliwkowski MX: **Untangling the ErbB signalling network.** *Nat Rev Mol Cell Biol* 2001, **2**:127-137.
50. Lemoine NR, Barnes DM, Hollywood DP, Hughes CM, Smith P, Dublin E, Prigent SA, Gullick WJ, Hurst HC: **Expression of the ERBB3 gene product in breast cancer.** *Br J Cancer* 1992, **66**:1116-1121.
51. Jackson-Fisher AJ, Bellinger G, Breindel JL, Tavassoli FA, Booth CJ, Duong JK, Stern DF: **ErbB3 is required for ductal morphogenesis in the mouse mammary gland.** *Breast Cancer Res* 2008, **10**:R96.
52. Lahlou H, Muller T, Sanguin-Gendreau V, Birchmeier C, Muller WJ: **Uncoupling of PI3K from ErbB3 impairs mammary gland development but does not impact on ErbB2-induced mammary tumorigenesis.** *Cancer Res* 2012, **72**:3080-3090.
53. Balko JM, Miller TW, Morrison MM, Hutchinson K, Young C, Rinehart C, Sanchez V, Jee D, Polyak K, Prat A, Perou CM, Arteaga CL, Cook RS: **The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium.** *Proc Natl Acad Sci USA* 2012, **109**:221-226.
54. Pechoux C, Gudjonsson T, Ronnov-Jessen L, Bissell MJ, Petersen OW: **Human mammary luminal epithelial cells contain progenitors to myoepithelial cells.** *Dev Biol* 1999, **206**:88-99.
55. Gudjonsson T, Villadsen R, Nielsen HL, Ronnov-Jessen L, Bissell MJ, Petersen OW: **Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties.** *Genes Dev* 2002, **16**:693-706.
56. Villadsen R, Fridriksdottir AJ, Ronnov-Jessen L, Gudjonsson T, Rank F, LaBarge MA, Bissell MJ, Petersen OW: **Evidence for a stem cell hierarchy in the adult human breast.** *J Cell Biol* 2007, **177**:87-101.
57. Eirew P, Kannan N, Knapp DJ, Vaillant F, Emerman JT, Lindeman GJ, Visvader JE, Eaves CJ: **Aldehyde dehydrogenase activity is a biomarker of primitive normal human mammary luminal cells.** *Stem Cells* 2011, **30**:344-348.
58. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK, Visvader JE: **Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways.** *Breast Cancer Res* 2010, **12**:R21.
59. Ginestier C, Liu S, Wicha MS: **Getting to the root of BRCA1-deficient breast cancer.** *Cell Stem Cell* 2009, **5**:229-230.
60. Liu S, Ginestier C, Charafe-Jauffret E, Foco H, Kleer CG, Merajver SD, Dontu G, Wicha MS: **BRCA1 regulates human mammary stem/progenitor cell fate.** *Proc Natl Acad Sci USA* 2008, **105**:1680-1685.
61. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, Helyer L, Pan L, Leidal A, Gujar S, Giacomantonio CA, Lee PW: **Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis.** *Stem Cells* 2011, **29**:32-45.
62. Ali HR, Dawson SJ, Blows FM, Provenzano E, Pharoah PD, Caldas C: **Cancer stem cell markers in breast cancer: pathological, clinical and prognostic significance.** *Breast Cancer Res* 2011, **13**:R118.
63. Nalwoga H, Arnes JB, Wabinga H, Aklsen LA: **Expression of aldehyde dehydrogenase 1 (ALDH1) is associated with basal-like markers and features of aggressive tumours in African breast cancer.** *Br J Cancer* 2010, **102**:369-375.
64. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, Dairkee S, Tokuyasu T, Ljung BM, Jain AN, McLennan J, Ziegler J, Chin K, Devries S, Feiler H, Gray JW, Waldman F, Pinkel D, Albertson DG: **Breast tumor copy number aberration phenotypes and genomic instability.** *BMC Cancer* 2006, **6**:96.
65. Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kolbl H, Gehrman M: **The humoral immune system has a key prognostic impact in node-negative breast cancer.** *Cancer Res* 2008, **68**:5405-5413.

doi:10.1186/bcr3334

**Cite this article as:** Shehata et al.: Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Research* 2012 **14**:R134.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

