Mammary stem cells have myoepithelial cell properties

Michael D. Prater[1], Valérie Petit[2,3], I. Alasdair Russell[1], Rajshekhar Giraddi[1], Mona Shehata[1], Suraj Menon[1], Reiner Schulte[1], Ivo Kalajzic[4], Nicola Rath[5], Michael F. Olson[5], Daniel Metzger[6], Marisa M. Faraldo[2,3], Marie-Ange Deugnier[2,3], Marina A. Glukhova[2,3] and John Stingl[1].

[1]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK.

[2]Institut Curie, Centre de Recherche, Paris, F-75248 France;

[3]CNRS, UMR144, Paris, F-75248, France;

[4]Reconstructive Sciences, University of Connecticut Health Center
263 Farmington Avenue, Farmington, CT 06030-3705, USA

[5]The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow, G61 1BD, UK

[6]Institut de Génétique et de Biologie Moléculaire et Cellulaire, (CNRS/INSERM/Université de Strasbourg/Collège de France), Illkirch Cedex, France

Correspondence: marina.glukhova@curie.fr (+33(0)156246331) and john.stingl@cruk.cam.ac.uk (+44(0)1223769568)

ABSTRACT

Contractile myoepithelial cells dominate the basal layer of the mammary epithelium and are considered to be differentiated cells. However, we observe that up to 54% of single basal cells can form colonies when seeded into adherent culture in the presence of agents that disrupt acin-myosin interactions, and on average, 65% of the single-cell-derived basal colonies can repopulate a mammary gland when transplanted *in vivo*. This indicates that a high proportion of basal myoepithelial cells can give rise to a mammary repopulating unit (MRU). We demonstrate that myoepithelial cells, flow-sorted using 2 independent myoepithelial-specific reporter strategies, have MRU capacity. Using an inducible lineage tracing approach we follow the progeny of α-smooth muscle actin-expressing myoepithelial cells and show that they function as long-lived lineage-restricted stem cells in the virgin state and during pregnancy.

INTRODUCTION

The mammary epithelium is composed of an outer layer of basal cells that reside on the basement membrane and an inner layer of luminal cells, and collectively these cells are organised as a series of branched ducts that drain alveolar structures during pregnancy. The basal cell layer is composed predominantly of contractile myoepithelial cells, a cell type that has both epithelial and smooth muscle cells features and function to eject milk from the lactating mammary gland[1,2]. Electron microscopy studies have shown that the basal cell layer also contains a subpopulation of small light cells (SLCs) that appear relatively undifferentiated and have been hypothesised to represent mammary stem cells[3,4]. The identity of mammary stem cells is controversial. Previous studies have demonstrated that a rare subset of basal cells have the ability to generate ductal-lobular outgrowths when transplanted into cleared mammary fat pads of recipient mice[5-7]. These engrafting cells are termed mammary repopulating units, or MRUs. However, it has not been possible to resolve these cells from the myoepithelial cells that constitute most of the basal cell compartment. Subsequent studies using a lineage tracing approach have demonstrated that the luminal and basal compartments of the mammary epithelium are maintained by their own lineage-restricted stem cells, and it was suggested that the multilineage potential of the MRUs is an artefact of the transplantation assay[8-10]. However, a more recent report has challenged the existence of lineage-restricted stem cells, and has reported that the mammary epithelium is maintained by a multilineage basal stem cell[11], although these two stem cell models need not be mutually exclusive.

In this manuscript, we demonstrate that a high proportion of basal cells, most of which are myoepithelial cells, can acquire MRU potential when cultured in the presence of a Rho protein kinase inhibitor. We also demonstrate, through the use of two different myoepithelial-reporter transgenic mouse strains, that freshly isolated myoepithelial cells have MRU capacity. The stem cell nature of myoepithelial cells was further confirmed in lineage tracing experiments using *Acta2-Cre-ER^{T2};Rosa26LacZ* mice, which demonstrated that

myoepithelial cell-derived clones expand within the basal cell layer during pubertal development and pregnancy, survive through involution and contribute to the basal layer of ducts and alveoli in multiple pregnancies.

RESULTS

The vast majority of basal cells are non-dividing myoepithelial cells

The basal cell population can be isolated to high purity using flow cytometry based on the differential expression of epithelial cell adhesion molecule (EpCAM) and alpha 6 integrin (CD49f)[12] (Fig. 1a, Supplementary Fig. 1a). The basal population can be subdivided into EpCAM^{high} (upper 20% of the population) and EpCAM^{low} (lower 80%) subpopulations (Fig. 1a), with the former containing a ~5-fold higher frequency of MRUs and ~60% of all MRUs (Fig. 1e). The vast majority of basal cells appear to be myoepithelial cells since ~97% of double-sorted basal cells expressed the myoepithelial marker alpha smooth muscle actin[13] (αSMA) (Fig. 1b). By contrast, only 0.33% (±0.13) of double-sorted luminal cells expressed αSMA (n=4). There was no difference in the proportion of αSMA^{+} cells between basal EpCAM^{high} and EpCAM^{low} cells; nor was there any difference in the level of myoepithelial-associated gene transcripts (*Acta2* and *Myh11*) or in the level of smooth muscle actin protein between these subpopulations (Fig. 1c, d). We hypothesised that dividing cells may express higher levels of EpCAM and thus may be enriched within the basal EpCAM^{high} subpopulation. After injecting mice with a synthetic nucleoside (either 5-iodo-2'-deoxyuridine (IdU) or 5-bromo-2'-deoxyuridine (BrdU)), we observed that the basal EpCAM^{high} cell subpopulation contained a 6-fold higher frequency of nucleoside^{+} cells compared to EpCAM^{low} cells (Fig. 1f). When corrected for population size, the majority (66%) of nucleoside^{+} cells localised within the basal EpCAM^{high} subpopulation (Fig. 1g). When freshly dissociated mammary cells were stained with the DNA binding dye, Hoechst 33342 to identify cells within the S/G2/M phases of the cell cycle (Supplementary Fig. 2), we observed that the basal EpCAM^{high} subpopulation was enriched 8-fold for S/G2/M cells compared to the EpCAM^{low} subpopulation (Supplementary Fig. 3a). The majority (65%) of S/G2/M cells

localised within the basal EpCAM^{high} subpopulation (Supplementary Fig. 3b). Hoechst^{4n} cells were visually confirmed as proliferating cells and not merely cell clumps (Supplementary Fig. 3c). Since the frequency and distribution of MRUs correlated with those of nucleoside^{+} cells within the basal cell population, we hypothesised that basal cell proliferation status may influence their mammary repopulating capacity. To test this, we assessed the MRU and progenitor capacity of freshly isolated G0/G1 and S/G2/M basal cells and observed that the S/G2/M basal cells had a significantly higher MRU frequency and colony forming efficiency (CFE), although when accounting for cell population size, 81% of MRUs were in G0/G1 phase (Supplementary Fig. 3d, e).

Short-term culture increases MRU frequency and number

To determine if basal cells had proliferative capacity *in vitro*, flow-sorted basal cells were seeded at low density in FAD media[14] together with irradiated feeder cells and allowed to form colonies over 7 days. In the absence of the Rho-kinase inhibitor, Y-27632, the colony-forming efficiency (CFE) of basal cells was 5.2±1.5%, but this increased to 29.1±1.5% when Y-26732 was included in the culture media (n=3). The observed CFEs is an underestimate of the true CFE since we observe a ~50% reduction in mammary epithelial cell CFE due to cell toxicity caused by DAPI (4', 6-diamidino-2-phenylindole) and the antibodies used for flow-sorting (Supplementary Fig. 1b). These basal cultures in the presence of Y-26732 could be maintained for 13 passages (~140 cell divisions) before failing to form colonies. Since basal cells showed extensive proliferative capacity *in vitro*, we sought to determine whether MRUs were maintained during culture. To test this, we transplanted non-cultured and 7-day-cultured mouse basal cells at limiting dilutions into cleared fat pads of C57BL/6J mice. Remarkably, there was a ~460-fold expansion in MRU numbers during this 7-day culture period (Fig. 2a). The ductal-lobular outgrowths generated from these 7-day cultured MRUs were morphologically indistinguishable from those generated from freshly isolated basal cells (Fig. 2b).

The majority of single-cell-derived basal colonies can repopulate a mammary gland

To determine whether the rapid MRU expansion that we observed is driven by myoepithelial cells acquiring MRU activity or expansion of existing MRUs, single-basal-cell-derived, 7-day-old colonies were transplanted into cleared fat pads. Mammary glands from C57BL/6J.CBA-Tg(*Actb-EGFP*) mice were used for donor tissue and since transgene-free mice of the same strain were not available, we used NOD/SCID *IL2Rγc*$^{-/-}$ pups as recipients. A single basal cell was flow-sorted into each well of a 96-well dish and each well was visually confirmed to contain a single GFP^{+} cell. It was observed that 54±4% of EpCAM^{high} and 30±5% EpCAM^{low} cells could form a colony after 7 days of culture in the presence of the Rho-kinase inhibitor Y-27632. Individual colonies were trypsinised and transplanted into separate cleared mammary fat pads. Approximately 83% of colonies derived from a basal EpCAM^{high} cell repopulated a mammary gland and 61% of the colonies derived from basal EpCAM^{low} cells engrafted, demonstrating that *de novo* acquisition of MRU potential occurred during culture (Fig. 3a). The engraftments from single-cell-derived basal colonies expressed luminal (Mucin 1) and basal (CK14 and αSMA) markers and produced β-casein during pregnancy (Fig. 3b). In addition, the primary outgrowths were capable of forming secondary engraftments when dissociated and re-transplanted into cleared fat pads, demonstrating that MRU self-renewal had occurred (Supplementary Table 1).

Cytoskeletal remodelling and inhibition of TGFβ significantly influence basal colony formation

In order to understand the molecular changes that might be responsible for MRU expansion, we performed gene expression profiling of non-cultured, 1-day-cultured and 7-day-cultured basal cells. There were ~12,000 differentially expressed genes (DEGs), at FDR<0.01, between non-cultured basal cells compared to 1 or 7-day-cultured basal cells and ~7,000 DEGs between 1-day and 7-day cultured basal cells. Pathway enrichment analysis of the microarray data using MetaCore (GeneGo Inc.) demonstrated that cytoskeletal remodelling and TGFβ pathways were significantly downregulated during culture (Supplementary Fig.

4a). Addition of TGFβ1 protein to FAD media significantly reduced basal cell CFE and this was rescued by adding an inhibitor of the TGFβ receptor, SB 431542[15], to the media (Supplementary Fig. 4b). To investigate the effect of cytoskeletal remodelling on basal cell colony formation, we used small molecule inhibitors to modulate actin dynamics. Latrunculin B and cytochalasin D, which inhibit filamentous (F)-actin polymerisation and increase the free pool of globular (G)-actin monomers[16,17], significantly increased basal cell CFE (Supplementary Fig. 4c). However, at a higher concentration (250 nM), cytochalasin D completely inhibits basal colony formation in the presence of Y-27632 (Supplementary Fig. 4c). Jasplakinolide, which stabilises F-actin[18], significantly reduced basal colony formation in the presence of Y-27632 (Supplementary Fig. 4c). To confirm that Rho kinase inhibition increases basal cell CFE we added a different Rho kinase inhibitor, H1152[19], to FAD media and observed that it significantly increased basal cell CFE to a similar level to that observed with Y-27632 (Supplementary Fig. 4d). Rho kinase inhibition has been shown to reduce apoptosis of dissociated embryonic stem cells by preventing actomyosin contraction[20,21]. To test whether the same mechanism was operating in mammary basal cells, we added a myosin II inhibitor, blebbistatin[22], to FAD media and observed that it significantly increased basal colony formation to a similar level to that obtained with the Rho kinase inhibitors (Supplementary Fig. 4d). The results show that actin cytoskeleton remodelling and downregulation of TGFβ signalling permit a high proportion of basal cells to form colonies.

Myoepithelial cells have mammary repopulating capacity and can undergo clonal expansion *in vivo*

To directly test the proliferative and MRU capacity of freshly isolated, non-cultured myoepithelial cells, we utilised two independent smooth muscle-specific reporter strategies (*Acta2-GFP* and *Myh11-Cre-GFP;Rosa26LacZ*) to isolate and sort a pure population of myoepithelial cells. Both α-smooth muscle actin (encoded by *Acta2*) and smooth muscle myosin (encoded by *Myh11*) are functional markers of myoepithelial cells and enhance contractile force generation during lactation[23]. Sections through the mammary glands of

Acta2-GFP transgenic mice showed colocalisation of GFP and αSMA expression (Fig. 4a). Using flow cytometry, we observed basal αSMA^+ and basal αSMA^- cells (Fig. 4b, Supplementary Fig. 5a-c). These SMA^- basal cells are epithelial in nature since 82±4% of these cells express CK14 or CK5 (Supplementary Fig. 5d). Approximately 30% of basal αSMA^+ cells had colony forming potential, but surprisingly less than 1% of basal αSMA^- cells could form colonies (Fig. 4c). To assess the MRU capacity of myoepithelial cells, basal αSMA^+ cells were transplanted at limiting dilution into cleared fat pads of C57BL/6J mice and ~1% of basal αSMA^+ cells had repopulating capacity (Fig. 4d). Basal αSMA^- cells were transplanted at cell doses proportion to their population size but no engraftments were observed despite transplanting a total of 860 cells. Engraftments derived from basal αSMA^+ cells were dissociated and non-sorted cells were re-transplanted at limiting dilution into cleared fat pads. Secondary engraftments were generated demonstrating that MRU self-renewal had occurred (Supplementary Fig. 5e). We also utilised a *Myh11-Cre-GFP;Rosa26LacZ* reporter system to mark myoepithelial cells and their descendants. Similar to αSMA^+ cells, Myh11+ cells and their progeny did not contribute to the luminal cell layer of mammary ducts and alveoli in these mice (Fig. 4e). Basal GFP^+ and total basal cells were sorted for progenitor and MRU assays (Fig. 4f). The vast majority (90%) of colonies derived from flow-sorted basal cells expressed lacZ, whereas all colonies derived from flow-sorted luminal cells were lacZ-negative (Fig. 4g). There was no difference in the mammosphere forming efficiency of basal GFP^+ and total basal cells (Fig. 4h, i). Furthermore, no statistically significant difference was observed between the MRU frequencies in basal GFP+ and total basal populations (Fig. 4j). Transplanted basal GFP^+ cells gave rise to ductal and alveolar luminal cells during mammary repopulation (Fig. 4k). The self-renewal capacity of these $Myh11^+$ MRUs was demonstrated by transplanting fragments of the primary outgrowths into secondary recipients, since the majority (8 out of 13) of fragments formed secondary engraftments (Supplementary Fig. 5f). These results demonstrate that αSMA^+ and $Myh11^+$ myoepithelial cells have MRU and self-renewal potential.

To determine whether myoepithelial cells proliferate and contribute to growth of mammary epithelium in intact glands *in vivo*, we performed genetic lineage tracing analysis. Mice expressing Cre-recombinase fused to the oestrogen-ligand binding domain ER^{T2} under control of the αSMA-promoter (i.e. strain *Acta2-Cre-ER^{T2}*) were crossed with a *Rosa26LacZ* reporter mouse. Administration of Tamoxifen to prepubertal 4-week-old females induced expression of LacZ in 15-20% of myoepithelial cells. One week after Tamoxifen injection, we detected mostly single LacZ-positive cells, small clones containing 2-4 cells and only few clones containing more than 4 cells (Fig. 5a-c). Remarkably, six weeks after injection, the amount of the clones containing more than 4 cells expanded to 15%, and numerous large clones consisting of 10 and more cells were detected. Furthermore, in glands from 8 day-old pregnant mice, multicellular clones of LacZ-positive cells were found in small lateral branches and alveolar buds (Fig. 5d). Consistent with other publications[10,24], myoepithelial cells and their progeny did not contribute to the luminal cell layer of mammary epithelium.

To further investigate the capacity for clonal expansion and contribution of myoepithelial cells to secretory alveoli during pregnancy, five sexually mature, eight-week-old, *Acta2-Cre-ER^{T2};Rosa26LacZ* mice were treated with Tamoxifen. In two weeks, three mice were mated, and 5-6 weeks after Tamoxifen injection, mammary glands were dissected for analyses at days 12, 15 and 17 of pregnancy (Fig. 5e). X-gal staining of the glands in whole-mount, revealed numerous LacZ-positive cells (Fig. 5f), whereas histological examination confirmed the presence of LacZ-positive cells in the basal layer of the ducts and alveoli. The glands from pregnant animals contained larger clones of LacZ-positive cells than those from virgin mice following the same chase period (Fig. 5f and g), since the frequency of clones containing more than 4 cells was approaching 20% in the pregnant animals, and only 7-8% in the virgin mice (Fig. 5g).

To determine whether myoepithelial cells contributed to the luminal cell layer, sorted basal and luminal cells from *Acta2-Cre-ER^{T2};Rosa26LacZ* mice were x-gal stained in suspension and LacZ-positive and negative cells were counted on cytospots (Supplementary Fig. 6a).

Basal cell populations from virgin and pregnant mice analysed 5-6 weeks after Tamoxifen injection (experimental schedule shown in Fig. 5e) contained from 8 to 15% of LacZ-positive cells, whereas the amount of labelled cells in the luminal population was very low, between 0.17 and 0.65% (Fig. 6a). These observations were confirmed by flow cytometry analysis of *Acta2-Cre-ER^{T2};R26^{mTmG}* mice obtained by mating *Acta2-Cre-ER^{T2}* and double-fluorescent reporter mice *R26^{mTmG}*. In this model, upon Cre-induced recombination, expression of the red fluorescent protein Tomato is replaced by that of GFP[25]. When two 13-week-old virgin, one 13-day-pregnant and one 1-day lactating *Acta2-Cre-ER^{T2};R26^{mTmG}* mice were analysed five weeks after Tamoxifen injection, fewer than 0.1% of the GFP$^+$ cells resided in the luminal compartment, whereas, the amount of GFP$^+$ basal cells was between 31.5% and 49.5% (Figure 6b-d and Supplementary Figure 6b). Altogether, these results strongly indicate that during pregnancy, basal myoepithelial cells expand within the basal compartment of the mammary epithelium only and do not contribute to the luminal layer. Finally, to trace the myoepithelial cells, labelled at puberty through long time periods including mammary gland remodelling during post-lactational involution, we analysed mammary glands of five *Acta2-Cre-ER^{T2};Rosa26LacZ* mice following two complete pregnancy cycles, 20-26 weeks after Tamoxifen injection (Fig. 6a, e-h and Supplementary Figure 6c). Mice were analysed on days 16 and 25 of 2nd involution and days 6-7 and 15 of 3rd pregnancy. Numerous LacZ-positive cells were found in all analysed glands showing that labelled myoepithelial cells were long lived (Fig. 6a, f, h and and Supplementary Figure 6c). In 3rd pregnancy, 23-26 weeks after Tamoxifen injection, the amount of LacZ-positive cells in the basal compartment was approximately 10%, i.e., close to that observed in the 1st pregnancy, after 5-6 weeks of chase (Fig. 6a). Sorted luminal population contained 0.17-0.2% of labelled cells only (Fig. 6a). Histological observations were in line with these data, no labelled luminal cells were revealed (Fig. 6h, right panel).

Altogether, these data show that myoepithelial cell-derived clones can expand in the basal cell layer of both mammary ducts and growing alveolar structures during normal mammary

gland homeostasis, survive through involution and contribute to the basal layer of ducts and alveoli in multiple pregnancies.

DISCUSSION

The basal cell population is perceived to be a mixed population consisting of rare, undifferentiated mammary stem cells interspersed with a population of relatively differentiated myoepithelial cells[5-7,26]. However, MRUs have proven to be exceptionally difficult to prospectively isolate from the myoepithelial cells that make up the bulk of the basal cell compartment by using flow cytometry, nor have dramatic differences in the gene signatures of MRU-enriched and myoepithelial-enriched cell fractions been observed[5]. This suggests that there is a close developmental relationship between MRUs and myoepithelial cells, and the differences between these cells may be subtle. Supporting this concept is the observation that MRU numbers can fluctuate up to 14-fold within the 2 to 3 days between the oestrus and dioestrus phases of the mouse oestrous cycle[27]. Such a large variation in MRU numbers within such a short time period suggests that pre-existing basal cells are being activated to become MRUs, rather than expansion of the MRU subpopulation via cell division. One possible mechanism behind this that myoepithelial cells could function as MRUs when primed for cell division; in support of this hypothesis, inducing basal cell proliferation in our cell culture system permitted a ~460-fold expansion of MRUs after 7 days in culture. Since the majority of single-cell-derived basal colonies contained a MRU, this demonstrates that *de novo* activation of MRUs occurred and that myoepithelial cells contributed to the MRU expansion. It is important to note, however, that culture may activate myoepithelial progenitors rather than differentiated myoepithelial cells, although if this were the case, the myoepithelial progenitor population would have to be quite large to account for the high proportion of basal cells that can become MRUs. A more plausible explanation is that culture itself may reprogram myoepithelial cells to a more primitive state.

We have demonstrated that myoepithelial cells can acquire multilineage MRU potential

during culture. Moreover, we have observed that freshly isolated Myh11$^+$ basal cells have MRU capacity and that all detectable MRUs are αSMA$^+$, again indicating the myoepithelial nature of MRUs. Currently it is not known if all cells of the myoepithelial cell lineage can function as MRU, or just a restricted subset. Furthermore, we demonstrate using a lineage tracing approach that cells of the myoepithelial lineage function as lineage-restricted stem cells in both the virgin state and in pregnancy. One plausible explanation for the discrepancy in the lineage potential of basal cells between lineage tracing studies vs. transplantation assays is that the intercellular and/or cell-extracellular matrix interactions within the intact mammary basal compartment favour the myoepithelial cell fate and prevent asymmetric divisions required for the acquisition of luminal phenotype, whereas disruption of these interactions during tissue dissociation and subsequent transplantation results in loss of the factors that normally inhibit the multipotential cell behaviour of basal cells.

Our lineage tracing data are in line with the results described by Van Keymeulen and colleagues, who reported that that basal compartment of the mammary epithelium harbours uni-potent stem cells restricted to basal myoepithelial lineage[8,10]. However a recent study by Rios et al.[11] provided evidence that mammary basal cells, traced using basal keratin gene promoters or Lgr5, contribute to both the basal and the luminal compartments of mammary ducts and alveoli. Similarly, tracing of Wnt-responsive cells suggested the existence of bipotent mammary stem cells residing in the basal cell layer[9]. Lineage-tracing assays provide valuable data on the hierarchy of mammary epithelial cells, however this experimental approach can be limited by the specific characteristics of distinct promoter constructs. Currently, little is known regarding the heterogeneity of the mammary basal epithelial compartment, and how different promoters may target different cell populations that have distinct developmental potentials. Future experiments will be required to clearly define and characterise the complexity of the mammary stem and progenitor cell populations.

In brief, we show that in culture, a high proportion of myoepithelial cells can give rise to

MRUs and that freshly sorted myoepithelial cells have MRU capacity. We also demonstrate that in intact virgin and pregnant mouse mammary glands, cells of the myoepithelial lineage function as unipotent long-lived stem cells and contribute to the basal cell layer only.

ACKNOWLEDGEMENTS

We thank Amel Saadi for contributing to data interpretation and Aurélie Chiche, Guillaume Carita, Thalia Makdessi and Amandine Di-Cicco for help with lineage tracing experiments. We are grateful to Pierre Chambon and Philippe Soriano for providing mice. We thank Michele Leeson and Loic Tauzin for assistance with cell sorting; James Atkinson and Jodi Miller for sectioning and immunohistochemistry; Michelle Osborne for conducting the microarrays, Jane Gray for help with Western blotting, Silvia Fre and Veronica Rodilla for advice on the cell sorting of *R*26^{mTmG} mouse mammary cells. We thank Fiona Watt and Simon Broad for useful discussions and providing FAD media. We thank Sussan Nourshargh, Michaela Finsterbusch and Jennifer Brown for access to mammary tissue from transgenic mice. This work was funded by Cancer Research UK, Breast Cancer Campaign, the University of Cambridge, Hutchison Whampoa Limited, La Ligue Nationale Contre le Cancer (Equipe Labelisée 2013) and a grant from Agence Nationale de la Recherche ANR-08-BLAN-0078-01 to M.A.G.

AUTHOR CONTRIBUTIONS

M.D.P conducted the experiments and co-wrote the manuscript. V.P., I.A.R., I.K, N.R., R.G. and M.S. conducted experiments. M.M.F. and M.A.D. conducted experiments, analysed data and contributed to data interpretation. M.F.O. provided mice and designed experiments. P.C. and D.M provided mice. S.M. analysed the microarray data. R.S. helped design flow cytometry experiments. M.A.G. and J.S. designed experiments, analysed data and co-wrote the manuscript.

COMPETING FINANCIAL INTERESTS

John Stingl is a paid consultant for StemCell Technologies Inc.

FIGURE LEGENDS

Figure 1. The vast majority of mouse basal cells express alpha-smooth muscle actin (αSMA). High EpCAM expression enriches for mammary stem cells and proliferating cells.

(**a**) Flow cytometry plot showing stromal (black), luminal (blue) and basal (red) cell populations. The basal population has been subdivided into $EpCAM^{high}$ (brightest 20%) and $EpCAM^{low}$ (remaining 80%) subpopulations (**b**) Double-sorted basal $EpCAM^{high}$ and $EpCAM^{low}$ cells stained by immunocytochemistry for αSMA and isotype control inset. Mean (± SEM) of 5 independent experiments. Scale bars = 50 μm. (**c**) Relative mRNA transcript abundance of *Acta2*, *Myh11* and *Oxtr* in basal $EpCAM^{high}$ and $EpCAM^{low}$ cells as detected by real-time PCR. Data normalised to *Actb* and *Rplp0* reference genes. Mean (± SEM) of 4 independent experiments. (**d**) Relative abundance of αSMA protein in basal $EpCAM^{high}$ and $EpCAM^{low}$ cells as detected by Western blot (left). Data normalised to cytokeratin 14 (CK14) abundance. Mean (± SEM) of 3 independent experiments. A representative blot (right) showing protein standards (red), CK14 (55kDa) and αSMA (42kDa) bands (green). (**e**) Mammary repopulating unit (MRU) frequency of sorted basal $EpCAM^{high}$ and $EpCAM^{low}$ cells. Data for Basal $EpCAM^{high}$ and Basal $EpCAM^{low}$ pooled from 5 independent experiments. ** $P = 0.0002$. (**f**) Percentage of flow-sorted basal $EpCAM^{high}$ and $EpCAM^{low}$ cells positive for IdU/BrdU. Data is presented as the mean (± SEM) of 9 independent experiments. * $P = 0.04$. (**g**) Distribution of $IdU/BrdU^{+}$ cells within the basal population.

Figure 2. Short-term culture increases MRU numbers by ~460-fold.

(**a**) MRU frequency in freshly isolated and post-cultured basal cells with estimated number (± SEM) of MRUs per dish. Data pooled from 5 independent experiments for non-cultured basal and from 4 independent experiments for 7-day-cultured basal cells. (**b**) Representative engrafted fat pads derived from non-cultured basal cells and 7-day-cultured basal cells. Scale bar = 2 mm.

Figure 3. A high proportion of single-cell-derived basal colonies contain a MRU.

(**a**) Table showing single-cell cloning efficiency of basal EpCAM^{high} and EpCAM^{low} cells and the proportion of single-cell-derived basal colonies that engrafted when transplanted into cleared mammary fat pads of NSG pups. Cloning efficiencies are presented as the mean± SEM, with data pooled from 4 independent experiments. (**b**) Representative images of a GFP^{+} basal colony and a GFP^{+} engraftment from a transplanted basal colony (which was derived from a single basal EpCAM^{high} cell); scale bars = 500 μm. Images of sections through an engrafted fat pad stained for various markers by immunohistochemistry; scale bars = 100 μm.

Figure 4. Myoepithelial cells have MRU activity.

(**a**) Sections of mammary glands from *Acta2-GFP* mice (C57BL/6J) stained for αSMA and GFP by immunohistochemistry. Representative image seen in 3 independent samples. Scale bar = 100 μm. (**b**) Flow cytometry dot plot showing GFP^{+} (αSMA^{+}) events back-gated onto the EpCAM and CD49f plot. (**c**) CFE of basal αSMA^{+} and basal αSMA^{-} cells. Mean (±SEM) of 4 independent experiments. ** P = 0.0003. (**d**) MRU frequency in basal αSMA^{+} and basal αSMA^{-} cells. Data pooled from 3 independent experiments. ** P = 0.0002. (**e**) Wholemount (left) and sections (center and right) of X-gal-stained mammary glands from *Myh11-Cre-GFP;Rosa26LacZ* virgin (left and center) and 15-day-pregnant (right) mice. Images were observed in 8 (left panel) and 4 (central and right panels) independent samples. Scale bar for left panel = 1.8 mm. Scale bar for central panel = 120 μm. Scale bar for right panel = 50 μm. (**f**) Flow cytometry dot plot showing resolution of total basal and basal Myh11^{+} cells. Five independent samples analysed. (**g**) X-gal-stained colonies derived from flow-sorted basal and luminal cells. Scale bar = 5 mm. (**h**) Mammospheres derived from basal Myh11^{+} and total basal cells. Scale bar = 290 μm. (**i**) Mammosphere forming efficiency of basal Myh11^{+} and total basal cells. Data showing the mean sphere-forming efficiency from one of two independent experiments is presented. Data is derived from 3 technical replicates. (**j**) MRU frequency in basal Myh11^{+} and total basal cells. Data pooled

from 3 independent experiments. (**k**) Wholemount and section of a primary engraftment (above) developed from basal Myh11^{+} cells in virgin (left and central) and 12-day-pregnant (right) hosts and a secondary engraftment (below) in a virgin host. Nineteen primary and three secondary outgrowths were x-gal stained in whole mount, of which three primary and three secondary outgrowths were sectioned. Scale bars for wholemounts = 1.8 mm. Scale bars for sections = 50 μm.

Figure 5. Myoepithelial cell derived clones can undergo expansion within the basal layer of intact mammary glands and survive after multiple pregnancies.
Experimental schedule for (b) and (c). (b) Sections of whole-mount x-gal stained mammary glands from *Acta2-Cre-ER^{T2};Rosa26LacZ* mice injected with Tamoxifen at 4 weeks and dissected 1 and 6 weeks later, as indicated. Scale bar = 52 μm (c) Graph showing the percentage of myoepithelial cell clones containing 1, 2-4 and > 4 LacZ-positive cells in the glands dissected 1 week (2 mice) and 6 weeks (3 mice) after Tamoxifen injection. The values shown are means+SEM from 4 mammary fat pads at each time point. *P < 0.01. For details see On Line Methods. (d) Section of whole-mount x-gal stained mammary gland from *Acta2-Cre-ER^{T2};Rosa26LacZ* mouse injected with Tamoxifen at 4 weeks and dissected 5 weeks later, on day 8 of pregnancy. The arrows point to LacZ-positive cells in the basal layer of small lateral branches and emerging alveolar buds. Scale bar = 90 μm. (e) Experimental schedule for (f) and (g). (f) Fragments (upper panels) and sections (lower panels) of whole-mount x-gal stained mammary glands from 13-week-old virgin, 12- (P12) and 17-day-pregnant (P17) *Acta2-Cre-ER^{T2};Rosa26LacZ* mice dissected 5 and 6 weeks after Tamoxifen injection, as indicated. Arrows point to LacZ-positive alveoli. D, mammary duct, BV, blood vessel. Scale bar = 0.35 mm for upper panels and 65 μm for lower panels. (g) Graph showing the percentage of myoepithelial cell clones containing 1, 2-4 and > 4 LacZ-positive cells in the ducts from virgin and pregnant *Acta2-Cre-ER^{T2};Rosa26LacZ* mice. V1 and V2, two 13-week-old virgin mice; P12, P15 and P17, 1st pregnancy days 12, 15 and 17,

respectively. The values represent means+SEM of 3 counts performed in different gland areas (for details see On Line Methods).

Figure 6. The progeny of myoepithelial cells is restricted to the basal cell layer and survives after multiple pregnancies.

(a) A graph showing the percentage of LacZ-positive cells in basal and luminal cell populations isolated from *Acta2-Cre-ER^{T2};Rosa26LacZ* mouse and x-gal stained in suspension. V, 13-week-old virgin, P15 and P17, 1st pregnancy days 15 and 17, respectively; 3rdP6 and 3rdP15, 3rd pregnancy, days 6-7 and 15, respectively. Experimental schedule for V, P15 and P17 is shown in Figure 5e, for 3rdP6 and 3rdP15, in (g). Data shown for P15, P17 and 3rdP15 were obtained with inguinal mammary glands from one mouse, in each case, for V and 3rdP6, with pooled inguinal glands from two 13-week-old virgin and two 6-7-day-pregnant mice, respectively. (b) A mammary gland fragment from virgin *Acta2-Cre-ER^{T2}*;R26^{mTmG} mouse analyzed five weeks after Tamoxifen injection. Scale bar = 0.35 mm. (c) and (d) Flow cytometry dot plots showing GFP expression in luminal and basal cell populations isolated from virgin (c) and 1-day-lactating (d) *Acta2-Cre-ER^{T2}*;R26^{mTmG} mice. Experimental schedule for (b-d) was same as shown in Figure 5e. Red ovals indicate luminal (L) and basal (B) cell populations. (e) and (g) Experimental schedules for (f) and (h), respectively. (f) Fragment of whole-mount x-gal stained mammary gland from *Acta2-Cre-ER^{T2};Rosa26LacZ* mice injected with Tamoxifen at 4 weeks and analysed 25 weeks later, on day 16 of the involution following 2nd pregnancy (2nd Inv16). Scale bar = 0.87 mm. In (e) and (g), 1st and 2nd P,L,Inv, first and second pregnancy, lactation and involution cycles, respectively. (h) A fragment (left) and a section (right) of x-gal stained mammary glands from *Acta2-Cre-ER^{T2};Rosa26LacZ* mice injected with Tamoxifen at 4 weeks and analysed 23 weeks later, on day 15 of third pregnancy (3rdP15). Scale bar = 0.87 mm, left panel and 50 mm, right panel. Arrows in (f) and (h) point to alveoli. BV, blood vessel; D, mammary duct.

REFERENCES

1 **Franke, W. W. *et al.* Intermediate-sized filaments of the prekeratin type in myoepithelial cells. *The Journal of Cell Biology* 84, 633-654, doi:10.1083/jcb.84.3.633 (1980).**

2 **Lazard, D. *et al.* Expression of smooth muscle-specific proteins in myoepithelium and stromal myofibroblasts of normal and malignant human breast tissue. *Proceedings of the National Academy of Sciences* 90, 999-1003 (1993).**

3 **Chepko, G. *et al.* Differential alteration of stem and other cell populations in ducts and lobules of TGF[alpha] and c-Myc transgenic mouse mammary epithelium. *Tissue and Cell* 37, 393-412 (2005).**

4 **Smith, G. H. & Medina, D. A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *Journal of Cell Science* 90, 173-183 (1988).**

5 **Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* 439, 993-997 (2006).**

6 **Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* 439, 84-88 (2006).**

7 **Sleeman, K. E. *et al.* Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. *J Cell Biol* 176, 19-26, doi:jcb.200604065 [pii] 10.1083/jcb.200604065 (2007).**

8 **Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479, 189-193, doi:10.1038/nature10573 nature10573 [pii] (2011).**

9 **van Amerongen, R., Bowman, A. N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* 11, 387-400, doi:10.1016/j.stem.2012.05.023**

S1934-5909(12)00342-6 [pii] (2012).

10 Taddei, I. *et al.* Beta 1 Integrin deletion from the basal compartment of the mammary epithelium affects stem cells. *Nat Cell Biol* 10, 716-722 (2008).

11 Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. In situ identification of bipotent stem cells in the mammary gland. *Nature* 506, 322-327, doi:10.1038/nature12948

nature12948 [pii] (2014).

12 Shehata, M. *et al.* Phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland. *Breast Cancer Research* 14, R134 (2012).

13 Gottlieb, C., Raju, U. & Greenwald, K. A. Myoepithelial cells in the differential diagnosis of complex benign and malignant breast lesions: an immunohistochemical study. *Mod Pathol* 3, 135-140 (1990).

14 Wu, Y.-J. *et al.* The mesothelial keratins: A new family of cytoskeletal proteins identified in cultured mesothelial cells and nonkeratinizing epithelia. *Cell* 31, 693-703 (1982).

15 Inman, G. J. *et al.* SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor-β Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. *Molecular Pharmacology* 62, 65-74, doi:10.1124/mol.62.1.65 (2002).

16 Flanagan, M. D. & Lin, S. Cytochalasins block actin filament elongation by binding to high affinity sites associated with F-actin. *Journal of Biological Chemistry* 255, 835-838 (1980).

17 Spector, I., Shochet, N., Kashman, Y. & Groweiss, A. Latrunculins: novel marine toxins that disrupt microfilament organization in cultured cells. *Science* 219, 493-495, doi:10.1126/science.6681676 (1983).

18 Bubb, M. R., Senderowicz, A. M., Sausville, E. A., Duncan, K. L. & Korn, E. D. Jasplakinolide, a cytotoxic natural product, induces actin polymerization and

competitively inhibits the binding of phalloidin to F-actin. *Journal of Biological Chemistry* 269, 14869-14871 (1994).

19 Ikenoya, M. *et al.* Inhibition of Rho-kinase-induced myristoylated alanine-rich C kinase substrate (MARCKS) phosphorylation in human neuronal cells by H-1152, a novel and specific Rho-kinase inhibitor. *Journal of Neurochemistry* 81, 9-16, doi:10.1046/j.1471-4159.2002.00801.x (2002).

20 Chen, G., Hou, Z., Gulbranson, D. R. & Thomson, J. A. Actin-Myosin Contractility Is Responsible for the Reduced Viability of Dissociated Human Embryonic Stem Cells. *Cell Stem Cell* 7, 240-248 (2010).

21 Ohgushi, M. *et al.* Molecular Pathway and Cell State Responsible for Dissociation-Induced Apoptosis in Human Pluripotent Stem Cells. *Cell Stem Cell* 7, 225-239 (2010).

22 Straight, A. F. *et al.* Dissecting Temporal and Spatial Control of Cytokinesis with a Myosin II Inhibitor. *Science* 299, 1743-1747, doi:10.1126/science.1081412 (2003).

23 Haaksma, C. J., Schwartz, R. J. & Tomasek, J. J. Myoepithelial cell contraction and milk ejection are impaired in mammary glands of mice lacking smooth muscle alpha-actin. *Biol Reprod* 85, 13-21, doi:10.1095/biolreprod.110.090639 biolreprod.110.090639 [pii] (2011).

24 Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479, 189-193, doi:http://www.nature.com/nature/journal/v479/n7372/abs/nature10573.html#supplementary-information (2011).

25 Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* 45, 593-605, doi:10.1002/dvg.20335 (2007).

26 **Chepko, G. & Smith, G. H. Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue and Cell* 29, 239-253, doi:10.1016/s0040-8166(97)80024-9 (1997).**

27 **Joshi, P. A. *et al.* Progesterone induces adult mammary stem cell expansion. *Nature* 465, 803-807 (2010).**

28 **Yokota, T. *et al.* Bone Marrow Lacks a Transplantable Progenitor for Smooth Muscle Type α-Actin–Expressing Cells. *Stem Cells* 24, 13-22, doi:10.1634/stemcells.2004-0346 (2006).**

29 **Wendling, O., Bornert, J. M., Chambon, P. & Metzger, D. Efficient temporally-controlled targeted mutagenesis in smooth muscle cells of the adult mouse. *Genesis* 47, 14-18, doi:10.1002/dvg.20448 (2009).**

30 **Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* 21, 70-71 (1999).**

31 **Young, L. J. T. in *Methods in Mammary Gland Biology and Breast Cancer Research* (eds Margot M. Ip & Bonnie B. Asch) 67-74 (Kluwer/Plenum, 2000).**

32 **Gentleman, R. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5, R80 (2004).**

33 **Cairns, J. M., Dunning, M. J., Ritchie, M. E., Russell, R. & Lynch, A. G. BASH: a tool for managing BeadArray spatial artefacts. *Bioinformatics* 24, 2921-2922, doi:10.1093/bioinformatics/btn557 (2008).**

34 **Dunning, M. J., Smith, M. L., Ritchie, M. E. & Tavaré, S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* 23, 2183-2184, doi:10.1093/bioinformatics/btm311 (2007).**

35 **Smyth, G. K. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Robert Gentleman *et al.*) 397-420 (Springer, 2005).**

36 **Hu, Y. & Smyth, G. K. ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of Immunological Methods* 347, 70-78 (2009).**

METHODS

Mice

The C57BL/6J.CBA-Tg(*Actb-EGFP*) and NOD/SCID *IL2Rγc*$^{-/-}$ mice were a kind gift from Fiona Watt. The *Acta2-GFP* mice were a kind gift from Sussan Nourshargh and Michaela Finsterbusch and were made by Jen-Yue Tsai and Sanai Sato[28]. Establishment of *Acta2-Cre-ER*^{T2} mouse strain has been previously described[29]. *Myh11-Cre-EGFP* mice were purchased from Jackson Laboratories; *Rosa26LacZ* reporter strain, carrying a *loxP-stop-loxP-LacZ* cassette, was kindly provided by Philippe. Soriano[30]. R26^{mTmG} double-fluorescent reporter mouse strain[25] obtained from Jackson ImmunoResearch Laboratories Inc. was kindly provided by Sylvia Fre. All experiments and procedures involving animals were in strict accordance with the French and European legislations for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes and were approved by the Departmental Direction of Populations Protection (approval number: B75-05-18). Husbandry and supply of animals as well as maintenance and care of the animals in exempt of pathogen species environments before and during experiments fully satisfies the animal's needs and welfare. No statistical method was used to predetermine sample size, and experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mammary gland dissociation into single cell suspension

The number 3 and 4 mammary glands were dissected from 10-15-week-old virgin, female C57BL/6J mice were dissociated for 14-16 h at 37°C in 1:1 DMEM/F12 (+ 2.5mM L-glutamine + 15mM HEPES; Gibco) + 1 mg/ml collagenase (Roche) + 100 U/ml hyaluronidase (Sigma) + 50 µg/ml gentamicin (Gibco). The mammary glands were then processed to single cells as previously described[6]. Mammary glands from *Myh11-Cre-GFP*, *Myh11-Cre-GFP;Rosa26LacZ*, *Acta2-Cre-ER*^{T2}*;Rosa26LacZ and*

Acta2-Cre-ER^{T2}*;R*26^{mTmG} mice were dissociated and processed for single cell suspension and flow cytometry as described elsewhere[8-10].

Flow cytometry and immunofluorescence

Mammary cells were incubated with 10% normal rat serum for 10 min on ice to pre-block before antibody staining. All antibody incubations were for 10 min on ice in HF media: Hank's Balanced Salt Solution (Gibco) + 10 mM HEPES (Sigma) + 2% fetal bovine serum (FBS; Gibco). Mouse mammary cells were stained with the following primary antibodies: 1 µg/ml CD31-biotin (clone 390, eBioscience); 1 µg/ml CD45-biotin (clone 30-F11, eBioscience); 1 µg/ml Ter119-biotin (clone Ter119, eBioscience); 1 µg/ml BP-1-biotin (clone 6C3, eBioscience); 1 µg/ml EpCAM-Alexa Fluor 647 (clone G8.8, Biolegend) and 2 µg/ml CD49f-Alexa Fluor 488 (clone GoH3, Biolegend). In experiments using cells from C57BL/6J.CBA-Tg(*Actb-EGFP*) and *Acta2-GFP* mice, 2 µg/ml CD49f-Pacific Blue (clone GoH3, Biolegend) antibody was used instead of CD49f-Alexa Fluor 488. Mouse cells were then stained with 0.4 µg/ml streptavidin-APC-Cy7 (Biolegend). 1 µg/ml DAPI (Invitrogen) was used to detect dead cells. Cells were filtered through a 30 µm cell strainer prior to sorting. For experiments using cells from *Myh11-Cre- EGFP* mice, the following conjugated antibodies were used: anti-CD24-PE-Cy5 (clone M1/69; BD Biosciences), anti-CD49f-PE (clone GoH3; BD Biosciences), anti-CD45-APC (clone 30-F11; Biolegend), anti-CD31-APC (clone MEC13.3; Biolegend). For separation of basal and luminal cells from *Acta2-Cre-ER*^{T2};Rosa26LacZ mice, anti-CD49f antibody was FITC-conjugated. For analysis of *Acta2-Cre-ER*^{T2};R26^{mTmG} mice, anti-CD24-PE (clone M1/69; BD Biosciences) and anti-CD49f-PECy7 (clone GoH3; Biolegend) were used. Sorting of cells was done using a FACSAria I (BD Biosciences) except for Figure 4f and Figures 6c-d, where a FACSVantage (BD Biosciences) was used. Single-stained control cells were used to perform compensation manually. Matched isotype control antibodies conjugated to each fluorochrome were used to control for

background staining and fluorescence-minus-one controls were used to set gates.

Flow cytometry data were analysed using FlowJo software (Tree Star, Inc.).

Instrument configuration:

FACSAria I lasers:

Lasers	Wavelength	Power
UV	375 nm	20 mW
Violet	405 nm	50 mW
Blue	488 nm	80 mW
Red	633 nm	25 mW
561	561 nm	100 mW

FACSAria I detectors:

Name	Dichroic filter	Band pass filter	Wavelength range	Dyes
UV-C	None	424/44	402-446	DAPI, Hoechst 33342
Violet-C	None	450/50	425-475	Pacific Blue
Blue-D	600LP	610/20	600-620	PI (Propidium Iodide)
Blue-E	550LP	575/25	562-588	PE
Blue-F	505LP	530/30	515-545	Alexa Fluor 488
Red-A	755LP	780/60	750-810	APC Cy7
Red-C	None	660/20	650-670	Alexa Fluor 647

LSRII lasers:

Lasers	Wavelength	Power
UV	355 nm	20 mW
Violet	405 nm	25 mW

Blue	488 nm	20 mW
Red	633 nm	25 mW

LSRII detectors:

Name	Dichroic filter	Band pass filter	Wavelength range	Dyes
UV-B	None	450/50	425-475	DAPI
Blue-F	505LP	530/30	515-545	FITC

Quality check:

The laser alignment was checked regularly using BD™ Cytometer Setup & Tracking Beads and 8-peak rainbow beads (BD Biosciences). The PMT settings were set such that the unstained cells were placed in the first decade of fluorescence.

Compensation:

Single stained control cells were used for each dye used. Compensation was performed manually. A representative compensation matrix of the FACSAria I is shown below:

	FSC PMT	UV C	Red C	Red A	Blue F	Blue A	Blue E	Blue B
FSC PMT	100%	0%	0%	0%	0%	0%	0%	0%
UV C	0%	100%	0%	0%	0%	0%	0%	0%
Red C	0%	0%	100%	24%	0%	0%	0%	1.1%
Red A	0%	0%	3%	100%	0%	1%	0%	0%
Blue F	0%	0%	0%	0%	100%	0%	45%	0%
Blue A	0%	0%	0%	11.5%	0%	100%	5.4%	0.9%
Blue E	0%	0%	0%	0%	1%	0.7%	100%	4.5%

Blue B	0%	0%	8%	0%	0%	15%	0%	100%

Gating strategy:

The following isotype controls were used to check for background staining: Rat IgG2a, κ-Alexa Fluor 488 (clone RTK2758, Biolegend); Rat IgG2a,k-Alexa Fluor 647 (clone RTK2758, Biolegend); Armenian Hamster IgG-PE (clone HTK888; Biolegend); Rat IgG2b, κ-PerCP (clone RTK2758, Biolegend). Representative sort plots with typical gates are shown below for the sorting of basal cells. Doublets, dead cells and contaminating haematopoietic, endothelial and stromal cells are gated out. High expression of EpCAM ($EpCAM^{high}$) is defined as the brightest 20% of the basal population.

Proliferation and apoptosis assays

To assess proliferation status in non-cultured cells, freshly dissociated mammary cells were incubated with 10 μg/ml Hoechst 33342 (Sigma) for 45 min at 37°C. Cells were then stained with antibodies as stated above except that 0.5 μg/ml propidium iodide (PI; Sigma) was used instead of DAPI to detect dead cells. The full gating strategy is shown below. $Hoechst^{2n}$ and $Hoechst^{4n}$ cells were visualised on an ImageStream imaging flow cytometer (Amnis). For synthetic nucleoside injections, mice were injected with a 150 mg/kg dose of BrdU or IdU (Sigma), were culled 1-7 h post-injection, their mammary glands removed and dissociated and basal cells isolated and fixed on slides in cold methanol for 5 min at -20°C. The cells were incubated with 2N HCl for 30 min at 25°C to denature the DNA. A primary antibody that recognises both BrdU and IdU (clone B44; BD Biosciences) was used at a 1 μg/ml and incubated overnight at 4°C followed by an anti-rabbit IgG-Alexa Fluor 555 secondary antibody (Invitrogen) at 1 μg/ml for 1 h at 25°C. Approximately 500 basal $EpCAM^{high}$ and $EpCAM^{low}$ cells for each independent sample were scored by eye on a Nikon upright confocal microscope at 60x magnification. To assess proliferation

status post-culture, cells were trypsinised and incubated with 10 μg/ml Hoechst for 45 min at 37°C and were then stained with EpCAM-Alexa Fluor 647 and PI. FlowJo was used for all cell cycle analysis.

Immunocytochemistry and immunofluorescence:

Purified basal cells were allowed to adhere to poly-L-lysine coated slides at 37°C for 15 min. Basal cells and colonies were fixed in 1:1 acetone/methanol and permeabilised with 0.025% Triton X-100 in TBS for 10 min before pre-blocking with 10% normal goat or donkey serum (Sigma) in TBS for 1 h at 25°C. For immunohistochemistry, paraffin-embedded tissue sections were de-waxed in xylene, gradually rehydrated and then boiled in citrate buffer pH6 for 10 min before pre-blocking with 10% normal donkey serum in TBS for 1 h at 25°C or with 1%BSA/0.1% Tween 20/PBS. Rabbit anti-mouse polyclonal antibodies against αSMA (cross reacts with human; Abcam), smooth muscle myosin heavy chain (Abcam), cytokeratin 5 (Abcam), cytokeratin 14 (cross reacts with human; Covance), cytokeratin 18-biotin (cross reacts with human; Abcam), mucin 1 (Abcam) and β-casein (gift from Christine Watson) were used at 2 μg/ml incubated overnight at 4°C. A chicken anti-mouse polyclonal antibody against GFP (Abcam) was used at 0.25 μg/ml. Cells and sections were then stained with a biotinylated donkey anti-rabbit IgG antibody or biotinylated donkey anti-chicken IgG antibody (Jackson ImmunoResearch) at 1 μg/ml for 1 h at 25°C. The Vectastain Elite ABC kit (Vector Labs) was used for avidin-horseradish peroxidase conjugation and 3,3'-diaminobenzidine (DAB) containing hydrogen peroxide (Vector Labs) was added and incubated for 2-10 min. For immunofluorescence, a goat anti-rabbit IgG-Alexa Fluor 555 (Invitrogen) was used at 1 μg/ml incubated for 1 h at 25°C. DAPI was used at 1 μg/ml to visualise cell nuclei.

Cell culture

Culture dishes were coated with Growth Factor Reduced (GFR) Matrigel (BD Biosciences) diluted 1:60 in PBS and incubated for 1 h at 37°C. Basal cells were cultured in FAD media: 3:1 DMEM/F12 (+1.8x10^{-4}M adenine + 1.8x10^{-3}M calcium) + 10% FBS (PAA) + 0.5 µg/ml hydrocortisone (Sigma) + 10^{-10}M cholera toxin (Enzo Life Sciences) + 10 ng/ml epidermal growth factor (EGF, Peprotech) + 5 µg/ml insulin + 50 µg/ml gentamicin. Cell culture media was supplemented with 10 µM Y-27632 (Sigma) unless stated otherwise. Irradiated NIH 3T3 fibroblasts were seeded at 10^4 cells/cm^2. Cultures were kept at 37°C in a 5% (vol/vol) CO_2, 5% (vol/vol) O_2 atmosphere for 7 days. For colony formation assays, mammary cells were seeded at 10-50 cells/cm^2 and cultured for 7 days (10 days for human cells) before fixation in 1:1 acetone/methanol. Colonies were stained with Wright-Giemsa (Fisher) and counted under a microscope at low magnification. For transplantations and analysis, cultured cells were trypsinised and sorted by flow cytometry using DAPI and EpCAM to isolate live epithelial cells and exclude dead cells and fibroblast cells. For some experiments, recombinant Human TGFβ1 (R&D Systems), SB 431542 hydrate (Sigma), blebbistatin (Sigma), latrunculin B (Enzo), cytochalasin D (Tocris Bioscience), jasplakinolide (Enzo) and H-1152 dihydrochloride (Tocris Bioscience) were added to cell culture media for some experiments. In Fig. 4g, freshly sorted cells were cultured in DMEM/F12 medium containing B27 (Gibco) + 1% FBS. For X-gal staining, colonies were fixed in 2% formaldehyde, 0.2% glutaraldehyde, 0.02% NP-40 in PBS at 4°C for 4 min, washed with PBS, incubated with the X-gal staining solution (0.025% X-gal, 3 mM K3Fe(CN)6, 3 mM K4Fe(CN)6, 1.5 mM MgCl2, 15mM NaCl, 40mM HEPES) overnight at 30°C, post-fixed with 4% formaldehyde, and counterstained with nuclear Fast Red. Colony-formation assay with basal and luminal cells from Myh11Cre-GFP; Rosa26 mice was performed three times with three technical replicates in each case. For mammosphere assays, freshly isolated cells were seeded at a density of 5,000 cells/well on ultralow-adherent 24-well plates (Corning) in DMEM/F12 containing B27 (Gibco), 20 ng/mL EGF (Invitrogen), 20

ng/mL bFGF (Gibco), 4 µg/mL heparin (Sigma), 10 µg/mL insulin (Sigma) supplemented with 2% GFR Matrigel. The mammospheres were dissociated every two weeks with 0.05% trypsin (Gibco) and were re-seeded at 5,000 cells/well.

Analysis of myoepithelial cell clonal expansion *in situ* and quantitative evaluation of LacZ-labeling in basal and luminal compartments

For analyses of clonal expansion and progeny tracing at puberty, 4-week-old Acta2-Cre-ER^{T2};Rosa26 females were i.p. injected twice with 1 mg of Tamoxifen within 24 hours. Mice were sacrificed one week (2 mice) and six weeks (three mice) after injection, glands dissected and processed for whole-mount x-gal staining as described elsewhere (Biology of the Mammary Gland, http://mammary.nih.gov). Between 339 and 511 clones were counted for 1-week chase time point, and between 260 and 528 for 6-week chase.

For analyses of clonal expansion and progeny tracing during pregnancy, 8-week-old *Acta2-Cre-ER^{T2};Rosa26LacZ* and *Acta2-Cre-ER^{T2};R26^{mTmG}* virgin females were i.p. injected twice with 0.75 mg of Tamoxifen within 24 hours, mated two weeks later and analysed 5-6 weeks after Tamoxifen injection at different pregnancy stages (Fig. 5e). For tracing the progeny of myoepithelial cells after two pregnancies, 4-week-old females were injected twice with 0.75 mg of Tamoxifen within 24 hours, mated at 8 weeks, allowed to lactate for one week, mated again after three weeks of involution and analysed after second lactation, either during involution (chase of 20 weeks), or at third pregnancy (chase of 22-26 week)

For histological analyses, x-gal-stained glands were embedded in paraffin, 7 µm-thick sections were cut, de-waxed and stained with nuclear Fast Red. For analyses of clonal expansion at pregnancy, clones were counted in 3 different gland areas and, at least, 9 sections per gland were analysed. 335, 281, 680, 648 and 492 clones were counted for V1, V2, P12, P15 and P17 samples, respectively.

To evaluate the amount of LacZ-positive cells in basal and luminal epithelial compartments, single cell suspensions were obtained from inguinal mammary glands of *Acta2-Cre-ER^{T2};Rosa26LacZ* mice as described above, basal and luminal cells isolated by flow cytometry, x-gal stained in suspension and LacZ-positive and negative cells were counted on cytospots. In each case, at least, 500 basal and 2,000 luminal cells were counted.

Cell injections into cleared fat pads

All animal work conducted in the UK was approved by the Cambridge Institute Local Ethics Committee and the Home Office and all animal work conducted in France was conducted in accordance with French veterinary guidelines and those formulated by the Council of Europe for experimental animal use (L358-86/609EEC). Cells were suspended in 65% HF media + 25% GFR Matrigel + 10% trypan blue solution (0.4%, Sigma), except for in Fig. 4j where 50% GFR Matrigel was used, at a concentration such that a 10 µl injection volume contained the desired cell dose. The endogenous mammary epithelium in the inguinal (number 4) glands of 3-week-old female C57BL/6J or NOD/SCID IL2Rγc$^{-/-}$ pups was cleared and cells were injected into cleared fat pads as previously described [31]. The mice were mated 3 weeks after surgery and the number 4 glands were removed during pregnancy. Glands were fixed overnight in Carnoy's fixative. GFP^{+} glands were visualised under a fluorescence microscope at low magnification and were then fixed in 4% paraformaldehyde for 1 h at 25°C before paraffin embedding. For whole-mount analysis, glands were stained overnight in carmine alum (Sigma) and analysed under a microscope at low magnification. After X-gal staining in whole-mount, glands were dehydrated and post-fixed in acetone for 30 min before carmine staining. For secondary transplantations, primary outgrowths (that contained GFP^{+} epithelium) were dissociated, made into a single cell suspension and were transplanted back into cleared mammary fat pads as described above except for in Supplementary Fig. 3f

where 1 mm^3 fragments, containing epithelial ducts that were clearly visible under a dissecting microscope, were re-transplanted into cleared mammary fat pads.

RNA extraction:

Total RNA was extracted from non-cultured, 1-day-cultured and 7-day-cultured basal cells using the RNeasy Plus Mini kit (Qiagen). The RNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific) and quality checked using a 2100 Bioanalyser (Agilent). All samples had a RNA integrity number ≥ 8.

Microarray summary:

Total RNA was prepared for microarray analysis using an Illumina TotalPrep RNA Amplification kit (Ambion). The cRNA was hybridised to a MouseWG-6 v2.0 Expression BeadChip (Illumina). All microarray data analyses were carried out on R using Bioconductor packages[32]. Raw intensity data from the array scanner were processed using the BASH[33] and HULK algorithms as implemented in the beadarray package[34]. Log2 transformation and quantile normalisation of the data from 4 independent experiments was performed. Differential expression analysis was carried out using the limma package[35]. Differentially expressed genes were selected using a p-value cut-off of <0.01 after global application of a false discovery rate correction for multiple contrast testing. The microarray data can be accessed via the Gene Expression Omnibus (GEO, series accession number GSE31347). Full microarray methods in accordance with the MIAME guidelines can be found at the GEO website.

Real-time PCR

Total RNA was extracted, quantified and quality checked as described above. Reverse transcription of the total RNA was performed using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Assuming a 1:1 conversion of RNA to cDNA, 2ng of

cDNA was used per reaction. TaqMan Fast Universal PCR Master Mix (Applied Biosystems) and TaqMan Gene Expression assays (Applied Biosystems) were used following the manufacturer's protocol. Reactions were done in 12.5µl volumes in triplicate using a 7900HT Fast Real-Time PCR System (Applied Biosystems) under the following conditions: 95°C for 20 sec followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec. Relative transcript levels were calculated using the comparative CT method and were normalised to the reference genes Actb and Rplp0. Full real-time PCR methods in accordance with the MIQE guidelines are detailed as follows:

MIQE checklist:

Experimental design: Experimental groups: Basal pre-culture, 1-day-cultured basal cells, 7-day-cultured basal cells. Control groups: No Template Control (NTC; no cDNA), RT- (no SuperScript). Technical replicates were done in triplicate. Relative transcript abundance was averaged from 3 independent biological replicates. Real-time PCR assays were performed by the lead author.

Sample: Mammary glands from 10-14 week old, female, virgin C57BL/6J mice were macrodissected and dissociated for 14-16 h at 37°C in DMEM/F12 containing collagenase and hyaluronidase. Mammary glands were treated with ammonium chloride, trypsin, dispase and DNase and then filtered to obtain a single cell suspension. Mammary cells were stained with fluorochrome-conjugated antibodies and the viability dye DAPI and were then passed through a cell sorter. Cultured basal cells were grown on Matrigel-coated cell culture plastic in FAD media supplemented with 10µM Y-27632 for 1-7 days at 37°C in a 5% (vol/vol) CO_2, 5% (vol/vol) O_2 atmosphere.

Nucleic acid extraction: Total RNA was extracted using the RNeasy Plus Mini kit (Qiagen) following the manufacturer's instructions. The RNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific) and quality checked using a

2100 Bioanalyser (Agilent). Most samples had an RNA integrity number (RIN) ≥8 (Supplementary Table 2).

Reverse transcription: The SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen; catalogue number 11754) was used following the manufacturer's protocol and reaction conditions. Reaction volumes were 20µl with 150ng of total RNA. Temperature and time: 25°C for 10 min, 42°C for 120 min, 85°C for 5 min. The vast majority of C_qs were undetermined for no-reverse transcription controls but where amplification did occur, C_q values were ≥37. C_q values with reverse transcription were between 20-35 depending on transcript abundance. All cDNA was stored at -20°C. The qPCR target information is reported in Supplementary Table 3.

qPCR protocol

TaqMan® Gene Expression Assays (Applied Biosystems) were used following the manufacturer's protocol. Final concentrations were 900nM for primers and 250nM for the TaqMan® MGB probe (6-FAM dye-labeled). Reaction conditions: TaqMan® Gene Expression Assay (20x) + TaqMan® Fast Universal PCR Master Mix (2x; contains DNA polymerase) No AmpErase® UNG (Applied Biosystems) + 2ng cDNA (0ng for NTC) + H_2O. Reactions were done in triplicate in 12.5µl volumes containing 2ng of RNA converted to cDNA. A 7900HT Fast Real-Time PCR System (Applied Biosystems) was used with the following thermocycling parameters: 95°C for 20 sec followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec.

qPCR validation:

Amplification efficiencies:

Gene Symbol	**Slope**	***y* intercept**	**PCR efficiency**	r^2

Acta2	-3.56	26.2	0.91	1
Actb	-3.65	23.1	0.88	1
Myh11	-3.63	23.7	0.89	1
Rplp0	-3.60	20.3	0.90	1

The linear dynamic range was from 0.02-2ng of cDNA per reaction. At 0.002ng of cDNA, C_q values for less abundant targets became variable (C_qs>35 and triplicates differed by >1 C_q).

Data analysis:

qPCR data were analysed and C_q values determined using the 7900HT Sequence Detection System software (Applied Biosystems). No template controls did not amplify and the *Myh11* no reverse transcription (RT^-) control did not amplify. The exceptions were amplification for *Acta2*, *Actb* and *Rplp0* RT^- controls with C_q values ≥37 which were 10-15 C_qs greater than those for samples with reverse transcribed template. *Actb* and *Rplp0* were used as reference genes based on their use in the literature and because their expression did not differ significantly between the samples according to the microarray data. The geometric mean of the 2 reference gene Ct values was used for normalisation.

Western blotting:

An equal number of flow-sorted basal $EpCAM^{high}$ and $EpCAM^{low}$ cells were centrifuged at 470 x g, the supernatant removed and cells lysed in 10µl of CHAPS lysis buffer (20 mM Tris, 8M Urea, 5 mM Magnesium Acetate, 4% CHAPS; pH to 8.0; Amersham). Protein lysates were sonicated for 30 sec and diluted with NuPAGE LDS sample buffer and NuPAGE reducing agent (Invitrogen) before being run on a NuPAGE Bis-Tris gel (Invitrogen) in NuPAGE MOPS SDS running buffer (Invitrogen) for 30 min at 60V and 90 min at 120V. The protein was transferred onto a

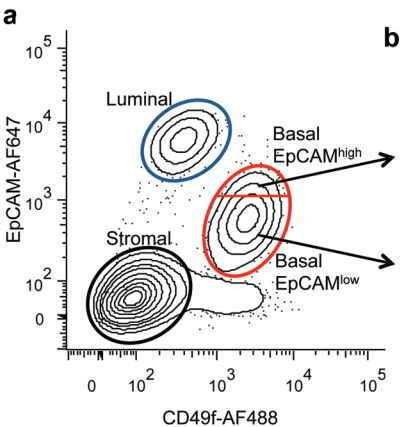
nitrocellulose membrane using the iBlot transfer system (Invitrogen). The membrane was incubated in Odyssey blocking buffer (LI-COR) for 1 h at 25°C before being incubated with a rabbit anti-mouse αSMA antibody (Abcam) at 2 μg/ml overnight at 4°C followed by a goat anti-rabbit IRDye 800CW secondary antibody (LI-COR) at 0.2 μg/ml for 45 min at 25°C. The membrane was then imaged on a LI-COR Odyssey CLx infrared imaging system. The blot was then re-probed for cytokeratin 14 (1 μg/ml; Covance) following the same protocol.

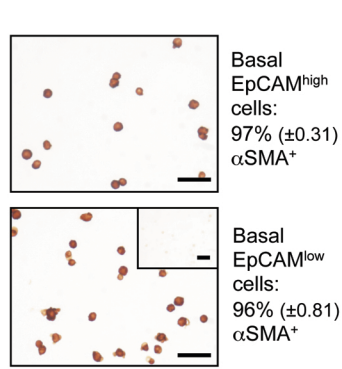
Statistical analysis

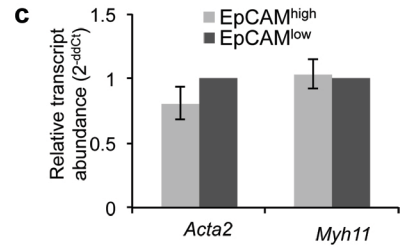
Data are presented as the mean of independent experiments with SEM. Comparisons between multiple groups were analysed using analysis of variance (ANOVA) followed by Bonferroni post-tests. Comparisons between 2 groups were analysed using 2-tailed, 2-sample Student's t-tests. Statistical significance was set at $P < 0.05$. All raw data can be found in the data source file.

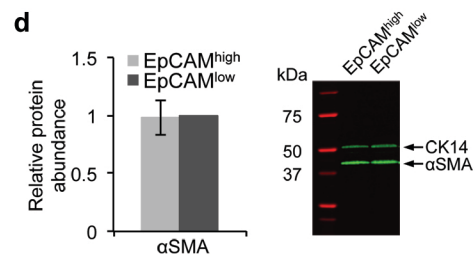
MRU frequencies between different cell populations were compared statistically using the Extreme Limiting Dilution Analysis (ELDA)[36] online tool (http://bioinf.wehi.edu.aU/software/elda/).

Figure 1 (Prater et al.)





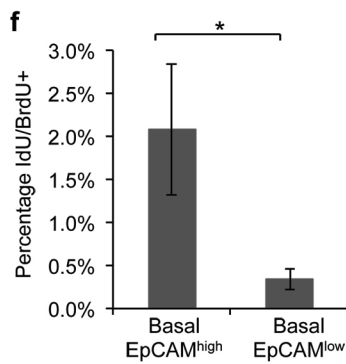




e

Cell population	Cell dose	Engrafts	MRU freq. [95% CI]	MRU distribution (%)
Basal EpCAM^{high}	50 100	7/12 5/6	1 in 57 [1/31 to 1/104]	57
Basal EpCAM^{low}	250 500	8/12 4/6	1 in 300 [1/163 to 1/552]	43

**



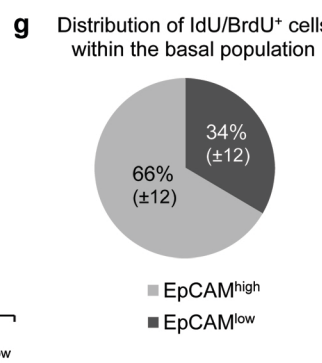
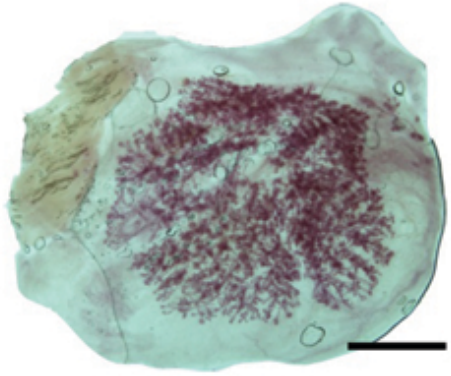


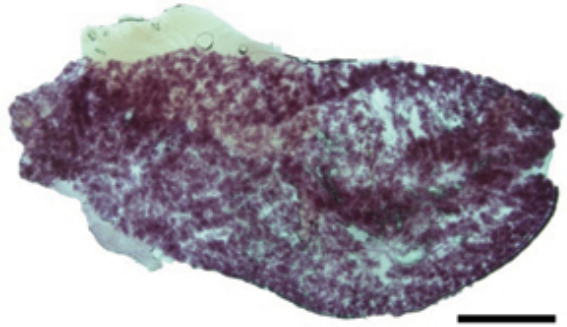
Figure 2 (Prater et al.)

a

Cells transplanted	Number of cells injected / fat pad	Number of engrafts	MRU frequency [95% CI]	Number of epithelial cells / dish	Number of MRUs / dish
Non-cultured basal	5 15 20 38 200 250 500	0/9 0/7 1/9 1/7 10/12 2/4 3/4	1 in 205 [1/122 to 1/343]	4,750 seeded (±250)	23 seeded (±1)
7-day-cultured basal	10 100 500 1,000	0/4 6/8 5/5 4/4	1 in 79 [1/35 to 1/177]	1,465,455 (±787,519)	10,551 (±1,418)

b





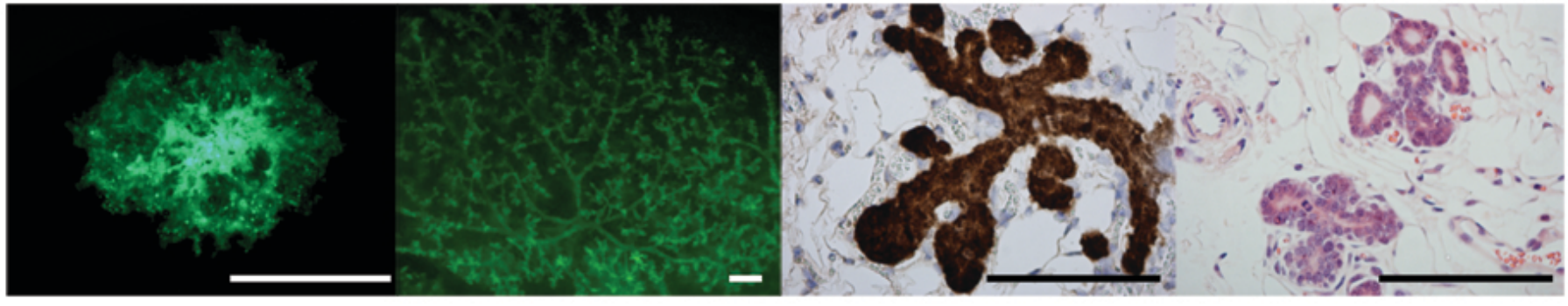
Cells transplanted: Non-cultured basal 7-day-cultured basal

Figure 3 (Prater et al.)

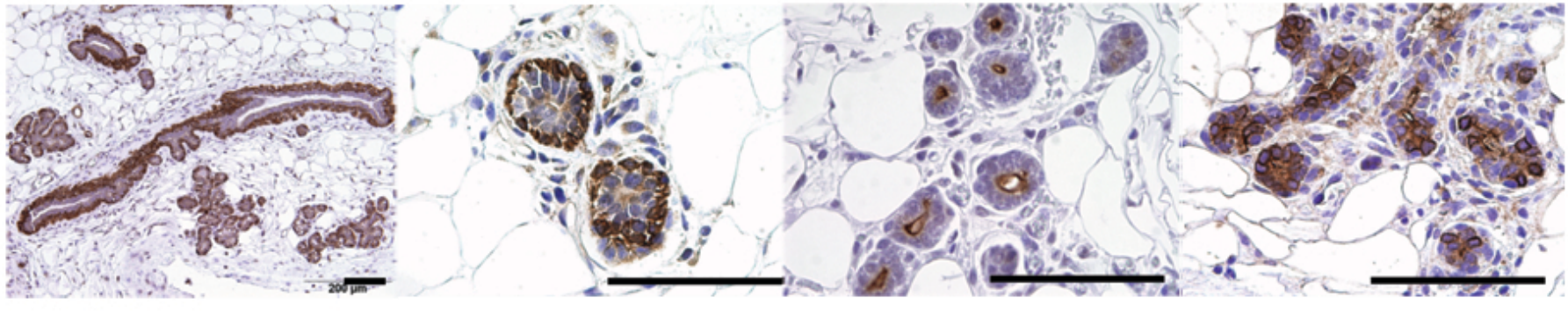
a

Cell population	Single-cell cloning efficiency (%)	% of single-cell-derived colonies that engrafted
Basal EpCAM^{high}	53.5 (±4.1)	83 (15/18)
Basal EpCAM^{low}	29.7 (±5.4)	61 (11/18)

b

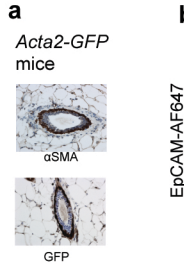


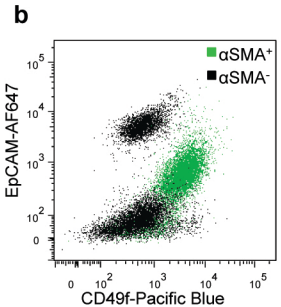
GFP^{+} basal colony | GFP^{+} engraftment | GFP | H&E

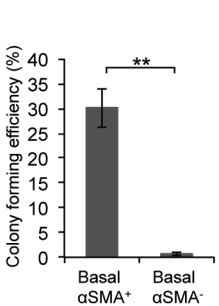


αSMA | Cytokeratin 14 | Mucin 1 | β-casein

Figure 4 (Prater et al.)





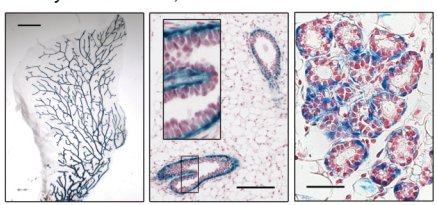


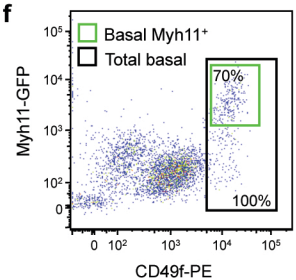
d

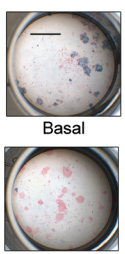
Cell population	Number of cells injected / fat pad	Number of engrafts	MRU frequency [95% CI]
Basal αSMA$^+$	50 100 200 500	1/5 6/8 8/9 5/5	1 in 93 [1/54 to 1/160]
Basal αSMA$^-$	10 20 25 50	0/3 0/4 0/10 0/10	NA [1/287 to 1/∞]

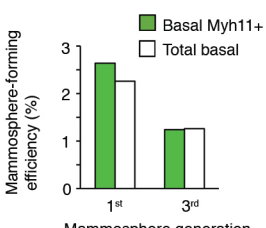
**

e *Myh11-Cre-GFP;Rosa26* mice









j

Cell population	Number of cells injected / fat pad	Number of engrafts	MRU frequency [95% CI]
Basal Myh11$^+$	50 200 500	5/7 7/8 10/10	1 in 67 [1/33 to 1/136]
Total basal	50 200 500	6/7 8/9 10/10	1 in 26 [1/10 to 1/64]

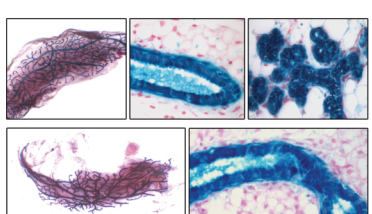


Figure 5 (Prater et al.)

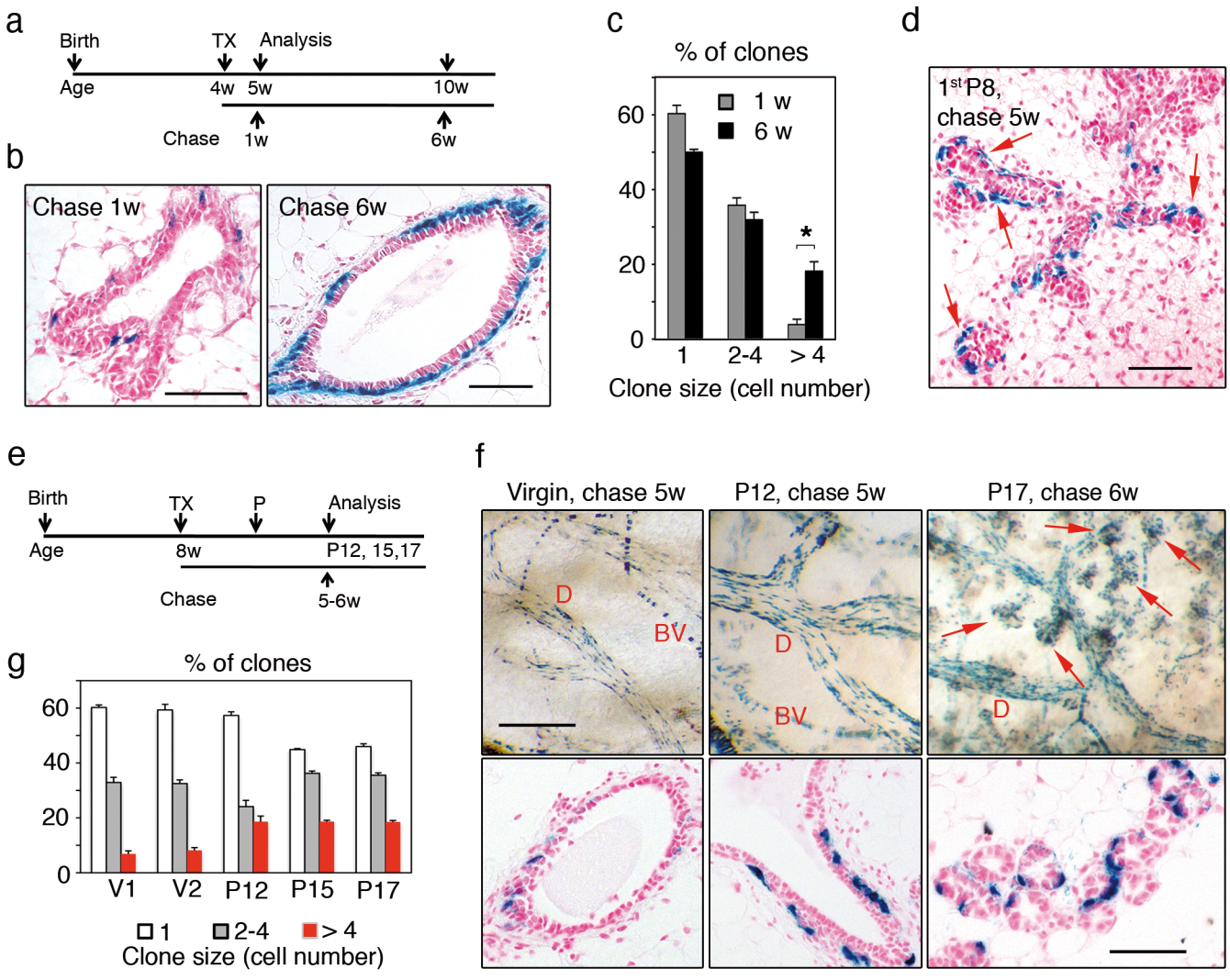


Figure 6 (Prater et al.)