Supporting Information

Fridriksdottir, Villadsen et al.

SI Materials and Methods

Fluorescence activated cell sorting (FACS): Gating included forward scatter and side scatter to remove debris and select for single cells (1). CD271 and CD326 are restricted to myoepithelial and luminal cells, respectively (2, 3), and screening of sorted populations by double immunofluorescence identified $CD271^{high}/CD326^{low}$ cells as $K19^{-}/K14^{+}$ and $K19^{-}/\alpha$-smooth muscle $actin^{+}$. In some experiments, stromal cells were prospectively excluded by incubation with a mixture of CD31 (JC70A; 1:50), CD34 (QBEnd/10; 1:50), CD45 (Bra-55; 1:250), and fibroblast surface protein (1B10; 1:50) recognizing stromal cells (3) without change of the experimental outcome. To purify myoepithelial cells from TDLUs and ducts, respectively, $CD326^{low}/CD271^{high}$ cells were further sorted into $PDPN^{high}$ and $PDPN^{low}$ cells by podoplanin-AF488 (NC-08, 1:50, BioLegend). For isolation of luminal progenitors from differentiated MEP cultures, cells were incubated with CD117 (K45, 1:50, Neomarkers) for 30 min at 4°C, washed twice and incubated with secondary antibody AF647 goat anti- mouse IgG2a (1:500, Life Technologies, A21131). Negative controls were incubated with secondary antibody only. Upon incubation, the cells were washed twice in HEPES/BSA/EDTA buffer and filtered through a 10 or 20 μm filter cup (Filcons). Propidium iodide (1:1000, Life Technologies) or Fixable Viability Stain 780 (1:1000, BD Biosciences, cat. no. 565388) was added to distinguish live from dead cells, and the cells were analyzed and sorted using a FACSAria or a FACSFusion flow cytometer (BD Biosciences).

Cell culture: MEPs grow in Myo medium (DMEM/F-12 (Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12, 1:1, Life Technologies), 2 mM glutamine (Invitrogen), 1 μg/ml

hydrocortisone, 9 μg/ml insulin, 5 μg/ml transferrin (Sigma-Aldrich), 5.2 ng/ml Na-Selenite (BD Industries), 100 μM ethanolamine (Sigma-Aldrich), 20 ng/ml basic fibroblast growth factor (PeproTech), 5 nM amphiregulin (R&D Systems), with the addition of 1.8 $x10^{-4}$ M adenine (Sigma-Aldrich), 10 μM Y-27632 (Y0503, Sigma-Aldrich or 1683, Axon Medichem) and 20 μl/ml B27 (Life Technologies), and non-myodifferentiated myoepithelial cells, NMMEPs, were propagated in Epi medium (Dulbecco´s modified Eagle´s medium (DMEM, high glucose, no calcium, Life Technologies):Ham´s F12 Nutrient Mixture (F12, Life Technologies), 3:1 v/v), 2 mM glutamine, 0.5 μg/ml hydrocortisone, 5 μg/ml insulin, 10 ng/ml cholera toxin (Sigma-Aldrich), 10 ng/ml epidermal growth factor (Peprotech), 1.8 $x10^{-4}$ M adenine, 10 μM Y-27632 and 5% FBS, modified from (4, 5)). Cultures had their medium changed three times a week and were passaged before confluence. NIH3T3 cells were cultured in Dulbecco´s modified Eagle medium, DMEM 1965 (Gibco) supplemented with 2mM glutamine and 10% FBS. Human breast fibroblasts were cultured in collagen-coated flasks (Nunc, 8 μg/cm^2 PureCol, Nutacon) in chemically defined medium, CDM3, supplemented with 20% FBS (6, 7) and used between passages 6 and 10. All cultures were maintained at 37 °C in a humidified atmosphere with 5% CO_2. All media included 0.5% gentamycin (Biological Industries). For phase contrast microscopy a Nikon Diaphot 300 microscope was used.

Cell propagation, differentiation and clonal colony formation: To determine whether loss of myoepithelial differentiation in Epi medium was irreversible, NMMEPs were exposed to Myo medium in second passage for 12 days prior to staining for myoepithelial and luminal markers. Parallel cultures of myoepithelial cells continuously run in Myo medium served as control.

To induce luminal differentiation, the feeder cells were removed by differential trypsination (exposure for 30 sec to 0.25% trypsin/1 mM EDTA at room temperature followed by removal of

detached fibroblasts, followed by addition of fresh trypsin to detach epithelial cells into single cells). The epithelial cells were plated on collagen-coated culture flasks or wells, 2,000 cells/cm^2, in MEGM complete medium (CC-3150, Lonza) supplemented with 2 mM glutamine, 4 μg/ml Heparin (Sigma-Aldrich), 20 ng/ml basic fibroblast growth factor, 20 ng/ml epidermal growth factor, 20 μl/ml B27 and a small molecule transforming growth factor-beta receptor inhibitor, RepSox (25μM, R0158, Sigma Aldrich). The cells were cultured for 7-10 days prior to immunocytochemical staining. As a read out for luminal differentiation, cultures were stained by immunoperoxidase for K19 (see below and Table S1), and the frequency of K19$^+$ luminal progenitor clones were quantified. Luminal progenitor clones (K19$^+$ CFUs) in differentiated cultures derived from TDLU-enriched biopsies (n=4) or duct-enriched biopsies (n=4) images of peroxidase stained samples were acquired on Leica DM5500B and then imported and analyzed in batch using in house designed macros (ImageJ version 1.50g). The macros used to determine the areas positive for K19 in stained samples were as follows: Images were converted to RGB followed by color deconvolution (plugin Color Deconvolution 1.7, H-DAB) and import of the Colour-2 image to the Trainable Weka Segmentation tool (v2.3.0). A sample specific classifier was generated to discriminate background from peroxidase (brown) stained areas by training on one image per sample. The resulting probability map image was subtracted by 0.5, binarized (plugin Make Binary) and measured, excluding areas smaller than 3,000 μm^2 (plugin Analyze Particles). The measured area outline was added to the corresponding original image for visualization of the final threshold (plugin Analyze Particles ROI manager). To determine the area covered by cells, original images were imported to the Trainable Weka Segmentation tool and a sample specific classifier trained to distinguish areas covered by cells. The probability maps were binarized, inverted, selected (plugin Create Selection) and measured (plugin Measure). Data are represented as the

total area of K19 stain relative to the total area covered by cells. To estimate progenitor cell frequency, we first calculated the mean area of the ten largest $K19^+$ CFUs in four samples and set the inclusion threshold at 95%, above which colonies with less than approximately 15 cells were considered abortive. The number of CFUs was divided by the number of cells initially seeded per analyzed area (2,000 cells/cm^2).

To estimate K19, K5, α-smooth muscle actin, vimentin, CD90 and CD271 protein expression relative to K14, images of peroxidase stained paired samples were acquired on Leica DM5500B and analyzed in batch using Trainable Weka Segmentation and a classifier. Probability maps were subtracted by 0.7 and binarized (plugin Make Binary) followed by area measurement (plugin Analyze Particles), excluding areas smaller than 300 μm^2. Data are presented as area of stain relative to area stained for K14. Estimation of α-smooth muscle actin and vimentin is based on cell populations from four biopsies, and estimation of K19, K5, CD90 and CD271 is based on three biopsies. The nuclear marker, p63, did not qualify to be estimated by this method.

Single cell preparations from each organoid collection from micro-collected ducts and TDLUs were incubated with CD271-APC and EpCAM-BV421 followed by Fixable-viability-780 to distinguish live from dead cells and sorted by FACS. Sorted myoepithelial cells were plated under standard MEP culture conditions (NIH-3T3 feeder, $4x10^3$ cells/cm^2, Myo medium) and cultured for 10-14 days followed by removal of residual feeder cells by differential trypsination followed by trypsination of epithelial islets for further experimentation in either luminal differentiation assays or for passaging on top of new feeders for use for *in vivo* transplantation.

***In vivo* transplantation:** 2x10^6 epithelial cells from either NMMEP (eight inoculations from three biopsies) or MEP cultures (thirteen inoculations from six biopsies), MEP derived from TDLU- or duct-enriched biopsies (eleven inoculations from four biopsies of each) or MEP derived from microcollected TDLUs or ducts (four inoculations of each from two biopsies) were mixed with 5x10^5 irradiated (~20 Gy) human breast fibroblasts and suspended in 1:1 collagen gel : growth-factor reduced Matrigel (BD Bioscience) and kept cold. 150 µl of the cell suspension was injected orthotopically to the 4th mammary gland of 5-10 week old female NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Sug} mice (NOG mice, Taconics). Either one or two inoculations were made per mouse. For transplantation of single cell NMMEP clones from TDLUs and ducts (inoculations from three clones of each derived from one biopsy), respectively, in some cases fewer epithelial cells (2.25 x 10^5 – 2x10^6) used per inoculum due to limited experimental material. A limited dilution series had shown that a minimum of 1x10^5 cells are needed for generation of structures. One week prior to the inoculation and throughout the experimental period, the drinking water was supplemented with 0.67 µg/ml of 17β-estradiol (Sigma-Aldrich). Mice were sacrificed after 8-9 weeks, mammary glands were excised and snap-frozen in -80° C n-hexane and mounted for cryostat sectioning to analyse epithelial-derived structures.

RNA isolation, RT-qPCR and RNA-seq analysis: Total RNA was isolated by Trizol (Invitrogen) according to the manufacturer´s protocol, and 2 ng of cDNA derived from reversely transcribed RNA using the High Capacity RNA-to-cDNA Kit (Applied Biosystems) was used to perform quantitative real- time (RT) PCR using the TaqMan® Gene Expression Assays (Applied Biosystems) in a CFX384 Touch™ RT-PCR Detection System (Bio-Rad). The primers are listed in Table S2, and RT- qPCR conditions were the following: 50^{o}C for 2 minutes and 95^{o}C for 10 minutes, followed by

40 cycles at 95°C for 10 seconds and 60°C for 30 seconds. For quantification of gene expression, each gene expression was normalized by the geometric mean of four reference gene expressions (glyceraldehyde-3-phosphate dehydrogenase (GAPDH), hypoxanthine phosphoribosyltransferase 1 (HPRT1), TATA-box binding protein (TBP) and phosphoglycerate kinase 1 (PGK1), based on the $2^{-\Delta\Delta Ct}$ method (8). RNA-sequencing and bioinformatics analysis was conducted by Beijing Genomics Institute (BGI), Hongkong. In short, DNAse treated total RNA samples were enriched for mRNAs using oligo dT magnetic beads. In turn, mRNAs were fragmented into 200 bp-size fragments and the first strands of cDNAs were synthesized by using random-hexamers. In order to generate the library products, after the second strand synthesis, double strands cDNAs were purified by magnetic beads followed by addition of adenine to the 3'ends and ligation of sequencing adaptors to the fragments. These fragments were amplified by PCR using ABI StepOnePlus Real-Time PCR System, and the quantity and quality was checked by Agilent 2100 Bioanalyzer. Sequencing was performed using Illumina HiSeq™ 2000 and 10M clean reads per sample was generated. Data processing included removing adaptors, any contamination and low-quality reads from raw reads. Based on the NOIseq method, a list of significantly differentially expressed genes (DEGs) was established using a cutoff value as fold difference higher than 2 between two groups and probability higher than 0.6. A bioinformatics resource DAVID 6.8 was used for overrepresentation of gene ontology analysis. We selected the gene-sets skin and keratinocyte (grouped as Skin/Keratinocyte) and the gene-sets contractile fiber, actin cytoskeleton and myofibril (grouped as Actin/Myosin). The genes within the selected gene-sets are presented in a heatmap with RPKM values on a $\log_2$ scale and have been row-scaled.

Immunohistochemistry and cytochemistry: A biopsy was given a positive score of heterogeneity if more than one third of the TDLUs in the stained section showed a patterned staining with K19.

Only biopsies with detectable TDLUs were used for scoring while ductal-enriched biopsies were excluded. The staining pattern observed by peroxidase staining used for quantification was validated by fluorescence staining with K19 and nuclear stain. To further ensure that the observed heterogeneous K19 expression of the luminal population by immunoperoxidase staining was not confounded by presence of $K19^-$ myoepithelial cells, the staining pattern was confirmed by double labeling immunofluorescence with K19 and K14 combined with nuclear stain. To confirm the luminal phenotype of $K19^+$ cells, organoids from three biopsies were trypsinized, smears were doublestained by immunoflourescence with K19 and MUC1, and the frequency of double-labelled cells were quantified on acquired digitized micrographs of three preparations from each biopsy (average and standard deviation of 3 x 100 cells). To determine the expression pattern of podoplanin (clone D2-40, Dako), a total of 45 biopsies was stained by immunoperoxidase and counterstained with hematoxylin.

For quantification of luminal contamination of sorted myoepithelial cells by three different FACS protocols, cellular smears were fixed for 5 min at -20°C in icecold methanol followed by blocking in 10% goat serum/PBS for 5 min. The smears were then stained against K14 (LL002, 1:50, Neomarkers) and K19 (A53-B/A2, Abcam) followed by appropriate fluorescent secondary antibodies, i. e. isotype-specific goat anti-mouse IgG AlexaFluor or goat anti-rabbit antibodies (Life Technologies) as used in all fluorescence stainings. For quantifications and imaging of fluorescence, slides were mounted with ProLong Gold antifade reagent with 4, 6-diamidino-2-phenylindole (DAPI; Life Technologies), and quantification of stained smears was routinely based on 3 x 100 cells, directly observed with a x 20 objective in a fluorescence microscope. Nuclei of peroxidase stained sections and cells were counterstained with hematoxylin. For antibodies and

further details see Table S1. To determine luminal heterogeneity, a panel of 38 biopsies was stained by immunoperoxidase for K19 (clone BA16, GenWay). Similar staining patterns in sections were obtained with either fixation in icecold methanol for 5 min (K19 antibody used 1:100 or 1:200) or at RT in formaldehyde and permeabilized in 0.01% Triton-X100 as described above (K19 antibody used 1:1000 or 1:2000). Smears of $PDPN^{high}$ and $PDPN^{low}$ cells were fixed at RT for 10 minutes in 3.7% formaldehyde, washed 3x in phosphate- buffered saline (PBS) and incubated in 0.01% Triton X-100 for 10 min at RT followed by 3x PBS washing. The fixed smears were blocked for 5 min with Ultra V Block (Lab Vision Corporation, TA-125-UB) followed by 5 min with 10% goat serum in PBS prior to staining for K17. In each preparation the frequency of $K17^{+}$ cells among $PDPN^{high}$ and $PDPN^{low}$ cells was quantified (average and standard deviation of 3 x 100 cells).

To identify TDLU- and duct-enriched biopsies, serial sections were made for a panel of biopsies, which were hematoxylin stained and scored for their TDLU content. That the observed histology of a given biopsy was representative for the entire biopsy was confirmed in minimum three samples of three of the biopsies. To quantify the contribution from ducts in TDLU-enriched and duct-enriched biopsies, respectively, the frequency of ducts (> 1500 μm) in collagenase digests were first scored by phase contrast microscopy, then the organoids were trypsinized and the released cells manually counted to estimate the number of large ducts per million cells. Four representative biopsies for each category were chosen for further experimentation.

For analysis of epithelial-derived structures in transplanted mammary fat pads, mammary fat pads were excised, snap frozen in -80°C n-hexane and sectioned at a setting of 8 μm in a cryostat. Human-derived structures were detected by their positive reaction against K19 or K17 (clone E3, Dako), which does not react with mouse mammary tissue. Immunofluorescence and peroxidase

stainings were evaluated, quantified and photographed using a laser-scanning microscope (LSM 700) and brightfield microscopes (Laborlux S or DM5500B, Leica), respectively.

The frequency of $K19^+$ profiles in transplants was determined as the number of $K19^+$ structures/total number of structures. For quantification of $K19^+$ structures derived from TDLU- or ductal-enriched biopsies, representative sections with the highest number of structures were imaged and analyzed. Two to three injections were performed for each biopsy. The different transplantations from the same biopsy gave comparable results. One representative transplantation from each biopsy was chosen for quantification. K19/K7 expression profile of structures generated from transplants were compared with structures of normal breast *in situ* by triple-labeling for K7 (LPK 5, Chemicon), K19 (BA16) and, to identify the outer layer of myoepithelium, α-smooth muscle actin (1A4, Sigma), and quantified in photos of structures derived from TDLUs and ducts, respectively (average and standard deviation of 3 x 100 cells).

Western blotting: Whole cell lysates were made by incubation in RIPA buffer (R0278, Sigma) with protease, and phosphatase II and III inhibitor cocktails (P8340, P5726, and P0044, Sigma) followed by passing through a 26G needle. Proteins were separated by a 12 % Novex® Bis-Tris pre-cast polyacrylamide gradient gel (Life technologies) and transferred overnight to a PVDF membrane. Molecular weight was indicated by using a pre-stained protein ladder (SM0671, Fermentas). After blocking one hour in room temperature, the membrane was incubated overnight with antibodies against p16 (1:1000, ab108349, abcam) or β-actin (1:5000, A-5441, Sigma) in blocking solution with gentle rocking at 4 °C, followed by incubation for one hour at room temperature with secondary horseradish peroxidase–conjugated antibodies (P0399 and P0447, DAKO). Western

blots were visualized using enhanced chemiluminescence solution (PerceECL 32106, Thermo Scientific) and a chemiluminescence imager (Amersham Image 600, GE Healthcare Life Sciences).

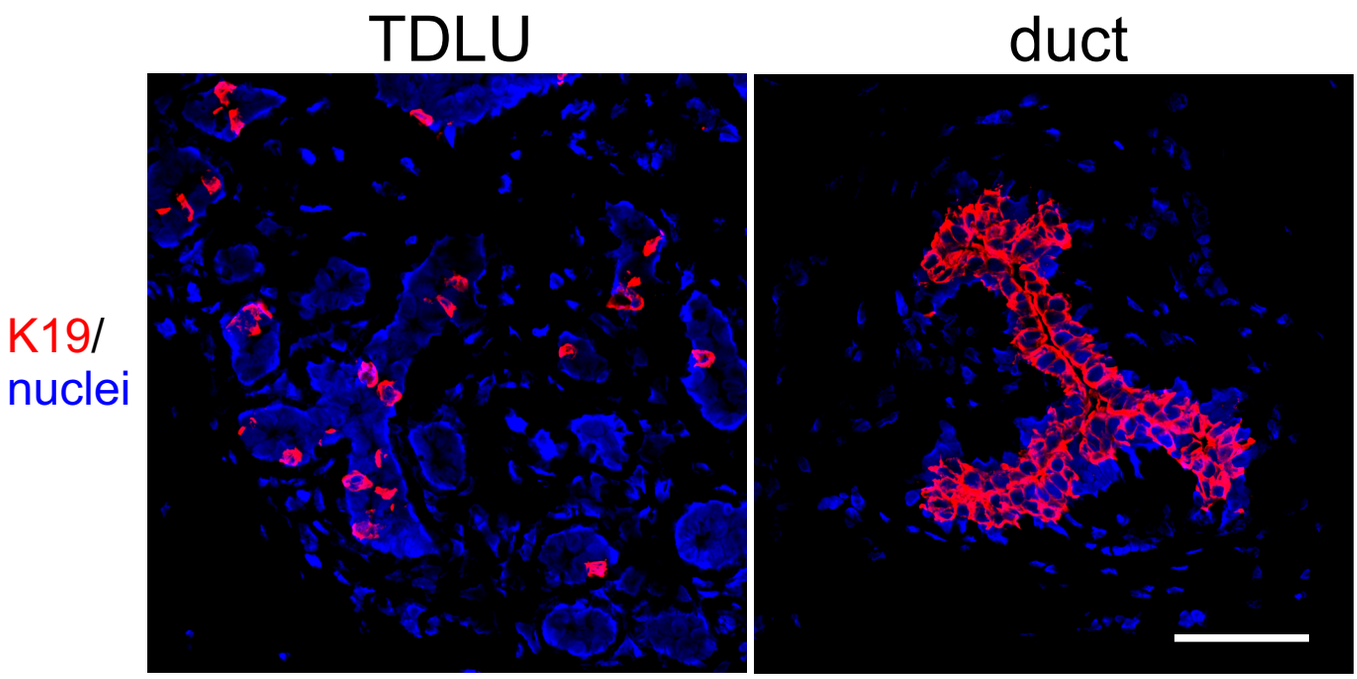

Fig. S1. Multicolor imaging of a TDLU and a duct stained with keratin K19. Multicolor imaging of a cryostat section of reduction mammoplasty biopsy #W430-2 including a TDLU (left) and a duct (right). Sections were stained by fluorescence for K19 (red) and nuclei (blue). The images represent fluorescence microscopy examples of typical heterogeneous and homogeneous stainings, respectively (Bar = 50µm).

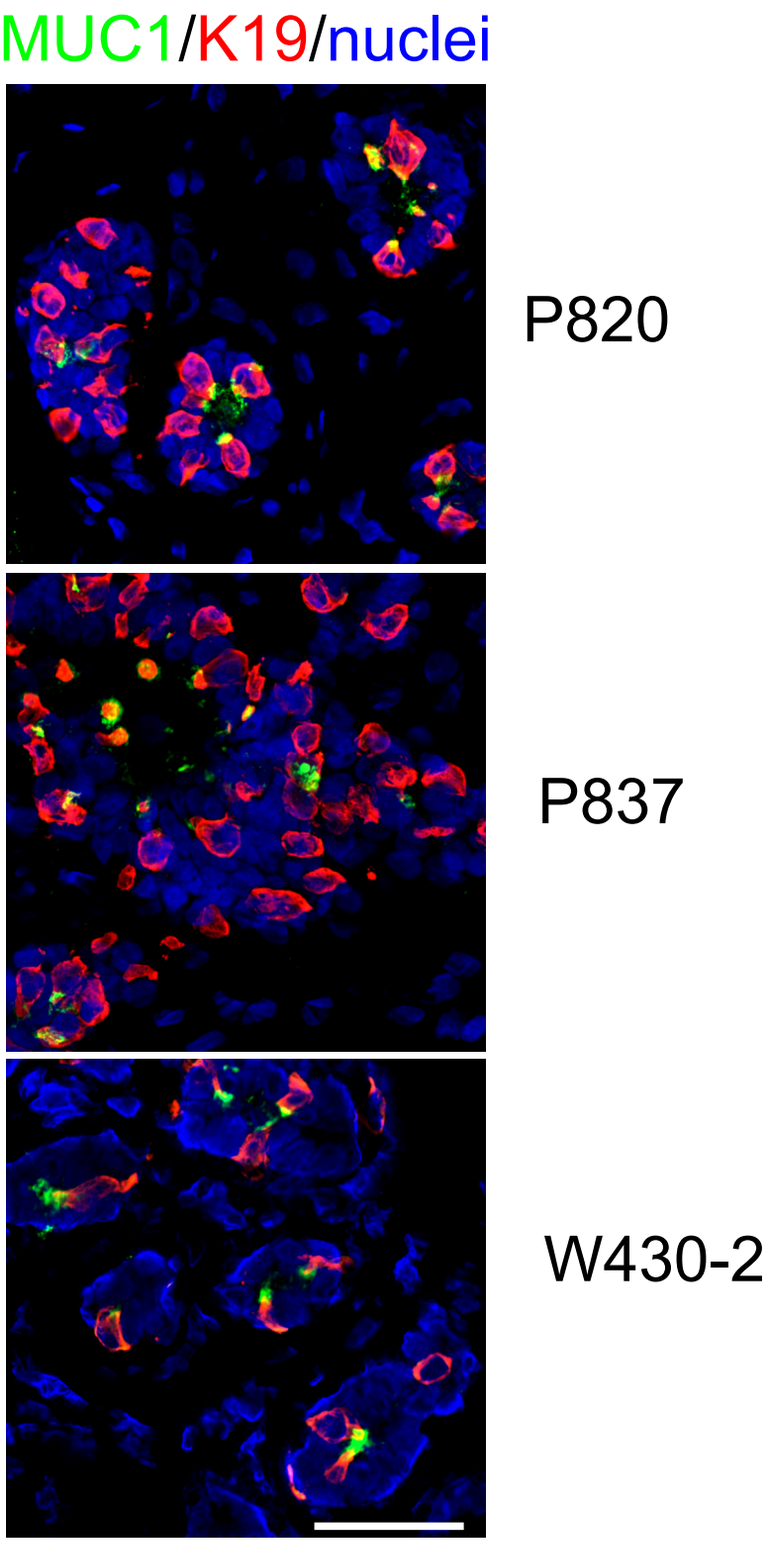

Fig. S2. MUC1 and keratin K19 are part of the same differentiation program in heterogeneous TDLUs. Multicolor imaging of cryostat sections of reduction mammoplasty biopsies #P820, #P837, and #W430-2. Sections were stained by fluorescence for MUC1 (green), K19 (red) and nuclei (blue). Note the coordinate expression of cytoplasmic keratin K19 and MUC1 at the apical plasma membrane when present in the focal plane (Bar = 50μm).

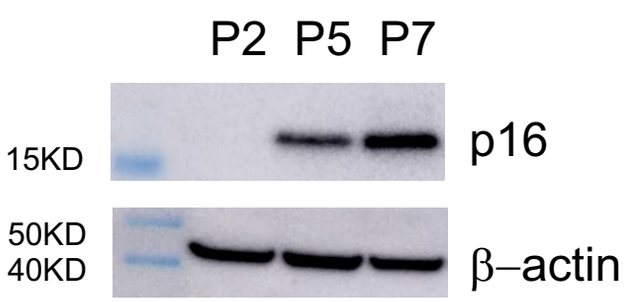

Fig. S3. Myoepithelial-derived cells reach p16 dependent stasis in passage five to seven. Western blotting of proteins extracted in passage 2 (P2), 5 (P5) and 7 (P7) of $CD326^{low}/CD271^{high}$ cells cultured on fibroblast feeders in Myo medium incubated with antibodies recognizing $CDKN2A/p16^{INK4A}$ (p16, upper panel) and β-nonmuscle actin (β-actin, lower panel) with molecular weight markers included (first lane). In low passage p16 is not expressed but is switched on by passage 5 and further increased by passage 7.

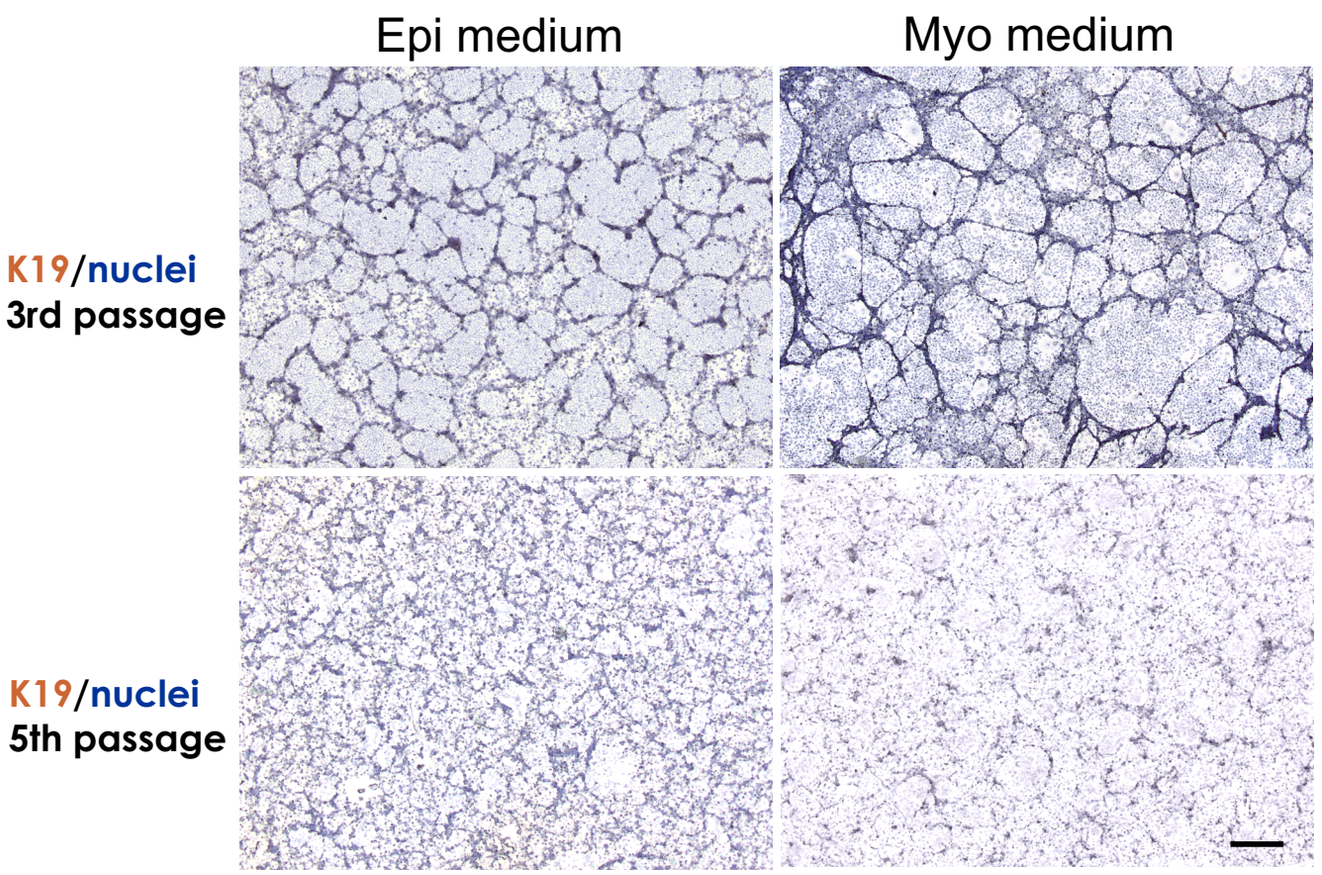

Fig. S4. Myoepithelial-derived cells in either Epi or Myo medium refrain from luminal conversion with subculture. Low magnification micrographs of (upper row) third passage and (lower row) fifth passage subconfluent cultures of myoepithelial-derived cells on (left column) Epi medium or (right column) Myo medium. Cultures were stained by immunoperoxidase for K19 and counterstained with hematoxylin (nuclei). There are no $K19^{+}$ cells present. (Bar = 500 μm).

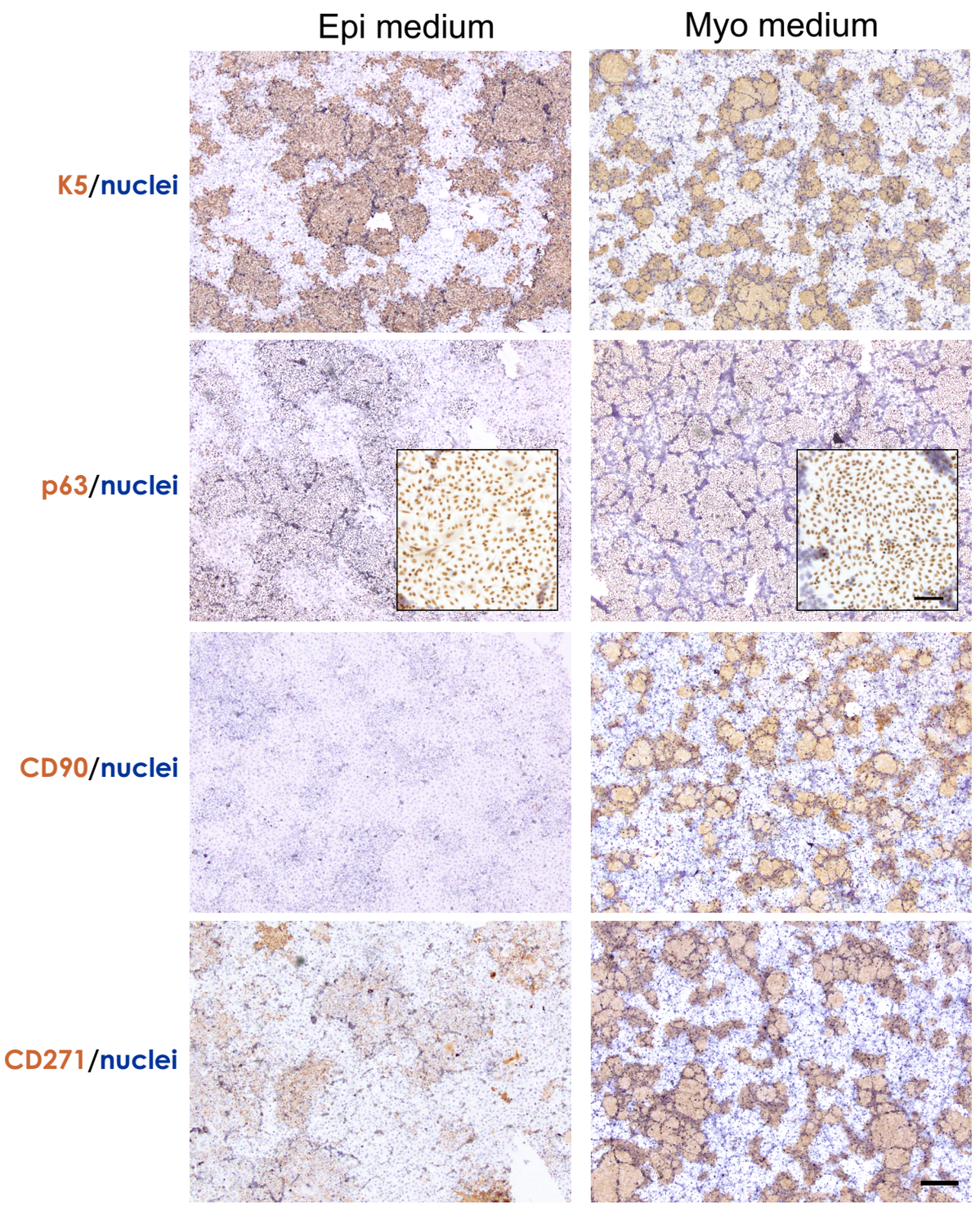

Fig. S5. Myoepithelial-derived cells in serial subculture express keratin K5 on both Epi and Myo medium but differ in their myodifferentiation profile. Immunoperoxidase stainings of subconfluent cultures of myoepithelial-derived cells on either (left column) Epi medium or (right column) Myo medium. Cultures were stained by immunoperoxidase for K5 (upper row), p63 (second row), CD90 (third row), and CD271 (lower row), and counterstained with hematoxylin (blue, nuclei). Whereas all cells stain with myoepithelial markers (K5 and p63), only cells in myoepithelial medium stain for CD90 and CD271. (Bar = 500 µm; inset: 100 µm).

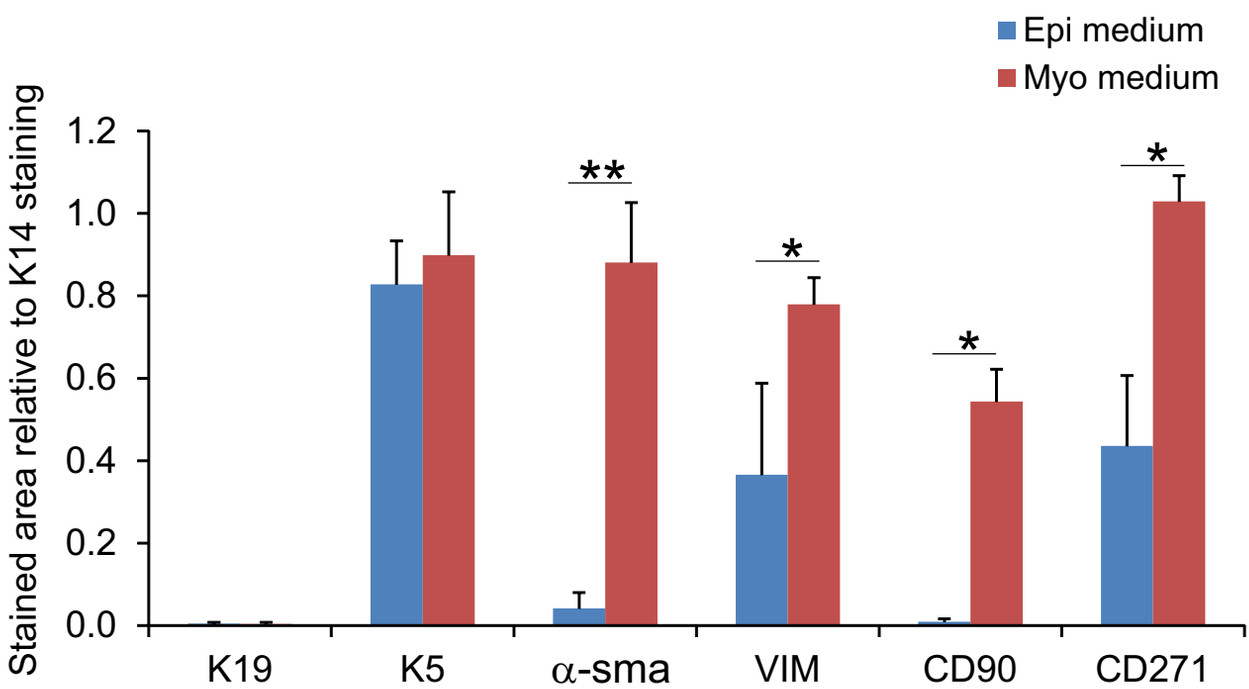

Fig. S6. Expression of epithelial and mesenchymal markers relative to keratin K14 expression as influenced by culture medium. Area stained for K19 (K19), K5 (K5), α-smooth muscle actin (α-sma), vimentin (VIM), CD90 (CD90) and CD271 (CD271) relative to area stained for K14. While Epi (blue bar) as well as Myo (red bar) culture medium maintain low K19 and high K5 expression, Myo medium significantly supports the expression of α-smooth muscle actin, vimentin, CD90 and CD271. *: $p<0.05$, **$p<0.005$ by Student's T test. Error bars indicate standard deviations.

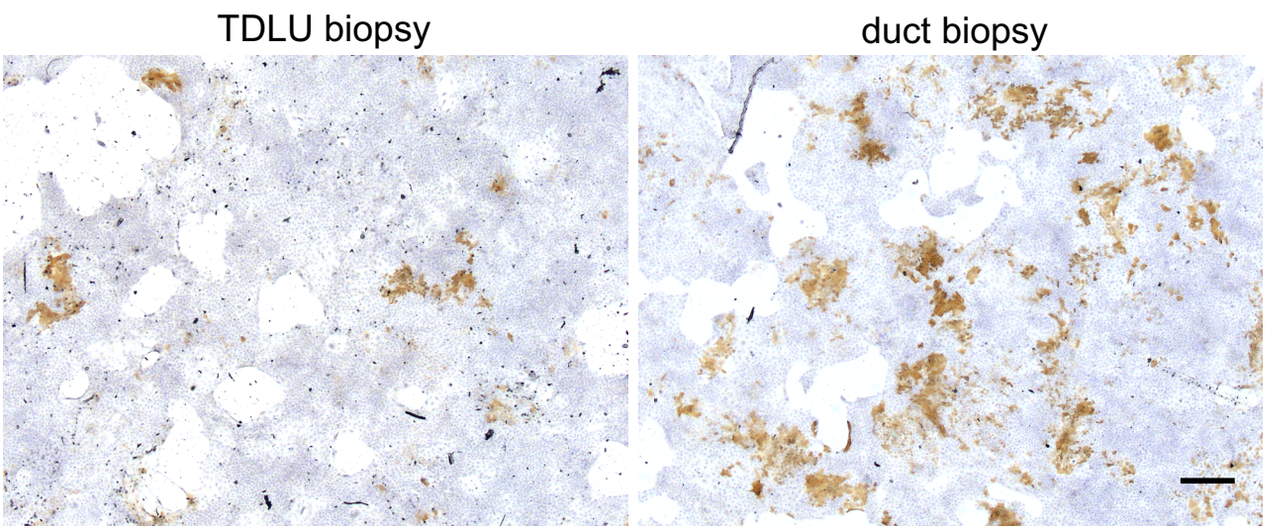

Fig. S7. TDLUs differ from ducts in the ability to generate $K19^{+}$ cells. Low magnification micrographs of cultures stained with immunoperoxidase against keratin K19 to illustrate the staining intensity behind digitized images for area quantifications. Cultures were counterstained with hematoxylin. (Bar = 500 μm).

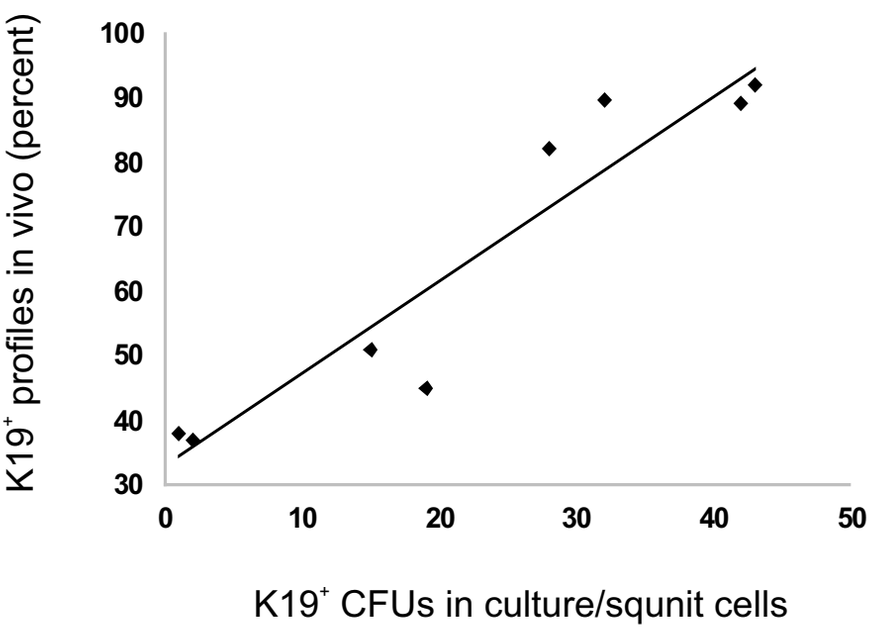

Fig. S8. K19^{+} colony forming ability correlates with K19^{+} profiles from transplanted MEPs. K19+ profiles generated in vivo from transplanted MEPs in percent of total number of profiles plotted against number of K19^{+} CFUs derived from MEPs in culture per square unit show a positive correlation by Spearman correlation test ($r=0.93$, $p<0.005$). Note that the leftmost four data originate from TDLUs and the rightmost data from ducts.

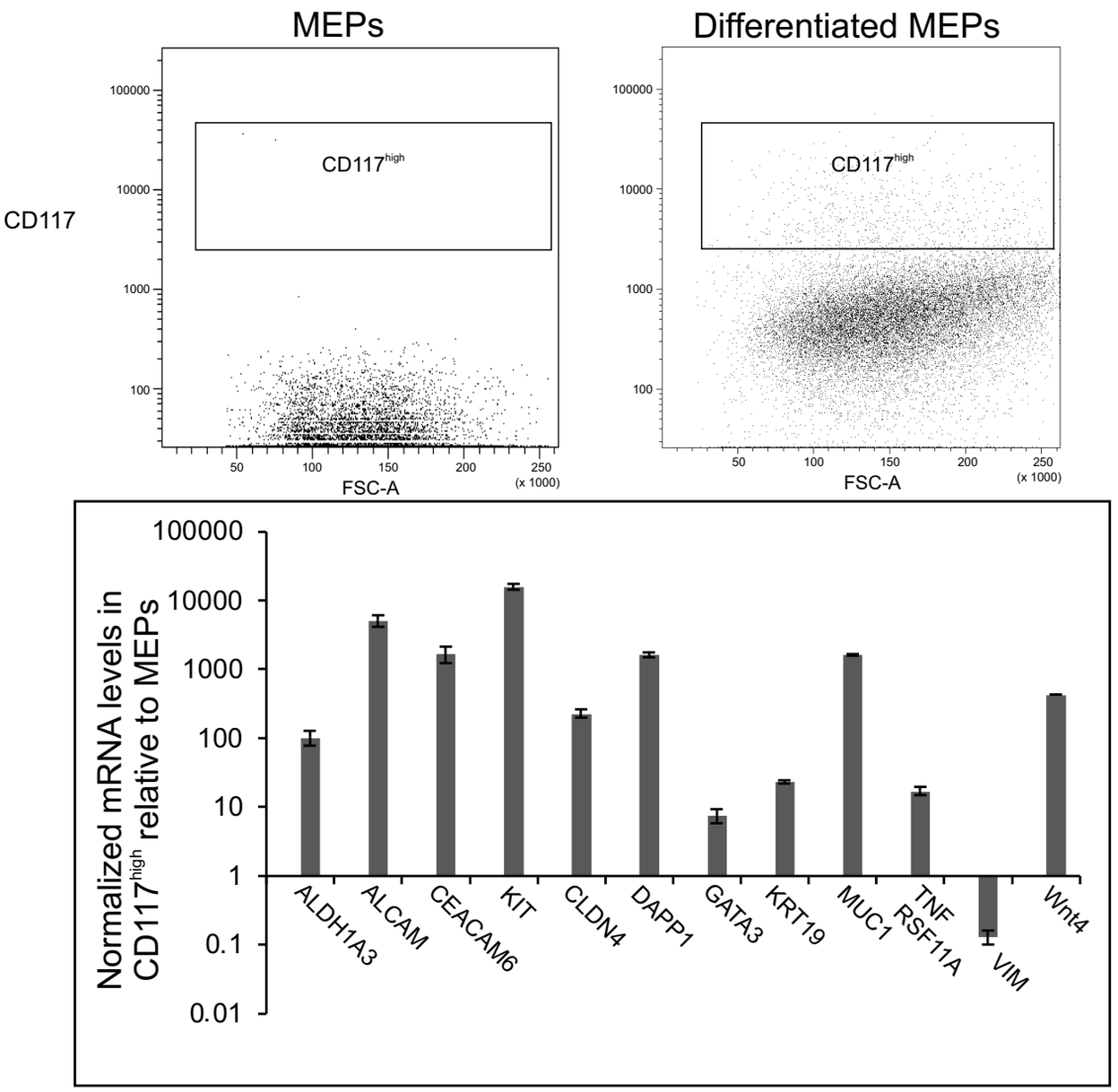

Fig. S9. Differentiation of MEPs induces expression of several luminal markers as evidenced by qPCR. Forward scatter (FSC-A) FACS diagrams show the CD117 expression status in MEPs (left) and differentiated MEPs (right), respectively. Box indicates gate used for isolation of CD117^{high} cells in differentiated MEPs, which are not present in MEPs. Bar diagram depicts mRNA levels normalized to four reference genes for number of luminal markers expressed in CD117^{high} cells as compared to MEPs. CD117^{high} cells express higher levels of luminal markers and a lower expression of vimentin (VIM). (Aldehyde Dehydrogenase 1 Family Member A3 (ALDH1A39), Activated Leukocyte Cell Adhesion Molecule (ALCAM), Carcinoembryonic Antigen Related Cell Adhesion Molecule 6 (CEACAM6), KIT proto-oncogene receptor tyrosine kinase (KIT), Claudin 4 (CLDN4), dual adaptor of phosphotyrosine and 3-phosphoinositides 1 (DAPP1), GATA binding protein 3 (GATA3), Keratin 19 (KRT19), Mucin 1, cell surface associated (MUC1), TNF receptor superfamily member 11a (TNFRSF11A), Vimentin (VIM), Wnt family member 4 (Wnt4)). (Error bars indicate standard deviation of technical triplicates).

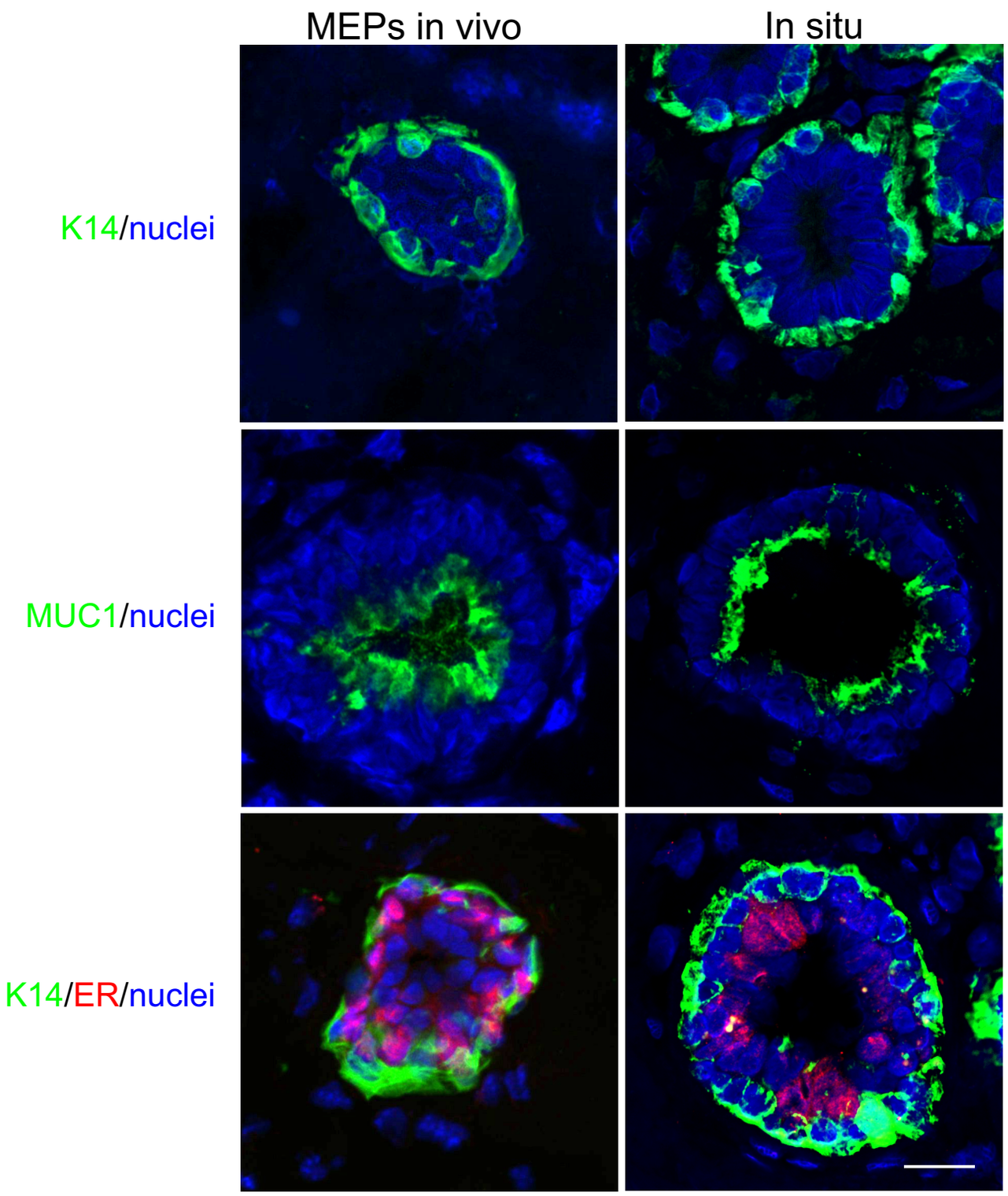

Fig. S10. Transplanted MEPs give rise to structures which histologically resemble human breast acini. Multicolor imaging of cryostat sections of MEPs in vivo (left column) and human breast tissue in situ (right column). Sections were stained by fluorescence for K14 and nuclei (top row), MUC1 and nuclei (middle row), and K14, ER and nuclei (bottom row). MEPs readily gave rise to breast morphogenic units consisting of an inner layer of polarized luminal cells and an outer layer of myoepithelial cells upon transplantation to the NOG mouse mammary gland (Bar = 20μm).

Table S1. List of antibodies used for immunostaining and/or FACS analysis.

Antibody	Clone	Company/Catalogue No.	Peroxidase	Fluorescence	FACS
α-sma	1A4	Sigma, A2547	1:500-1:1000	1:1000	-
CD29, AF488	TS2/16	BioLegend, 303016	-	-	1:50
CD49f, FITC	GoH3	BD Biosciences, 555735	-	-	1:50
CD90	aThy-1A1	Abcam, ab20147	1:100	-	-
CD117	K45	Neomarkers, MS-289-P0	-	-	1:50
CD271	ME20.4	BioLegend, 345102	1:100	-	-
CD271, APC	ME20.4	Cedarlane, CL10013APC	-	-	1:50
CD326, AF488	9C4	BioLegend, 324210	-	-	1:50
CD326, BV421	EBA-1	BD Biosciences, 56180	-	-	1:50
CD326, AF647	9C4	BioLegend, 324212	-	-	1:50
ERα	SP1	Labvision, RM-9101-S	-	1:10-1:25	-
Keratin 5	XM26	Novocastra, NCL-CK5	1:250	-	-
Keratin 7	LPK5	Chemicon, CBL194	-	1:10	-
Keratin 10	DE-K10	Neomarkers, MS-611-P0	-	1:25	-
Keratin 14	LL002	NeoMarkers, MS-115-P	1:50	1:25-1:50	-
Keratin 14	LL002	Monosan, MONX10687	-	1:50	-
Keratin 17	E3	Dako, M7046	1:200	1:25-1:50	-
Keratin 19	BA16	Abcam, ab20210	1:100-1:200	1:50-1:100	-
Keratin 19	BA16	GenWay, GWB-22664E	1:100-1:2000*	1:50-1:100	-
Keratin 19	A53-B/A2	Abcam, ab7754	-	1:100	-
MUC1	115D8	Biogenesis, 1510-5025	-	1:10-1:25	-
p63	7JUL	Novocastra, NCL-L-p63	1:10-1:25	-	-
Podoplanin	D2-40	Dako, M3619	1:25-1:100	-	-
Podoplanin, AF488	NC-08	BioLegend, 337006	-	-	1:50
Vimentin	Vim 3B4	Dako, M7020	1:250	-	-

* For staining of tissue sections following formaldehyde fixation, BA16 against K19 was diluted 1:1000-1:2000. Otherwise, a dilution of 1:100-1:200 was used (methanol fixation).

Table S2. List of primers used for RT-qPCR

Gene symbol	Assay ID	Category
ALDH1A3	**Hs00167476_m1**	**Luminal marker**
CD166	**Hs00977640_m1**	**Luminal marker**
CEACAM6	**Hs03645554_m1**	**Luminal marker**
KIT	**Hs00174029_m1**	**Luminal marker**
CLDN4	**Hs00976831_s1**	**Luminal marker**
DAPP1	**Hs00183937_m1**	**Luminal marker**
GATA3	**Hs00231122_m1**	**Luminal marker**
KRT19	**Hs00761767_s1**	**Luminal marker**
MUC1	**Hs00159357_m1**	**Luminal marker**
TNFRSF11A	**Hs00187192_m1**	**Luminal marker**
VIM	**Hs00185584_m1**	**Basal marker**
WNT4	**Hs01573504_m1**	**Luminal marker**
GAPDH	**Hs02758991_g1**	**Reference**
HPRT1	**Hs99999909_m1**	**Reference**
TBP	**Hs00427621_m1**	**Reference**
PGK1	**Hs00943178_g1**	**Reference**

References

1. Fridriksdottir AJ*, et al.* (2015) Propagation of oestrogen receptor-positive and oestrogen receptor-responsive normal human breast cells in culture. *Nat Commun* 6:8786-8797.
2. Popnikolov NK, Cavone SM, Schulttz PM, & Garcia FU (2005) Diagnostic utility of p75 neurotrphin receptor ($p75^{NTR}$) as a marker of breast myoepithelial cells. *Mod. Pathol.* 18:1535-1541.
3. Villadsen R*, et al.* (2007) Evidence of a stem cell hierarchy in the adult human breast. *J Cell Biol* 177:87-101.
4. Liu X*, et al.* (2012) ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am J Pathol* 180:599-607.
5. Tan DWM*, et al.* (2013) Single-cell gene expression profiling reveals functional heterogeneity of undifferentiated human epidermal cells. *Development* 140:1433-1444.
6. Petersen OW & van Deurs B (1987) Preservation of defined phenotypic traits in short-time cultured human breast carcinoma derived epithelial cells. *Cancer Res* 47:856-866.
7. Rønnov-Jessen L & Petersen OW (1996) A function for filamentous α-smooth muscle actin: Retardation of motility in fibroblasts. *J Cell Biol* 134:67-80.
8. Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25:402-408.