# Supporting Information

## Roy et al. 10.1073/pnas.1218682110

### SI Methods

**Dissociation of Breast Tissue.** The acquisition of human breast tissues from disease-free women undergoing reduction mammoplasty and the experiments using these tissues were approved by the UCSF Committee on Human Research under Institutional Review Board protocol No. 10-01532. The human tissues used in this study were obtained after informed consent. All information obtained about human tissue samples met Health Insurance Portability and Accountability Act (HIPAA) guidelines. All tissues used in this study were devoid of visible disease, bacterial, fungal, or viral contamination and exhibited a normal diploid 46, XX karyotype. Tissue was dissociated mechanically and enzymatically, as previously described (1). Briefly, tissue was minced and dissociated in RPMI 1640 with L-glutamine and 25 mm Hepes (Fisher; MT10041CV) supplemented with 10% (vol/vol) FBS (JR Scientific; 43603), 100 units/mL penicillin, 100 μg/mL streptomycin $SO_4$, 0.25 μg/mL fungizone, gentamycin (Lonza; CC4081G), 0.88 mg/mL collagenase (Worthington; CLS-2), and 0.40 mg/mL hyaluronidase (Sigma; H3506-SG) at 37 °C for 16 h. The cell suspension was centrifuged at 394 × *g* for 10 min followed by a wash with RPMI 1640/10% FBS. Clusters enriched in epithelial cells (referred to as organoids) were recovered after serial filtration through a 150-μm nylon mesh (Fisher; NC9445658), and a 40-μm nylon mesh (Fisher; NC9860187). The final filtrate contained the mammary stromal cells, consisting primarily of fibroblasts and endothelial cells. Following centrifugation at 290 × *g* for 5 min, the epithelial organoids and filtrate were frozen for long-term storage. To generate single-cell suspensions, epithelial organoids were further digested for 5 min in 0.5 g/L trypsin-0.2 g/L EDTA-0.58 g/L $NaHCO_3$ and 1 min in dispase-DNase I (StemCell Technologies; 7913 and 7900, respectively) then filtered through a 40-μm cell strainer (Fisher; 087711). *Mycoplasma* testing was performed by PCR analysis at Bionique Testing Laboratories (Saranac Lake, NY).

**Karyotypic Analysis.** Karyotyping was carried out at Molecular Diagnostic Services (San Diego). Briefly, primary breast cells obtained from the filtrate fraction described above and their corresponding R1 cells were allowed to grow to 80% confluency. Mitotic division was arrested by treating cells with 75 ng/mL colcemid for 18.5 h. Following treatment, cells were harvested with trypsin-EDTA, treated with a hypotonic solution, and fixed in methanol/acetic acid. Metaphase spreads were prepared from fixed cells and stained to observe chromosomal G bands. For each tissue sample, 20 metaphase spreads were counted, 5 of which were analyzed and karyotyped. Representative karyotypic analyses from two different donor samples are shown in Fig. 5 and Fig. S2. In total, six independent donor samples were karyotyped. All samples analyzed yielded a diploid 46, XX karyotype (Fig. 5 and Fig. S2).

**ALEXA-Seq Database Mining for Selective Cell Surface Markers Regulated by Repression of $p16^{INK4a}$.** Normalized gene expression levels for 13 cell surface markers in human breast luminal, myoepithelial, or naturally repressed $p16^{INK4a}$ (variant human mammary epithelial cells, vHMECs) cells were queried using the publicly available ALEXA-Seq database (Genome Sciences Center, Vancouver, BC, Canada) at http://www.alexaplatform.org/alexa_seq/. Briefly, gene expression analysis was based on alignment of short sequence reads (30–100 bp in length) generated by massive parallel RNA sequencing, a method that allows identification and quantification of both known and novel mRNA isoforms genome-wide (2).

**Plasmids and Retroviral Gene Transfer.** Lentiviral suspensions for short hairpin *$p16^{INK4a}$* and nontargeting control were collected from transfected 293T cells as previously published (3). Mammary epithelial cells expressing $p16^{INK4a}$ were transduced by exposing them to lentiviral suspensions in the presence of 4 μg/mL polybrene (Sigma-Aldrich) for 5 h. This step was repeated 24 h later to increase transduction efficiency. Cells were maintained in the appropriate medium for 72 h after initial transduction and then selected in the presence of 2 μg/mL puromycin (Sigma). Cells were expanded in primary mammary epithelial medium (MEBM, Lonza; CC-3151, supplemented with MEGM SingleQuots, Lonza; CC-4136) for an additional 72 h, trypsinized, and both total RNA and cell pellet were isolated for q-PCR and flow cytometry analysis, respectively. q-PCR was performed using a primer probe set for *$p16^{INK4a}$* (custom probe ID: 4331348) obtained from ABI. Glucuronidase B (*GusB*; IDT) expression was used to normalize for variances in input cDNA. The cell pellet was analyzed for expression of CD73 and CD90 cell surface markers by flow cytometry using antibodies as described below.

**Flow Cytometry Staining for Cell Sorting.** The single-cell suspension obtained as described above was stained for cell sorting with two human-specific primary antibodies, anti-CD73 labeled with PE (phycoerythrin) (BD Biosciences; 550257) and anti-CD90 labeled with APC (allophycocyanin) (BD Biosciences; 559869) and biotinylated antibodies for lineage markers, anti-CD2, -CD3, -CD16, -CD64 (BD Biosciences; 555325, 555338, 555405, and 555526), -CD31 (Invitrogen; MHCD3115), -CD45 and -CD140b (BioLegend; 304003 and 323604) to specifically remove hematopoietic, endothelial, and leukocyte lineage cells. Sequential incubation with primary antibodies was performed for 20 min at room temperature in PBS with 1% BSA, followed by washing in PBS with 1% BSA. Biotinylated primary antibodies were revealed with an anti-human secondary antibody labeled with streptavidin-Pacific Blue conjugate (Invitrogen; S11222). Cell sorting was performed using a FACSAria II cell sorter (BD Biosciences). All cell populations exhibited a normal diploid 46, XX karyotype.

**Quantitative Real-Time PCR for Pluripotency-Associated Genes.** Total RNA was extracted from sorted R1–R4 subpopulations, R1 colonies grown on feeder layers, R1 colonies grown in expansion media (α-MEM medium with glutamine supplemented with 15% (vol/vol) ES-FBS (Omega Scientific; FB-05), 18% (vol/vol) Chang B, and 2% (vol/vol) Chang C (Irvine Scientific; C-100 and C-106, respectively) and 1× penicillin/streptomycin), undifferentiated human embryonic stem cells (hESCs) H7 and H9, human induced pluripotent stem cells (iPSCs), or human mesenchymal stem cells (MSCs) using the PicoPure RNA Isolation kit (Molecular Devices; KIT0204). RNA purity and concentration were determined using a 2100 Bioanalyzer (Agilent Technologies). Quantitative real-time PCR was performed using 7 ng input RNA on a Custom Human RT2 Profiler PCR Array (Qiagen) following the manufacturer's instructions. *P* values were generated using Student *t* test with the online software provided along with the RT2 Profiler PCR Array (Qiagen). The human iPSC RNA was a generous gift from Paul Cheng (University of California San Francisco, UCSF).

**Mammosphere Culture.** Mammosphere culture was performed as previously described (4). Single cells from R1–R4 subpopulations were directly sorted into 96-well ultra-low attachment plates (Corning; 3474) at a density of 1 cell per well or plated in ultra-low attachment plates (Corning; 3473) at a density of 10,000 viable

cells/mL in primary culture and 1,000 cells/mL in subsequent passages. For mammosphere culture, cells were grown in a serum-free mammary epithelial basal medium (MEBM) (Lonza; CC-3151), supplemented with B27 (Invitrogen; 17504044), 20 ng/mL EGF (Lonza; CC-4017G), 20 ng/mL bFGF (Sigma; F0291-25UG), and 4 μg/mL heparin (Sigma; H1027). Mammospheres were collected by gentle centrifugation (99 × *g*) after 7–10 d in culture and dissociated enzymatically for 5–10 min in 0.5 g/L trypsin-0.2 g/L EDTA-0.58 g/L $NaHCO_3$. Dissociated cells were passed through a 40-μm sieve, stained with 0.4% Trypan Blue solution (Sigma; T8154) to assess cell viability and analyzed microscopically to confirm complete cell dissociation.

**Culture Conditions for in Vitro Multilineage Mammary Differentiation.** Mammospheres derived from R1–R3 subpopulations were dissociated into single cells and cultured in suspension to test for subsequent mammosphere-forming capacity and multilineage potential by plating on collagen-coated coverslips at colony-producing densities (2,000 viable cells/10 cm diameter dish) (Fig. S3). Cells were grown in Ham's F-12 medium with 5% (vol/vol) FBS, 5 μg/mL insulin (Lonza; CC-4021G), 1 μg/mL hydrocortisone (Lonza; CC-4031G), 10 μg/mL cholera toxin (Sigma; C8052-2MG), 10 ng/mL EGF (Lonza; CC-4017G), and gentamycin (Lonza; CC-4081G). After 5 d, a layer of Matrigel (BD Biosciences; 356230) was added along with 1 μg/mL prolactin (Sigma; L4021-50UG) in the case of assays for alveolar differentiation. Cells were fixed and collected for immunostaining after 12 d. Single cells from dissociated mammospheres (R1–R3) were cultured at colony-producing densities in 3D Matrigel. 3D cultures were performed as previously described (5). Briefly, single cells resuspended in the above growth medium were seeded at colonogenic density on a 1–2-mm thick solidified layer of growth factor reduced Matrigel (BD Biosciences; 356230). Acinar and branched-acinar structures forming in Matrigel were photographed and immunoblot analysis performed after 14 d.

**Immunostaining and Immunoblotting.** To assess the lineage composition of the colonies in Fig. 1*D* and Fig. S3*B*, lineage composition was assayed via flow cytometry after cell trypsinization from coverslips and fixation for 20 min at room temperature in 2% (vol/vol) PFA solution in PBS. Samples were stained with the anti-CD49f/α-6-integrin coupled to FITC and MUC-1 (mucin 1) primary antibodies for 20 min at room temperature in PBS containing 1% BSA followed by an incubation with a secondary goat anti-mouse IgG1 antibody labeled with Tricolor (Invitrogen; M32006). After incubation, cells were washed once with PBS containing 1% BSA. Flow cytometry analysis, using a BD LSRII flow cytometer (BD Biosciences), enabled us to determine the distribution of different mammary epithelial cell types (by percentage) from mammosphere-derived cells in successive passages: R1 (passages 1–3): myoepithelial ($CD49f^+/MUC\text{-}1^-$): 11.43 ± 0.48, 1.02 ± 0.18, and 1.42 ± 0.19; luminal ($CD49f^-/MUC\text{-}1^+$): 60.22 ± 1.57, 62.57 ± 1, and 23.62 ± 0.91; bipotent ($CD49f^+/MUC\text{-}1^+$): 20.32 ± 0.63, 31.81 ± 0.76, and 65.54 ± 1.08; R2 (passages 1–2): myoepithelial: 15.45 ± 0.94 and 20.34 ± 0.94; luminal: 81.23 ± 0.49 and 75.99 ± 0.57; bipotent: 0.26 ± 0.07 and 0.11 ± 0.07; R3 (passages 1–2): myoepithelial: 0.07 ± 0.02 and 0.04 ± 0.07; luminal: 94.92 ± 0.57 and 96.6 ± 0.69; and bipotent: 4.01 ± 0.69 and 2.32 ± 0.18. Data from five tissue samples are expressed as average ± SEM ($n = 5$). Cells grown on coverslips were also fixed for 20 min at room temperature in PBS + 2% (vol/vol) paraformaldehyde (PFA) then stained with primary anti-CD49f coupled to FITC (BD Biosciences; 555735) and anti–MUC-1 (Millipore; 05–652), which were used as myoepithelial and luminal markers, respectively. A secondary antibody labeled with Alexa Fluor 568 (Invitrogen; A-11031) was used to detect binding of anti–MUC-1. Nuclei were counterstained and mounted with DAPI/antifade (Invitrogen; P36935). Coverslips were examined with a fluorescent microscope (Zeiss LSM 510 NLO) with a 20× objective and analyzed using LSM Image Browser software.

Differentiation toward the alveolar lineage was assessed in lysates of cells that had been layered with Matrigel and prolactin. Briefly, cells were pelleted at 453 × *g* for 3 min at 4 °C and washed once in ice-cold wash buffer (25 mM Tris, pH 7.5, 250 mM sucrose, 2.5 mM $MgCl_2$, 10 mM benzamidine, 10 mM NaF, 1 mM sodium vanadate, 10 μg/mL leupeptin, 10 μg/mL aprotinin, 1 μg/mL pepstatin, and 1 mM PMSF). Pellets were resuspended in lysis buffer (20 mM Hepes·KOH, pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 250 mM sucrose, 1 mM sodium vanadate, 1 mM DTT, 25 μg/mL leupeptin, 25 μg/mL aprotinin, 2.5 μg/mL pepstatin, 1 mM PMSF, 10 mM benzamidine, and 20 mM NaF). Protein concentration of Dounce homogenized cell extracts was determined with bicinchoninic acid (Pierce Biotechnology) using BSA as a standard (Sigma). Cell extracts were heat denatured in loading buffer containing 5% (vol/vol) β-mercaptoethanol and separated by gradient (4–20%) polyacrylamide gel electrophoresis (Cambrex). Proteins were transferred onto Hybond-P membranes (GE Healthcare Bio). Membranes were probed with mouse monoclonal antibodies against human β-casein (Santa Cruz Biotechnology; SC-53189), or mouse anti–β-actin (Sigma; AC-15), followed by a horseradish peroxidase-conjugated goat anti-mouse antibody (Biomeda). β-Actin was used as a normalization loading control. Staining was developed with the SuperSignal West Pico chemiluminescence detection kit (Pierce).

**Animal Model for in Vivo Multilineage Mammary Differentiation.** Female nonobese diabetic (NOD)/SCID mice were used to assess the in vivo mammary reconstitution properties of the sorted R1–R4 epithelial subsets from three disease-free breast tissue samples. This animal model used for xenotransplantation of normal mammary epithelial cells has been previously described by Proia et al. (6). The mammary fat pads were cleared prepuberty and humanized by injecting 35 μL of a 1:1 mixture of irradiated (4Gy) and nonirradiated immortalized primary human mammary fibroblasts (500,000 cells total per fat pad). Fibroblasts, immortalized with human telomere reverse transcriptase (hTERT) and GFP (RMF/EG), were a generous gift from Charlotte Kuperwasser (Tufts University School of Medicine, Boston). Sorted epithelial cells were mixed with 500,000 RMF/EG fibroblasts in 35 μL of a 1:3 mixture of Matrigel-Collagen I (BD Biosciences) and implanted in the mammary fat pads 2–4 wk after clearing and humanization. Animals were euthanized 12 wk after cell injection and ductal outgrowths were analyzed. Human β-casein production was examined in animals injected with sorted cells, mated, and euthanized at day 18 of pregnancy. Fat pads were fixed in formalin and embedded in paraffin for histological analysis. Evaluation of the outgrowth potential of each cell population was analyzed by H&E staining. The animal studies were conducted in accordance with the UCSF Institutional Animal Care and Use Committee (IACUC) approved animal protocol AN086757/AN079997.

**Immunohistochemistry and Immunofluorescence.** Immunohistochemistry was performed on formalin-fixed paraffin-embedded tissues. Five-micrometer thick sections were deparaffinized, rehydrated through graded alcohols, and subjected to antigen retrieval. Sections were incubated with mouse monoclonal antibodies against anti-human smooth muscle actin (α-SMA) (Dako; M0851) and anti-human cytokeratin 8/18 (CK8/18) (Leica Microsystems; RTU-5D3), and a rabbit polyclonal antibody against anti-human β-casein (a generous gift from Charles Streuli, University of Manchester, Manchester, UK). Immunocomplexes were visualized by the avidin-biotin-peroxidase complex (ABC) method and sections were counterstained with hematoxylin. Stained sections were scanned at 20× magnification for image acquisition on a digital slide scanner (Aperio). Scanned images were viewed using

the Aperio's Spectrum software and regions of interest (ROI) were chosen at different magnifications. For fluorescent double staining, samples were incubated for 2 h at room temperature with Alexa Fluor 594 (Invitrogen; A11020) and Alexa Fluor 488 (Invitrogen; A21121) labeled secondary antibodies diluted 1/500. Nuclei were counterstained with Vectashield-DAPI and coverslipped. Immunostained sections were examined with a fluorescent microscope (Zeiss LSM 510 NLO) with a 20× objective and analyzed using LSM Image Browser software.

**In Vitro Endodermal Lineage Differentiation for Definitive Endoderm and Pancreatic Derivatives.** Sorted R1–R4 subsets were cultured under conditions previously reported to induce pancreatic lineage in human ESCs (7). Sorted cells seeded on feeder layer were subjected to either 3 d (definitive endoderm) or 12 d (pancreatic lineage) of differentiation. Flow cytometric C-X-C chemokine receptor type 4 (CXCR4) enrichment was not performed in these experiments to facilitate endoderm differentiation. Immunofluorescence analysis was performed with rabbit anti-human SOX17 (sex determining region Y-box 17) (Santa Cruz Biotechnology; SC-20099), goat anti-human HNF-3β/FOXA2 (forkhead box A2) (R&D Systems; AF2400), goat anti-human Brachyury (R&D Systems; AF2085), goat anti-human PDX1 (pancreatic and duodenal homeobox 1) (Santa Cruz Biotechnology; SC14662), and goat anti-human NKX6.1 (NK6 homeobox 1) (Santa Cruz Biotechnology; SC15030) primary antibodies followed by donkey anti-rabbit coupled to Alexa Fluor 546 (Invitrogen; A10040) and donkey anti-goat coupled to Alexa Fluor 488 (Invitrogen; A11055) secondary antibodies to ascertain differentiation toward endodermal lineage.

**In Vitro Mesodermal Lineage Differentiation for Cardiomyocyte, Adipocyte, and Endothelial Cell Derivatives.** To differentiate the sorted R1 subset into spontaneously beating cardiomyocytes, sorted cells were first grown on human placental fibroblast feeders (a generous gift from Susan Fisher, UCSF) in serum-free mammosphere medium (as described above). After 14 d, colonies appearing on the feeder layer were manually dissected and allowed to form embryoid bodies (EBs) in suspension in cardiac differentiation medium containing knockout DMEM (Invitrogen), 20% (vol/vol) FBS, nonessential amino acids, glutamine and β-mercaptoethanol. After 4 d in suspension, EBs were plated onto gelatin-coated 24-well plates and fed fresh medium every day. Monitoring of beating EBs was carried out using time-lapse video microscopy in an environmental chamber controlled by the Improvision's Open laboratory software in real time. Approximately 5% of EBs began beating after 12–14 d of culture. Differentiated cells were fixed in ice-cold methanol and analyzed by immunofluorescence with primary antibodies obtained from Santa Cruz Biotechnologies, including mouse monoclonal IgG2a anti-human GATA-4 (GATA binding protein 4) (SC-25310), goat polyclonal anti-human MEF-2 (myocyte enhancer factor-2) (SC-13917), and rabbit polyclonal anti-human Nkx2.5 (NK2 homeobox 5) (SC-14033). Staining was completed with secondary antibodies including goat anti-mouse IgG2a coupled to Alexa Fluor 488 (Invitrogen; A21131), donkey anti-goat IgG coupled to Alexa Fluor 488 (Invitrogen; A11055), and goat anti-rabbit IgG coupled to Alexa Fluor 488 (Invitrogen; A11008). Slides were examined with a fluorescent microscope (Zeiss LSM 510 NLO) with a 20× objective and analyzed using the LSM Image Browser software.

Adipocyte differentiation: sorted R1–R4 cells were expanded for 2 wk in α-MEM medium with glutamine supplemented with 15% (vol/vol) ES-FBS (Omega Scientific; FB-05), 18% (vol/vol) Chang B, and 2% (vol/vol) Chang C (Irvine Scientific; C-100 and C-106, respectively), and 1× penicillin/streptomycin. Cells were then seeded into 24-well chamber slides and placed under growth conditions (expansion medium) or differentiation conditions (Gibco StemPro Adipogenesis Differentiation kit; A10070-01) for 9 d to assay for Oil Red O staining and quantitative real-time PCR analysis. Media was changed every 3–4 d. Quantitative real-time PCR was performed using standard methods. Primer probe sets for *FABP4* (fatty acid binding protein 4, adipocyte) (Hs01086177_m1), *LEPTIN* (Hs00174877_m1), and *PPARγ* (peroxisome proliferator-activated receptor gamma) (Hs01115511_m1) were obtained from ABI. Glucuronidase B (*GusB*; IDT) expression was used to normalize for variances in input cDNA.

Sorted R1–R4 subsets were cultured under conditions previously described to induce endothelial cell differentiation (8). Briefly, cells were cultured in endothelial medium as previously described and analyzed for expression of the CD31/PECAM1 (platelet endothelial cell adhesion molecule 1) cell surface marker by flow cytometry after 2 wk. R1 cells yielded 2% CD31/PECAM1$^+$ cells when cultured under these conditions. The CD31/PECAM1$^+$ cells were isolated by flow cytometry and seeded at 50,000 cells per 500 μL of culture medium in a Matrigel differentiation assay. Cord formation was evaluated by phase contrast microscopy 24 h after seeding the cells. HUVECs (human umbilical vein endothelial cells) and primary mammary epithelial cells were used as positive and negative differentiation controls, respectively.

**Teratoma and Mammary Tumor Formation Assay.** Six- to 7-wk-old immunocompromised female SCID/beige mice (Charles River) were used to test teratoma forming capability of MDA-MB231 breast cancer cells, freshly sorted R1 cells, culture expanded R1 subclones, and H7 human ESCs. Cells were grafted under the renal capsule according to a published protocol (9). Mice were euthanized 8 wk (H7 cells) or 16 wk (R1 cells) after injection. The ePS-cell–derived teratomas were much smaller in volume (~10 $mm^3$) than the teratomas derived from hESCs (2 $cm^3$) when the growths were harvested at 16 wk vs. 8 wk, respectively. These differences in size may be due to several factors including: the initiating cell number (the number of R1 cells injected was 200,000–700,000 vs. 2.8–3.7 million for hESCs because of the difficulty in isolating a large number of R1 cells due to their rarity in primary human tissue), cell doubling time (~40 h for ePS cells vs. ~30 h for hESCs) and mortality (unlike hESCs, the ePS cells are mortal and therefore posses a limited number of population doublings). Of particular note, only ePS cells at early passages (up to passage 5) formed teratomas, whereas cells used at later passages failed to do so. Teratomas were surgically removed, fixed in formalin, embedded in paraffin and processed for immunohistochemistry as described below.

**Human Tissue Samples.** Disease-free breast tissues were obtained from the Cooperative Human Tissue Network (Nashville, TN) and the University of Pennsylvania (Philadelphia), Kaiser Foundation Research Institute (Oakland, CA), and the Comprehensive Breast Health Center Aurora Sinai Medical Center (Milwaukee) in accordance with the UCSF Committee on Human Research under Institutional Review Board protocol 10–01532. Samples were identified through unlinked codes in accordance with federal HIPAA guidelines.

**Histochemistry and Immunohistochemistry for Teratoma and Human Mammary Tumor Analysis.** Tissues included teratomas, MDA-MB231 tumors, human disease-free breast tissues, and various human tissues used as positive controls or mouse kidney used as a negative control. Paraffin-embedded mammary tumors and teratomas were cut into 4-μm serial sections, deparaffinized, and rehydrated using standard procedures. All steps were carried out at room temperature except when noted. Following antigen retrieval by microwaving in citrate buffer pH 6.0 for 10 min, sections were incubated with primary antibodies against lamin A/C (Epitomics; 2966–1, clone EPR4100), GFAP (Dako; M0761),

HAPLN1 (Sigma-Aldrich; HPA019105), PDX1 (Epitomics; 3470–1, clone EPR3358), AFP (ready to use; Dako; IR500), and TFF3 (Epitomics; 3178–1, clone EPR3973) for 1 h. Staining was visualized after incubation with the HRP polymer kit (Ultravision LP kit, Thermo Scientific) for 30 min and with diaminobenzidine substrate (Genemed; 520017) for 5 min. For mouse kidney sections stained with the mouse monoclonal anti-GFAP antibody, an additional peroxidase blocking step (3% H202 for 10 min) was added before antigen retrieval and the Mouse on Mouse kit (Vector Laboratories; BMK2202) was used instead of the Ultravision LP kit. For Alizarin Red staining, the slides were briefly rinsed with water and stained for 2 min with shaking. Slides were dipped in acetone (20 dips), followed by xylene (20 dips) and coverslipped. Stained sections were scanned at 20× magnification for image acquisition on a digital slide scanner (Aperio). Scanned images were viewed using the Aperio's Spectrum software and were chosen at different magnifications. Staining controls are shown in Fig. S6.

**Single-Cell–Derived R1 Clone Cell Culture.** Single R1 cells were expanded on human placental fibroblasts used as a feeder layer (a generous gift from Susan Fisher, UCSF). Fibroblasts from human placenta (at week 6.4 of gestation) were grown in media containing DMEM, M199 (Invitrogen; 11150–059) and 10% (vol/vol) FBS, gamma irradiated at 5,100 rads and frozen for long-term storage. Before use, feeders were thawed at 37 °C, washed, and plated on gelatin-coated 24-well tissue culture plates. Single R1 cells were plated on feeders 24 h after feeders were seeded and cultured for up to 14 d with medium that contained serum-free mammary epithelial basal medium (MEBM) (Lonza; CC-3151), supplemented with B27 (Invitrogen; 17504044), 20 ng/mL EGF (Lonza; CC-4017G), 20 ng/mL bFGF (Sigma; F0291-25UG), and 4 μg/mL heparin (Sigma; H1027). Cultures were monitored daily to confirm that colonies came from a single R1 cell. Single colonies obtained from individual R1 cells were manually dissociated, trypsinized, and split in three parts to probe for the differentiation potential of the subclones according to the procedures mentioned above.

**Human Embryonic Stem Cell Culture.** Human embryonic stem cells (hESCs) H7 and H9, routinely maintained in culture with replacement of frozen stocks every 10 passages (a generous gift from Susan Fisher, UCSF), were expanded on mouse embryonic fibroblasts (MEFs) used as a feeder layer (Millipore; PMEF-CFL). The fibroblasts were grown in media containing DMEM, M199 (Invitrogen; 11150–059) and 10% (vol/vol) FBS. The feeders were gamma irradiated at 5,100 rads and frozen for long-term storage. Feeders were thawed at 37 °C, washed, and plated on gelatin-coated six-well tissue culture plates. hESCs H7 and H9 were thawed, washed, and plated onto confluent feeder cells for up to a week. The hESC culture medium contained knockout DMEM, 20% (vol/vol) knockout serum replacement (Gibco; 10828–028), 10 ng/mL bFGF nonessential amino acids, glutamine, β-mercaptoethanol, and penicillin/streptomycin. Cultures were monitored daily to confirm that clusters of hESCs were adhering to the feeders and spreading out into typical hESC colonies and to determine passaging. When colonies reached an average size of 300–400 cells, colonies were manually dissected and passaged.

**Human Mesenchymal Stem Cell Culture.** Human mesenchymal stem cells (Lonza; PT-2501) from four different donors were seeded at a recommended density of 5,000–6,000 cells per centimeter squared and fed 3–4 d after seeding with MSCGM medium (Lonza; PT-3001) and subcultured according to the manufacturer's instructions.

**Expansion of a Single-Cell–Derived R1 Clone and Cell Cycle Analysis.** R1 sorted cells were expanded for 2 wk in α-MEM medium with glutamine supplemented with 15% (vol/vol) ES-FBS (Omega Scientific; FB-05), 18% (vol/vol) Chang B, and 2% (vol/vol) Chang C (Irvine Scientific; C-100 and C-106, respectively), and 1× penicillin/streptomycin. Cells were then trypsinized and plated at limiting dilution to generate single-cell–derived subclones. Single-cell–derived colonies obtained after 2 wk in culture were trypsinized using cloning rings and expanded again for a week to generate a growth curve (Fig. 5*B*). Population doublings were calculated using the following equation: PD = log (*A*/*B*)/log 2, where *A* is the number of cells collected and *B* is the number of cells plated initially.

Cells were metabolically labeled with 10 mmol/L bromodeoxyuridine (BrdU) for 4 h before harvest. Cells were isolated by standard trypsinization, resuspended in PBS, and fixed by addition of ice-cold ethanol to a final concentration of 70% (vol/vol). Nuclei were isolated and stained with propidium iodide and a FITC-conjugated anti-BrdU antibody (BD Biosciences) (3). Flow cytometry was carried out on a LSRII cytometer (BD Biosciences) using FlowJo software for analysis. All analyzed events were gated to remove debris and aggregates. A minimum of 20,000 events were collected for each analysis.

**Telomerase Reverse Transcriptase Expression Measurement and Telomerase Activity Assay.** Human telomerase reverse transcriptase (hTERT) expression levels were assessed by q-PCR using a primer probe set for *hTERT* (Hs00162669_m1) obtained from ABI. Glucuronidase B (*GusB*; IDT) expression was used to normalize for variances in input cDNA. Telomerase activity was assayed using a highly sensitive and nonisotopic version of the telomeric repeat amplification protocol (TRAP) assay, i.e., the fluorescence-based TRAPeze XL Telomerase detection kit (Millipore). Lysates (1,000 cell equivalents) from 184A1 (human mammary cell line), Wi-38 (human fibroblast line), HeLa, H7 hESCs, and single-cell–derived R1 subclones from PDs 44.5 and 55.2 were mixed with TRAPeze XL reaction mix containing Amplifuor primers and incubated at 30 °C for 30 min. Amplified fluorescently labeled telomerase products were quantitated with a fluorescence plate reader. Telomerase activity, expressed as total products generated (TPG) units, was calculated by comparing the ratio of telomerase products to an internal standard for each lysate, as described by the manufacturer.

**DNA Fingerprinting.** DNA fingerprinting (short tandem repeat, STR analysis) was carried out at Molecular Diagnostic Services (San Diego). Three nanograms of genomic DNA isolated from each cell population was amplified using the PowerPlex 1.2 or CellID short tandem repeat genotyping system (Promega) according to the manufacturer's instructions. DNA amplification was performed on an Applied Biosystems 2720 thermocycler. Following amplification, reactions were denatured with Hi-Di formamide and resulting fragments were separated and detected on an ABI Prism 3130 capillary electrophoresis platform with POP7 polymer (Applied Biosystems). Analysis and allelic assignment of the respective loci were performed using the GeneScan and Genotyper (Applied Biosystems) and the PowerTyper 12 macro (Promega) software packages.


1. Romanov SR, et al. (2001) Normal human mammary epithelial cells spontaneously escape senescence and acquire genomic changes. *Nature* 409(6820):633–637.
2. Griffith M, et al. (2010) Alternative expression analysis by RNA sequencing. *Nat Methods* 7(10):843–847.
3. Zhang J, Pickering CR, Holst CR, Gauthier ML, Tlsty TD (2006) p16INK4a modulates p53 in primary human mammary epithelial cells. *Cancer Res* 66(21):10325–10331.
4. Dontu G, et al. (2003) In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* 17(10):1253–1270.

5. Debnath J, Muthuswamy SK, Brugge JS (2003) Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* 30(3):256–268.
6. Proia DA, Kuperwasser C (2006) Reconstruction of human mammary tissues in a mouse model. *Nat Protoc* 1(1):206–214.
7. Kroon E, et al. (2008) Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat Biotechnol* 26(4):443–452.
8. Levenberg S, Golub JS, Amit M, Itskovitz-Eldor J, Langer R (2002) Endothelial cells derived from human embryonic stem cells. *Proc Natl Acad Sci USA* 99(7): 4391–4396.
9. Prokhorova TA, et al. (2009) Teratoma formation by human embryonic stem cells is site dependent and enhanced by the presence of Matrigel. *Stem Cells Dev* 18(1): 47–54.

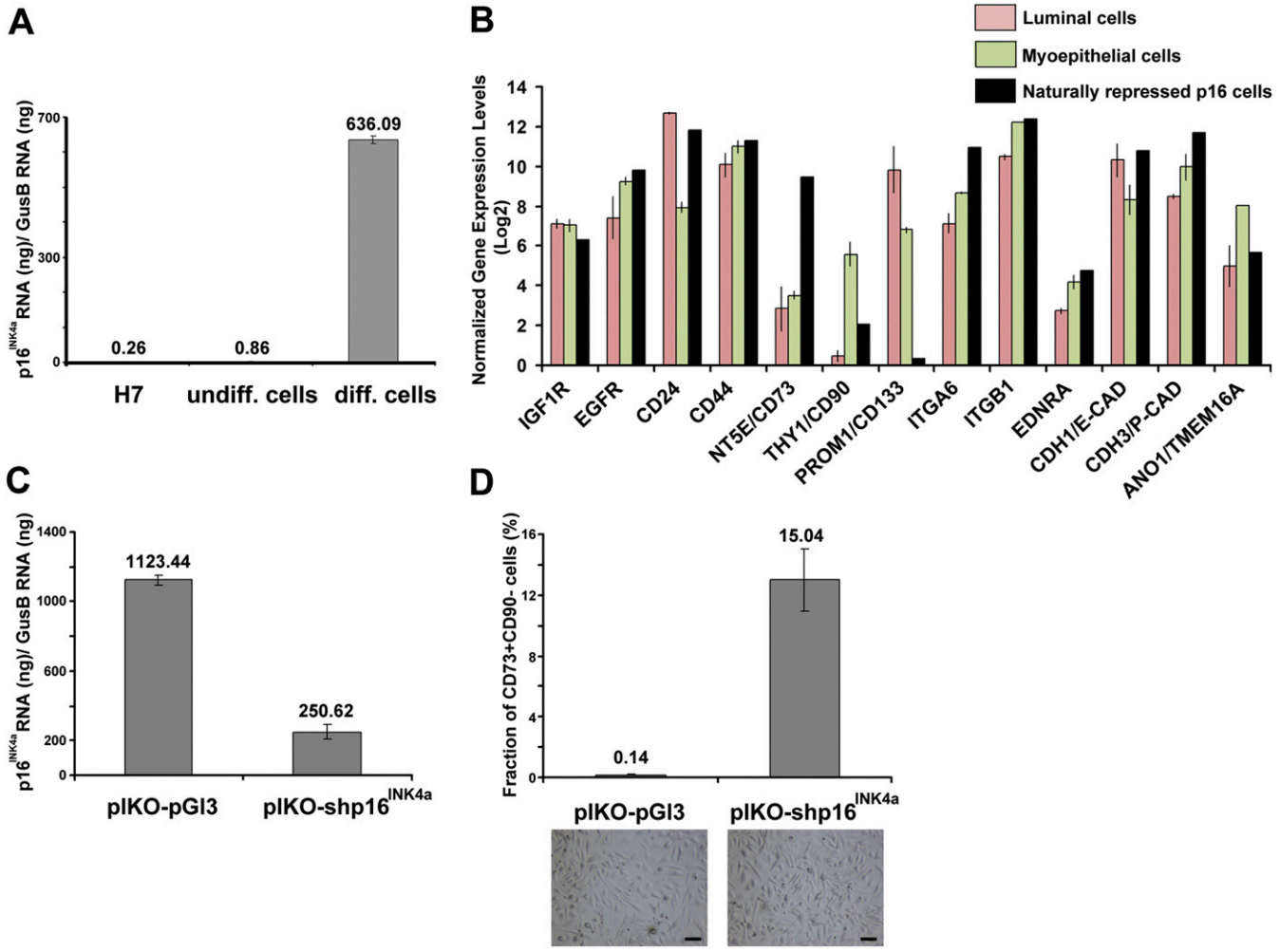



**Fig. S1.** Repression of *p16*$^{INK4a}$ and modulation of expression of cell surface markers CD73 and CD90. (*A*) Transcript levels of *p16*$^{INK4a}$ were measured in human embryonic stem cells, hESCs (H7), undifferentiated ($CD73^{+}CD90^{-}$) cells (undiff. cells), and human mammary differentiated myoepithelial cells (diff. cells). Expression levels were normalized to the housekeeping gene *GusB*. (*B*) Normalized gene expression levels for 13 cell surface markers in human breast luminal, myoepithelial, or naturally repressed *p16*$^{INK4a}$ cells were queried using the publicly available ALEXA-Seq database (1). *NT5E/CD73* and *THY1/CD90* were chosen for subsequent experiments. (*C*) Transcripts levels of *p16*$^{INK4a}$ were measured in human mammary epithelial cells, expressing CD90 but not CD73, transduced with control vector (plKO-pGl3) or short hairpin to *p16*$^{INK4a}$ (plKO-shp16$^{INK4a}$). Expression levels were normalized to the housekeeping gene *GusB*. (*D*) Percentage of $CD73^{+}CD90^{-}$ cells in the populations described above measured by FACS analysis. Phase contrast images of cells transduced with plKO-pGl3 and plKO-shp16$^{INK4a}$ vectors. (Scale bars, 100 μm.) Error bars indicate SDs ($n = 3$).

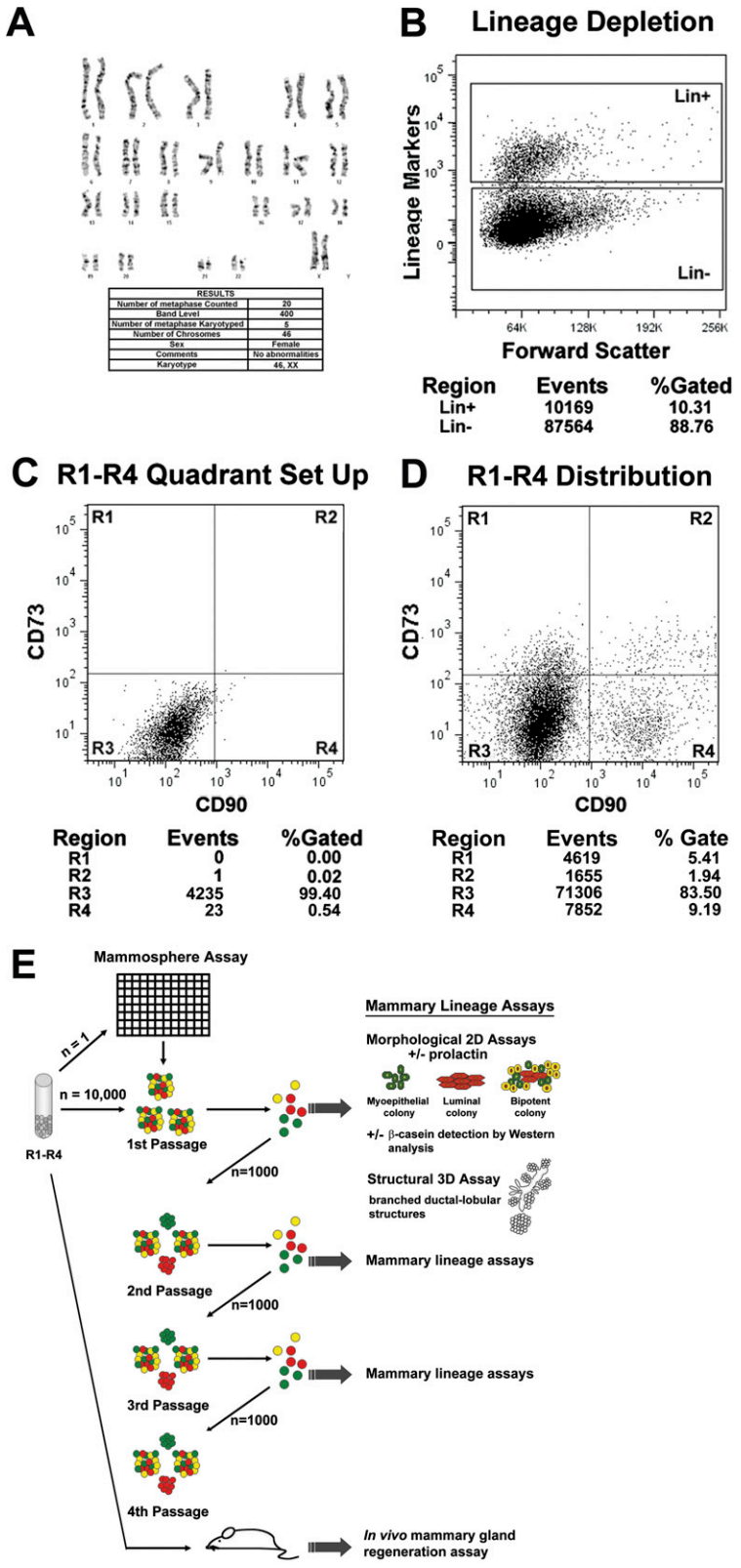



**Fig. S2.** Isolation and fractionation of human mammary epithelial subpopulations using cell surface marker-based FACS sorting. Human mammary epithelial subpopulations were characterized in 10 reduction mammoplasty samples. Representative results are shown for one sample. (*A*) Karyotype analysis exhibiting diploid 46, XX (normal) status. (*B*) FACS analysis of lineage cell surface markers CD2, CD3, CD16, CD31, CD45, CD64, and CD140b expression on human mammary single-cell gated populations. Cells negative or positive for the above markers were defined as the lineage-negative ($Lin^-$) and the lineage-positive ($Lin^+$) populations, respectively. (*C*) Representative FACS analysis of $Lin^-$ unstained single cells to establish the gates for CD73 and CD90 (R1–R4) subpopulations. (*D*) Representative scatterplot for R1–R4 subpopulations. (*E*) Experimental design for assessing the self-renewal and differentiation potential of sorted R1–R4 subsets.

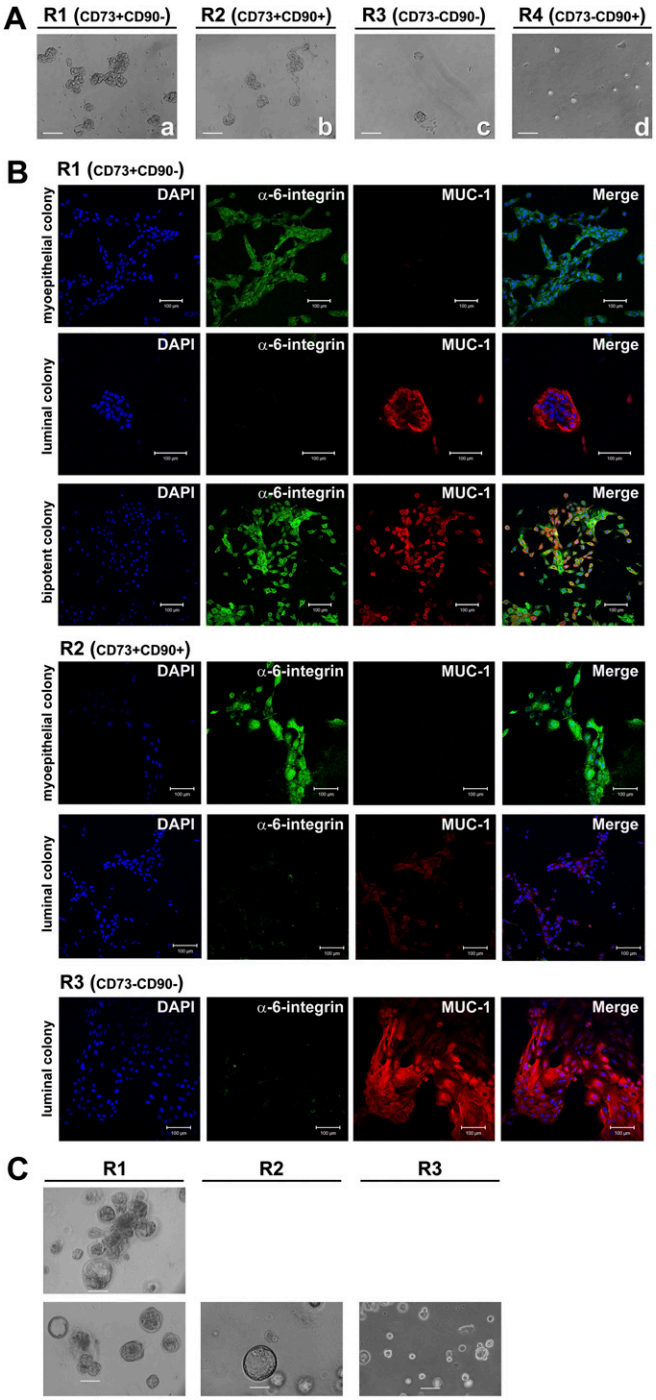



**Fig. S3.** Mammosphere-forming capacity and ability to recapitulate human mammary ductal-alveolar structures in vitro for R1–R4 human mammary epithelial subsets. (*A*) Representative images of mammosphere-forming ability in suspension for R1–R4 subsets (*a*–*d*). Growth was monitored daily to ensure that each structure was generated from a single cell and that individual structures did not merge. Mammosphere frequency was: 3.86 ± 0.13, 3.44 ± 0.16, 3.54 ± 0.14, and 3.46 ± 0.19 for R1 cells at passages 1–4; 0.47 ± 0.01, 0.56 ± 0.05, and 0.54 ± 0.05 for R2 cells at passages 1–3; 0.58 ± 0.01 and 0.48 ± 0.04 for R3 cells at passages 1–2; and 0.44 ± 0.05, 0.38 ± 0.04, and 0.4 ± 0.05 for unsorted cells at passages 1–4. No mammosphere formation was observed for R4 cells. Data are expressed as mean ± SEM ($n = 5$). (*B*) Representative images of first passage R1–R3 mammospheres dissociated and grown in differentiation conditions for 14 d and immunostained for α–6-integrin/CD49f and MUC-1 (mucin 1). R1 mammosphere-derived cells generate monolineage $CD49f^{+}/MUC\text{-}1^{-}$ myoepithelial colonies, monolineage $CD49f^{-}/MUC\text{-}1^{+}$ luminal epithelial colonies, and bipotent $CD49f^{+}/MUC\text{-}1^{+}$ colonies. R2 mammosphere-derived cells generate monolineage $CD49f^{+}/MUC\text{-}1^{-}$ myoepithelial colonies and monolineage $CD49f^{-}/MUC\text{-}1^{+}$ luminal epithelial colonies. R3 mammosphere-derived cells generate only monolineage $CD49f^{-}/MUC\text{-}1^{+}$ luminal epithelial colonies. (*C*) Only R1 cells formed ductal–acinar and acinar structures in colonogenic 3D Matrigel culture. R2 mammosphere-derived cells formed acinar structures and failed to form ductal–acinar structures. R3 mammosphere-derived cells fail to generate either structure in these conditions. (Scale bars, 100 μm.) Data are representative of three reduction mammoplasties.

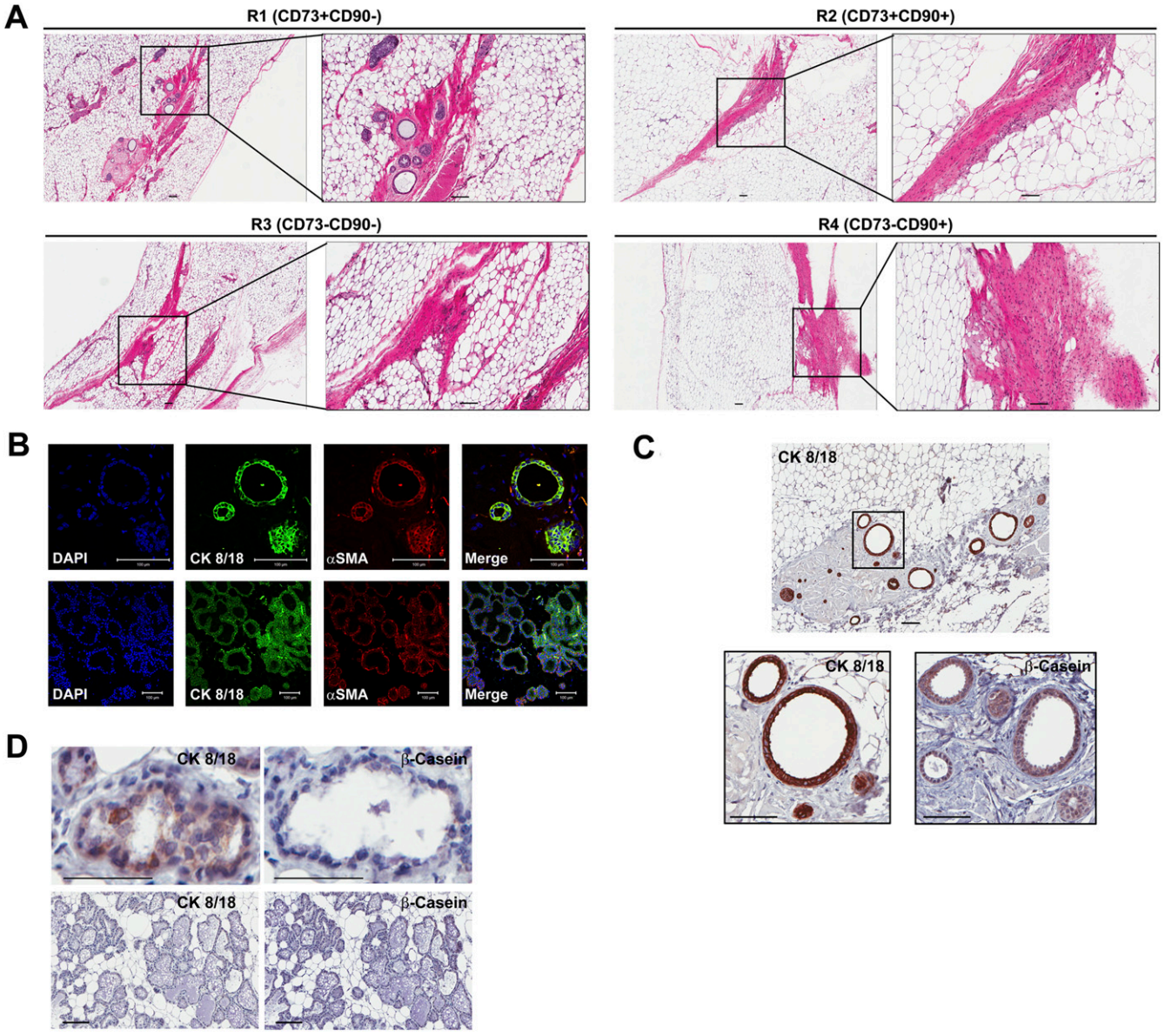



**Fig. S4.** In vivo outgrowth potential of R1–R4 human mammary epithelial subsets transplanted into cleared and humanized NOD/SCID mouse mammary fat pads. (*A*) Hematoxylin and eosin staining of ducts generated by R1–R4 subsets. Number of cells injected for each subset were as follows: R1: 11,000, 15,000, and 15,000; R2: 20,000, 38,000, and 63,000; R3: 200,000, 300,000, and 500,000; and R4: 20,000, 50,000, and 50,000. R2–R4 subsets did not generate any outgrowth when transplanted into mouse mammary fat pads. (*B*) Representative images of acini and ducts consisting of an inner luminal cell layer expressing CK8/18 (green) and an outer myoepithelial layer expressing α-SMA (red) stained with human-specific antibodies documenting the human origin of the structures formed by R1 cells in the mouse fat pad. (*C*) Human-specific anti-CK8/18 antibody documenting the human origin of the R1-derived secretory epithelial cells and alveolar lumen with specific staining for human β-casein milk protein in the mouse fat pad. (*D*) R1 (human)-derived ducts in nonpregnant mouse fat pad stained with human-specific anti-CK8/18 and human β-casein milk protein antibodies documenting the human origin of the R1-derived epithelial cells but lacking β-casein milk protein expression (*Upper*). Endogenous murine mammary glands of a pregnant mouse exposed to human-specific CK8/18 and human-specific β-casein milk protein antibodies demonstrating lack of staining in murine mammary glands and further documenting the human specificity of the antibodies assayed (*Lower*). Data are representative of experiments performed on R1–R4 subsets isolated from three reduction mammoplasties. (Scale bars, 100 μm.)

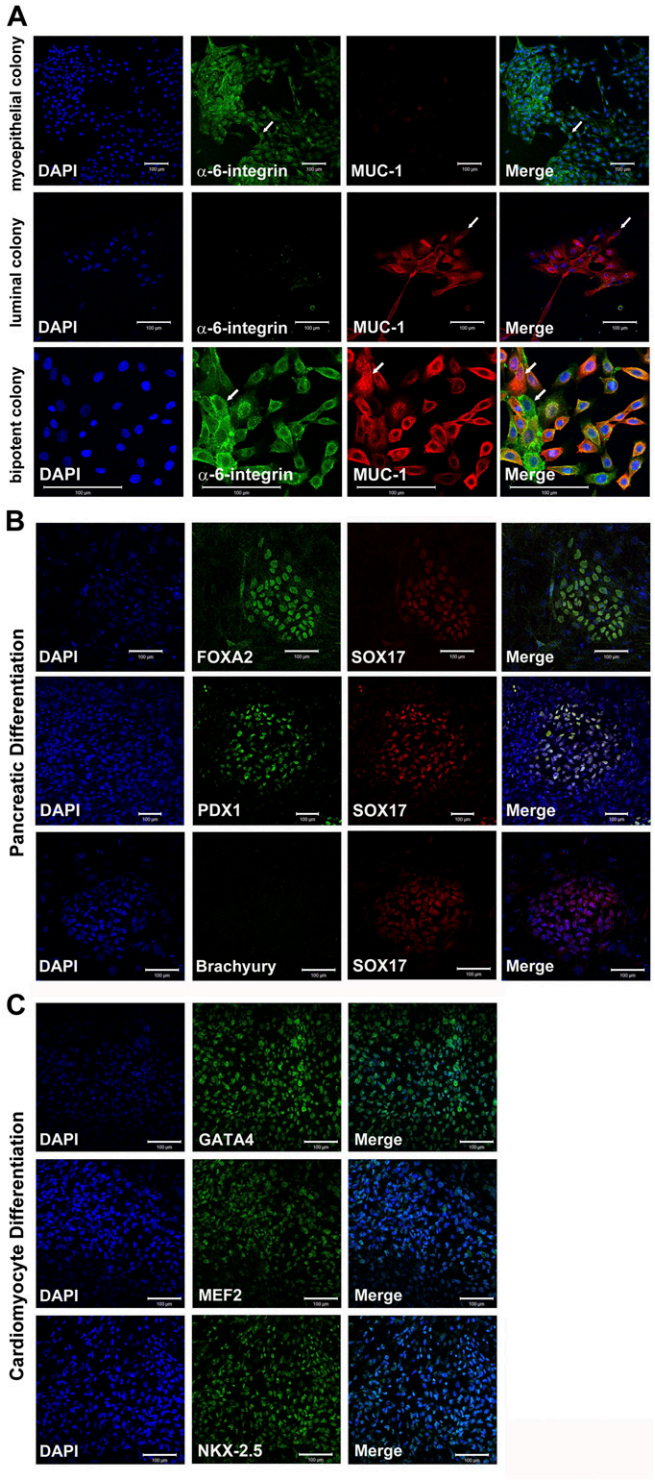



**Fig. S5.** In vitro differentiation potential of single-cell–derived R1 subclones toward ectodermal, endodermal, and mesodermal lineages. Single R1 cells from three independent patient samples were expanded as subclones on feeder layers. Each subclone was divided into three parts and cells were assessed for differentiation potential toward ectodermal, endodermal, and mesodermal lineages. (*A*) Cells exposed to conditions for mammary differentiation expressed α–6-integrin (myoepithelial) and/or MUC-1 (luminal) markers. Arrows indicate cells expressing only one marker. In bipotent progenitors, the majority of cells express both markers. (*B*) Cells exposed to conditions for pancreatic differentiation expressed FOXA2, SOX17, and PDX1. The lack of expression of Brachyury in $SOX17^{+}$ cells documents further commitment toward the endodermal lineage. (*C*) Cells exposed to conditions for cardiomyocyte differentiation generated beating cardiomyocytes which, upon subsequent analysis, expressed cardiomyocyte-specific markers GATA4, MEF2, and NKX2.5. (Scale bars, 100 μm.) Results were obtained in three of three analyses.

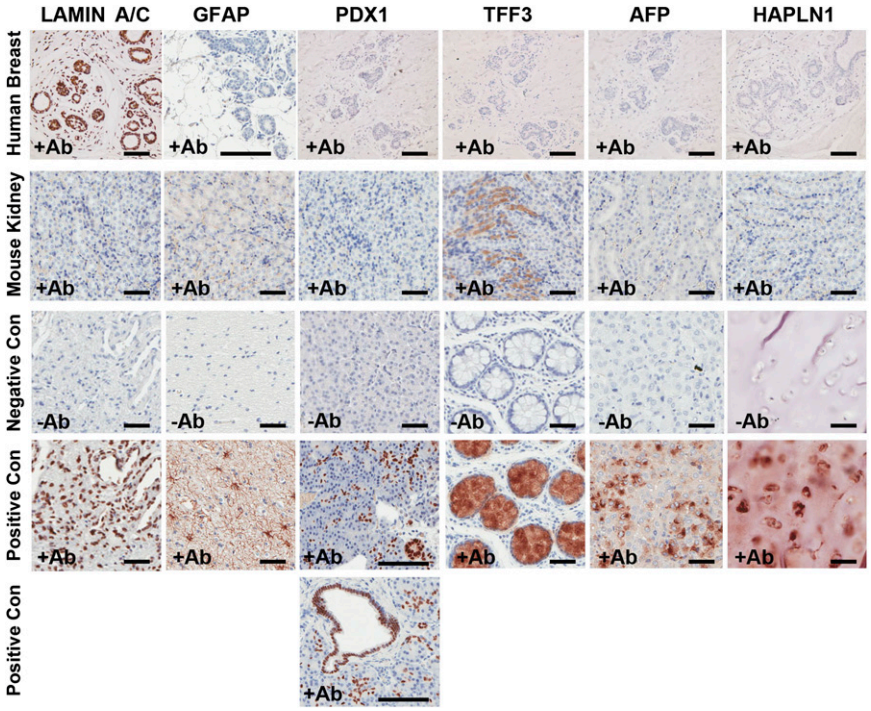



**Fig. S6.** Specificity of antibodies used to stain human-specific lineage derivatives within teratomas. Antibodies selected for teratoma analysis were tested on disease-free human breast tissues to demonstrate that epithelial cells in the breast tissue are negative for neuroepithelium (GFAP), endoderm (PDX1, TFF3, and AFP) and mesoderm (HAPLN1) lineage markers. Antibodies were also tested on mouse kidney sections to demonstrate that the antibodies lacked cross-reactivity with murine tissues. Only the anti-TFF3 antibody showed reactivity with mouse kidney but the human origin of TFF3-expressing structures in R1-derived teratomas was documented with the highly specific pan-human anti-lamin A/C antibody (Fig. 4). Human tissues were stained in the absence (–Ab) or presence (+Ab) of the primary antibodies as described in SI Methods. Finally, human tissues, previously documented to express the markers of interest, were used as antibody-specific positive controls (testis for anti-lamin A/C, brain for anti-GFAP, pancreas for anti-PDX1, colon for anti-TFF3, liver tumor for anti-AFP, and cartilage for anti-HAPLN1). Note the different positive staining pattern of PDX1 in normal pancreas: nuclei of single scattered cells, cells organized in small acini, as well as in ductal–acinar structures. (Scale bars, 100 μm.)

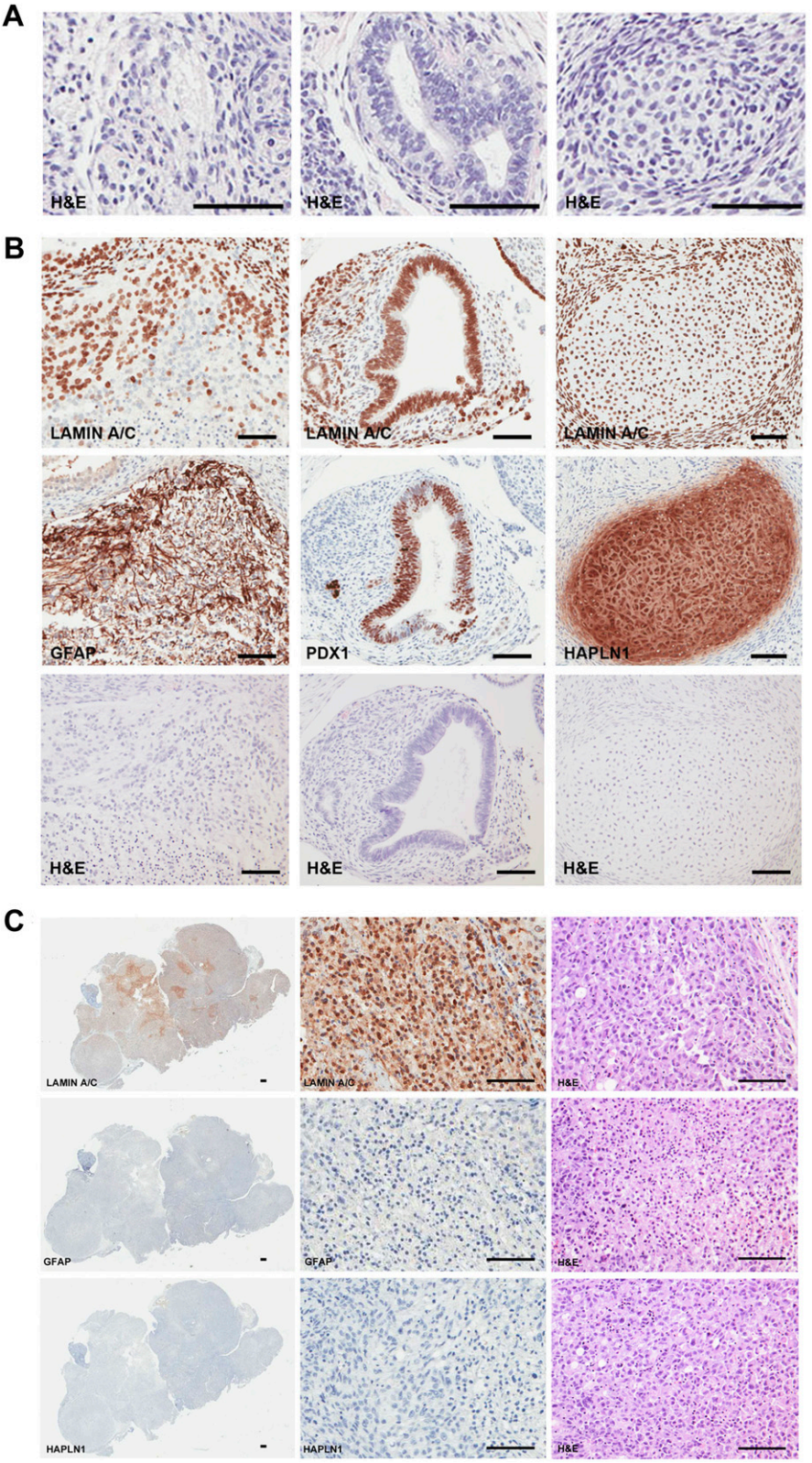
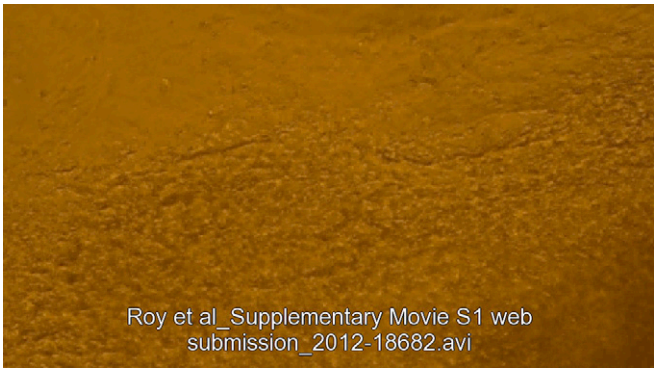



**Fig. S7.** Controls for R1- and hESC H7-derived teratomas and MDA-MB231 mammary tumors generated in immunocompromised mice. (*A*) Hematoxylin and eosin (H&E) staining of serial sections of the paraffin-embedded single-cell–derived R1-generated teratoma stained for the neuroepithelial (ectodermal) marker GFAP, the pancreatic (endodermal) marker PDX1, and the cartilage (mesodermal) marker HAPLN1, respectively, as shown in Fig. 4. (*B*) hESC H7 or (*C*) MDA-MB231 breast cancer cells were grafted under the renal capsules of 7- to 8-wk-old female SCID/beige mice. hESC H7-derived teratomas and MDA-MB231 mammary tumors were harvested 8 wk after injection. (*B*) Views of representative fields of an hESC H7-derived teratoma. Serial sections from the paraffin-embedded hESC H7-derived teratoma were stained for the pan-human–specific marker lamin A/C to document the human origin of these structures (*Top*), for the lineage-specific markers GFAP, PDX1, and HAPLN1 (*Middle*), and for H&E (*Bottom*). (*C*) Views of representative fields of MDA-MB231 mammary tumors

Legend continued on following page

stained for lamin A/C, GFAP, and HAPLN1. Lower magnification views (*Left* column), higher magnification views (*Center* column), and H&E staining (*Right* column) are shown. The hESCs generated teratomas in 4 of 10 analyses. (Scale bars for lower magnification images, 500 μm and for higher magnification images, 150 μm.)




**Movie S1.** Spontaneous beating of cardiomyocytes after R1 differentiation into the myocardial lineage. The time lapse covers a period of about 10 s.

Movie S1


**Movie S2.** Single R1 cells from three independent patient samples were expanded as subclones on feeder layers. Each subclone was divided into three parts and assessed for differentiation potential toward ectodermal, endodermal, and mesodermal lineages. This movie shows spontaneous beating of cardiomyocytes after the part of the subclone was exposed to conditions for cardiomyocyte differentiation. The time lapse covers a period of about 10 s.

Movie S2

### Table S1. Short tandem repeat analysis

Table S1

Genomic DNA was extracted from parental mammary cells dissociated from reduction mammoplasty organoids as well as beating cardiomyogenic cultures derived from the corresponding parental $CD73^{+}CD90^{-}$ (R1) cell population. Analysis was conducted for two individual tissue samples (samples 1 and 2) with sample 2 being analyzed from two different independent differentiation experiments. DNA samples from hESCs H7 at passage 56 and K562 cells were included as internal controls for the genotyping reaction. Loci, including D5S818, D13S317, D7S820, D16S539, vWA, TH01, TPOX, CSF1P0, and the sex-chromosome marker amelogenin, were analyzed using the PowerPlex 1.2 or CellID genotyping kits (Promega). Allelic assignment of all nine loci analyzed was identical between each parental breast cell population and its corresponding mesodermally differentiated (cardiomyogenic) R1 derivative in all cases but differed between donors and from that of H7 and K562 at several loci.

### Dataset S1. Distribution profiles for 10 reduction mammoplasty samples obtained from disease-free women

Dataset S1

Percentages of lineage negative cells expressing CD73 and CD90 in the indicated combinations from 10 reduction mammoplasty samples. Age and ethnicity of tissue donors are provided.

**Dataset S2. Relative expression of selected pluripotency, stress response, and reprogramming genes in pertinent cell populations**

Dataset S2

A custom qPCR-array (Qiagen) was used to analyze transcript levels in hESC H7, hESC H9, human iPSCs from three donors, human MSCs from four donors, R1–R4 cells from four donors, single-cell–derived R1 colonies grown on feeder layers from three donors, and single-cell–derived R1 clones grown under expansion conditions from three donors. Each sample was probed as technical duplicates in two different experiments. Analysis was performed with the RT2 Profiler PCR Array online software (Qiagen). Gene expression levels were normalized to *GAPDH* expression and were expressed as averages relative to those in hESC H9. *P* values are provided for statistical significance.