Supporting Information

Chaffer et al. 10.1073/pnas.1102454108

SI Materials and Methods

Cell Culture. HME cells and all derivatives were cultured in MEGM media as previously described (1). HMECs were isolated from primary tissue as previously described (2) and cultured in M87A+X (3). For cell proliferation assays, cell numbers were determined from triplicate measures using the CellTiter 96 AQueous One Solution Cell Proliferation Assay (Promega) according to the manufacturer's instructions.

Vectors and Viral Infections. pBabe SV40-ER (Zeocin), pBabe H-Ras (Puromycin), PRRL-GFP, pLV-Tomato vectors, production of virus, and infection of target cells have been previously described (1, 4). Infected cells were selected with Zeocin (100 μg/mL) and Puromycin (2 μg/mL).

Immunostaining. Cells were grown on Labtek II Chamber slides (Nunc), fixed using 4% paraformaldehyde, and permeabilized with 0.2% Triton-X 100/PBS. For tissue sections, tumor samples were formalin fixed, paraffin embedded, and cut into 5-μm sections. Antigen retrieval was performed with citrate buffer (pH 6.0) followed by boiling. Alexafluor-594 and -488 secondary antibodies (Invitrogen) were used for detection.

Western Blot Analysis. Protein was extracted from cell lysates using RIPA buffer. Western blots were performed using standard protocols. Blots were developed using ECL (Dura or Femto; Pierce). Horseradish peroxidase-conjugated secondary antibodies were used (Jackson Immunoresearch).

3D Culture on Matrigel. 3D culture of mammary epithelial cells on Matrigel was done as previously described (5).

Mammosphere Culture. Mammosphere culture was performed as previously described (6). For frozen sectioning, mammospheres were collected, incubated in 15% sucrose solution for 15 min, followed by 30% sucrose solution for 15 min, fixed in 4% paraformaldehyde for 30 min, and frozen in O.C.T. Compound (Tissue-Tek). Five-micrometer sections were cut. For flow cytometry analysis, mammospheres were dissociated into single cells by trypsinization.

Flow Cytometry. Cells were prepared according to standard protocols. Samples were sorted on a BD FACSAria SORP and analyzed on a BD LSRII using BD FACSDiva Software (BD Biosciences).

Animal Studies. All mouse studies were performed under the supervision of MIT's Division of Comparative Medicine in accordance with protocols approved by the Institutional Animal Care and Use Committee. Athymic female nude mice were purchased from Taconic Laboratories and NOD/SCID mice were bred in house. Mice were 2–4 mo of age at time of injections. Tumor cells were resuspended in 10% Matrigel/MEGM (20 μL) for mammary fat pad injections. Tumors were dissected at the end of the experiment and weighed. GFP-positive lung metastases were counted from individual lobes by fluorescent microscopy.

Statistical Analysis. Data are presented as mean ± SEM. Student's *t* test (two-tailed) was used to compare two groups ($P < 0.05$ was considered significant) unless otherwise indicated. Limiting dilution analysis was performed as previously described (7).

Antibodies. See Table S1.

Microarray and Gene Set Enrichment Analysis. Total RNA was isolated using the RNeasy Micro kit (Qiagen). Quality was determined by the Agilent Bioanalyzer 2100 (Agilent Technologies). Affymetrix human U133 Plus 2.0 arrays were used. Raw data from Affymetrix arrays was background corrected and normalized using robust multiarray average (RMA) from the affy R package (8). The gene set enrichment analysis (GSEA) was performed by the GSEA-P desktop application (9), using as gene sets the differentially expressed transcripts identified by Raouf et al. (10), from the comparisons of the bipotent CFC-enriched cell fraction with the mature myoepithelial cell fraction (table S1, sets 1–4) and luminal-restricted CFC-enriched with mature luminal cell fractions (table S1, sets 5–8).

Determining Switching Rates (k_s). ***Proliferation model.*** The proliferation rates of for $CD44^{hi}$ (H) and $CD44^{lo}$ (L) cells are assumed to be

$$\dot{N}_H = k_H N_H + k_s N_L \quad \text{[S1]}$$

$$\dot{N}_L = (k_L - k_s) N_s, \quad \text{[S2]}$$

where k_H and k_L are the growth rates of $CD44^{hi}$ and $CD44^{lo}$ cells, respectively, and k_s is the transition rate for cells from the $CD44^{lo}$ to the $CD44^{hi}$ state. Transitions from the $CD44^{hi}$ to the $CD44^{lo}$ state are not observed. All cell types (HME, HMLE-flopc, etc.) have distinct rate constants.

To solve the proliferation model, we use the canonical eigenvalue approach and write the equations as a matrix equation:

$$\frac{d}{dt}\begin{bmatrix} N_H \\ N_L \end{bmatrix} = \begin{bmatrix} k_H & k_s \\ 0 & k_L - k_s \end{bmatrix}\begin{bmatrix} N_H \\ N_L \end{bmatrix}. \quad \text{[S3]}$$

The eigenvalues and vectors are

$$\lambda_{H'} = k_H, X_{H'} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{[S4]}$$

$$\lambda_{L'} = k_L - k_s, X_{L'} = \begin{bmatrix} \chi \\ 1 \end{bmatrix}, \quad \text{[S5]}$$

where λ and X are the eigenvalues and vectors, respectively, and

$$\chi \equiv \frac{k_s}{k_L - k_s - k_H}. \quad \text{[S6]}$$

The proliferation dynamics are therefore described by the equations

$$\begin{bmatrix} N_H(t) \\ N_L(t) \end{bmatrix} = A_{H'} X_{H'} \exp(\lambda_H, t) + A_{L'} X_{L'} \exp(\lambda_L, t), \quad \text{[S7]}$$

where the two A parameters are chosen to match the initial populations.

Fraction of newly derived $CD44^{hi}$ cells. In the flow cytometry experiments, the observable is the fraction of $CD44^{hi}$ cells. This fraction can be written explicitly in terms of the proliferation solution Eq. **S7**,

$$X_{\mathrm{H}} = \frac{N_{\mathrm{H}}(t)}{N_{\mathrm{H}}(t) + N_{\mathrm{L}}(t)} \quad \text{[S8]}$$

$$= \frac{\sin\alpha \exp(\lambda_{\mathrm{H}}, t) + \chi \cos\alpha \exp(\lambda_{\mathrm{L}}, t)}{\sin\alpha \exp(\lambda_{\mathrm{H}}, t) + (1+\chi)\cos\alpha \exp(\lambda_{\mathrm{L}}, t)}, \quad \text{[S9]}$$

where α is defined as the angle made by the vector $[A_{\mathrm{L}'}; A_{\mathrm{H}'}]$. In the steady-state limit the fraction of $CD44^{hi}$ cells is

$$X_{\mathrm{H}} = \begin{cases} 1, & k_{\mathrm{H}} \le k_{\mathrm{L}} - k_{\mathrm{s}} \\ (1+\chi^{-1})^{-1}, & k_{\mathrm{H}} < k_{\mathrm{L}} - k_{\mathrm{s}}. \end{cases} \quad \text{[S10]}$$

Determining the model parameters. For cell types (HME-flopc, HMLE-flopc, and HMLER-flopc) where proliferation data were collected for $CD44^{hi}$ and $CD44^{lo}$ cells independently, all four parameters ($\alpha, k_{\mathrm{L}}, k_{\mathrm{H}'}$, and k_{s}) were determined simultaneously by a χ^2 fit to the newly derived $CD44^{hi}$ data, the asymptotic $CD44^{hi}$ fraction data, and the $CD44^{hi}$ and $CD44^{lo}$ proliferation data.

For cell types (HME and HMLER) where only the bulk proliferation rate was measured, all four model parameters ($\alpha, k_{\mathrm{L}}, k_{\mathrm{H}'}$, and k_{s}) were determined simultaneously by a χ^2 fit to the newly derived $CD44^{hi}$ data, the asymptotic $CD44^{hi}$ fraction data, and the bulk proliferation data. The growth rate of the bulk population was assumed to be the largest eigenvalue in Eq. **S4**.

Characterization of Primary Human Mammary Tissue. A schematic outline of mammary epithelial cell enrichment and a multicolor FACS procedure for the CD24, CD44, and ESA antigens were performed. Leukocytes, endothelial cells, and fibroblasts were removed by using the indicated antibody-coupled magnetic beads. Enriched mammary epithelial cells were subsequently analyzed by multicolor FACS. Only live cells (determined by propidium iodide staining) were analyzed by APC-conjugated mouse anti human-CD24, PE-conjugated mouse anti-human CD44, and FITC-conjugated mouse anti-human ESA antibody (Fig. S6).

1. Elenbaas B, et al. (2001) Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* 15:50–65.
2. Stingl J, Eaves CJ, Zandieh I, Emerman JT (2001) Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* 67:93–109.
3. Garbe JC, et al. (2009) Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res* 69:7557–7568.
4. Shaner NC, et al. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 22: 1567–1572.
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* 30:256–268.
6. Dontu G, et al. (2003) In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* 17:1253–1270.
7. Hu Y, Smyth GK (2009) ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* 347:70–78.
8. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307–315.
9. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP (2007) GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 23:3251–3253.
10. Raouf A, et al. (2008) Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* 3:109–118.

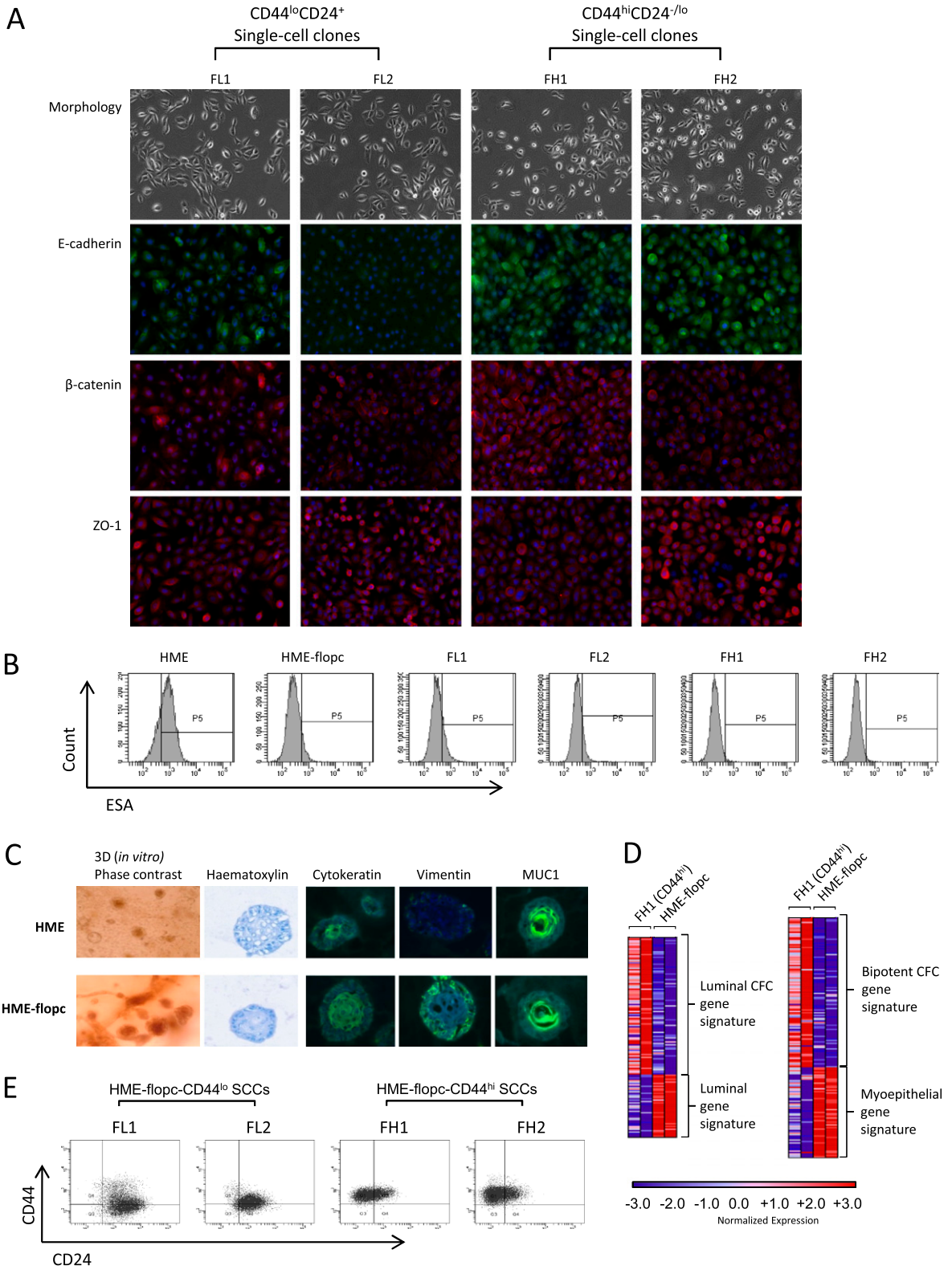


Fig. S1. (*A*) Bright phase micrographs (20×) showing morphology of two HME-flopc-CD44^{lo} single-cell clones (SCCs, FL1 and FL2) and two HME-flopc-CD44^{hi} SCCs (FH1 and FH2). Immunofluorescence (20×) demonstrates that E-cadherin, β-catenin, and ZO-1 rarely localize to cell–cell junctions in HME-flopc SCCs. (*B*) Flow cytometry analysis to determine ESA expression in the various cell populations, demonstrating that HME bulk cells contain ESA-positive and -negative cells, whereas bulk HME-flopc cells and SCCs are largely ESA negative. (*C*) 3D culture of mammary epithelial cells (HME and HME-flopc) on Matrigel illustrating their epithelial origins. Phase contrast micrographs (20×), hematoxylin staining (40×), and immunofluorescence (40×) are shown for cytokeratin, vimentin, and MUC1. (*D*) Gene set enrichment analysis comparing gene expression signatures of pure HME-flopc-CD44^{hi} cells (SCC FH1) and bulk HME-flopc cells to published gene expression profiles of primary luminal colony-forming cells (CFCs), bipotent CFCs, differentiated luminal cells, and differentiated myoepithelial cells (19). HME-flopc-CD44^{hi} cells are enriched for bipotent and luminal CFC gene signatures. (*E*) Flow cytometry plots showing CD44/CD24 profile of HME-flopc SCCs.

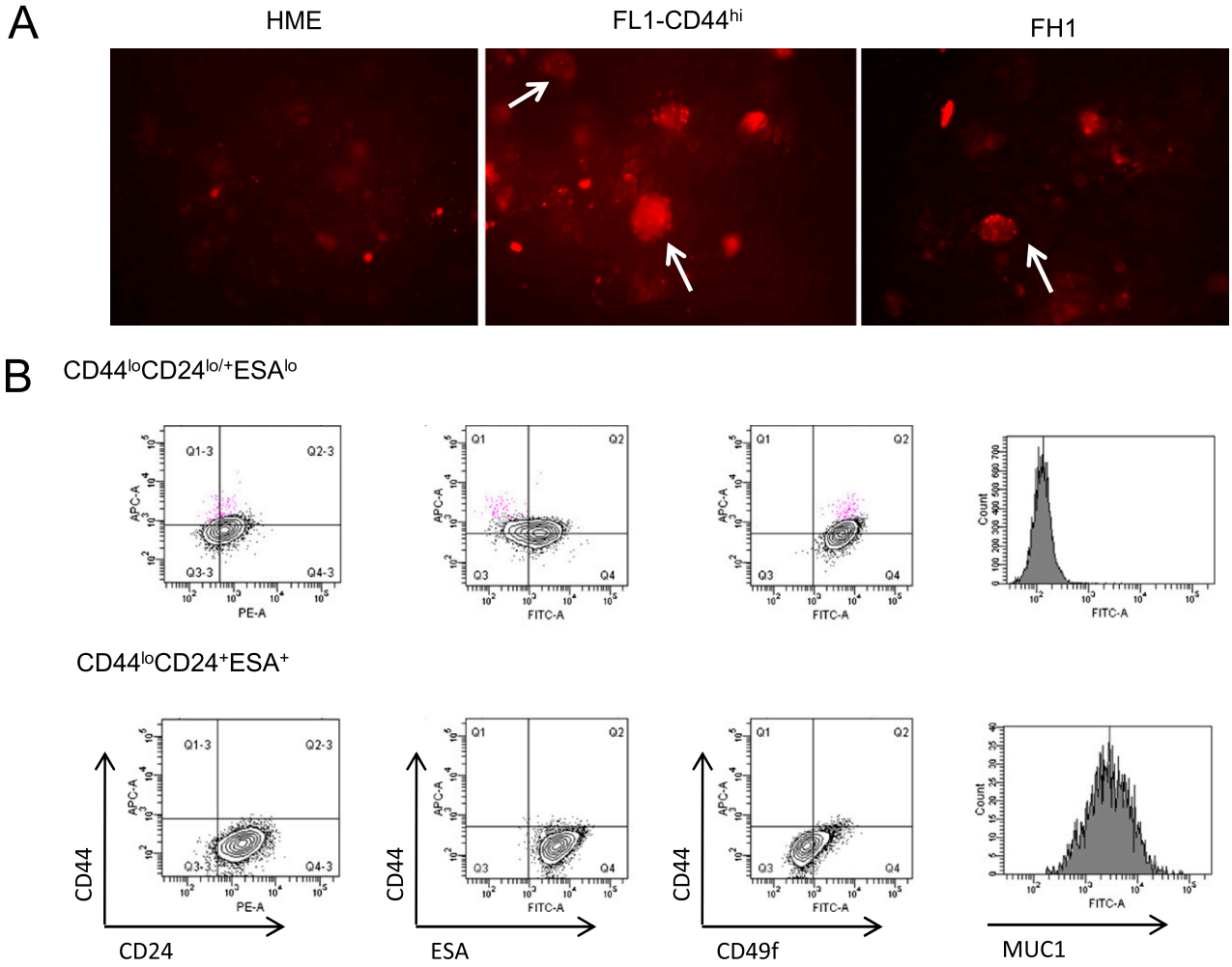


Fig. S2. (*A*) Single-cell suspensions of bulk HME cells, FL1-CD44^{hi} cells (CD44^{hi} cells that arose de novo from a CD44^{lo} cell), or FH1 cells (preexisting CD44^{hi} cells) expressing the red fluorescent protein (pLV-Tomato) were injected into humanized mouse mammary fat pads. Representative fluorescent images are shown of whole-mount mammary fat pads demonstrating the human origin of the epithelial structures that developed (100×). Large oblong-shaped structures (reminiscent of ductal structures) developed from FL1-CD44^{hi} and FH1 cells (white arrows). Small clusters of cells were evident in fat pads containing HME cells. (*B*) Flow cytometry profile of subpopulations of primary human mammary epithelial cells. CD44^{lo}CD24$^{lo/+}$ESA^{-} cells are CD49f^{hi}MUC1^{-}, whereas CD44^{lo}CD24^{+}ESA^{+} cells are CD49f^{lo}MUC1^{+}.

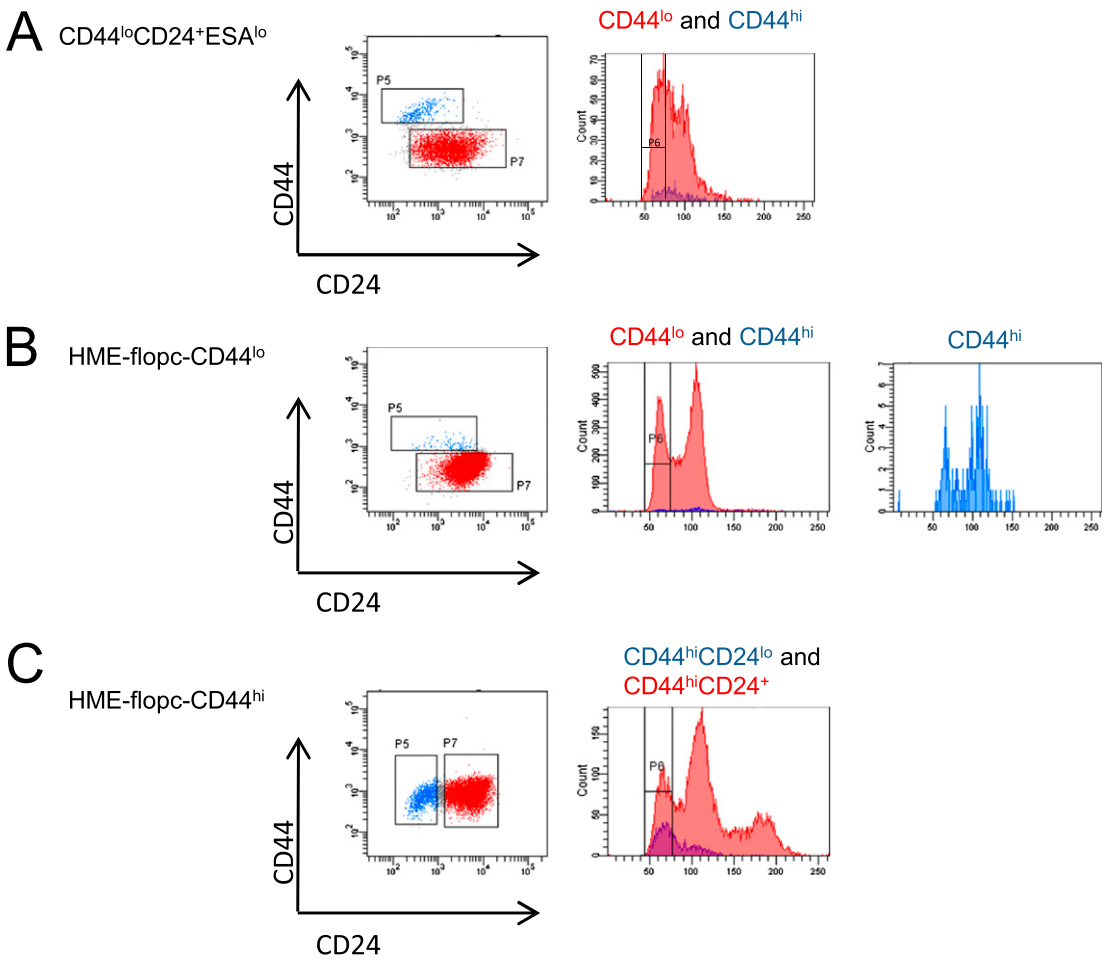


Fig. S3. (*A*) Hoechst staining to determine ploidy of different subpopulations of primary human mammary epithelial cells (HMECs). Subpopulations were purified by flow cytometry from the bulk HMEC population. After 12 d, when a population of de novo-derived $CD44^{hi}CD24^{lo}ESA^{-}$ cells had arisen from the $CD44^{lo}CD24^{+}ESA^{-}$ population, the cells were stained with Hoechst (1:1,000, 30 min). $CD44^{lo}CD24^{+}ESA^{-}$ cells and de novo-derived $CD44^{hi}CD24^{lo}ESA^{-}$ cells are diploid. (*B* and *C*) Hoechst staining to determine ploidy of different subpopulations of HME-flopc cells. (*B*) HME-flopc-$CD44^{lo}$ cells were purified by flow cytometry from the HME-flopc bulk population. After 12 d when a population of de novo-derived HME-flopc-$CD44^{hi}$ cells had arisen, the cells were stained with Hoechst (1:1,000, 30 min) and analyzed by flow cytometry for DNA content. Both HME-flopc-$CD44^{lo}$ and spontaneously arising HME-flopc-$CD44^{hi}$ cells are diploid. (*C*) Preexisting HME-flopc-$CD44^{hi}$ cells were purified by flow cytometry from the HME-flopc bulk population and analyzed by flow cytometry for DNA content. Preexisting HME-flopc-$CD44^{hi}$ cells contain a diploid ($CD44^{hi}CD24^{lo}$) and a tetraploid ($CD44^{hi}CD24^{+}$) population. Karyotype analysis demonstrated that the HME-flopc single-cell clones were tetraploid (Cell Line Genetics). We note that the biology observed between the normal HMECs, diploid HME-flopc-$CD44^{lo}$ cells, and the tetraploid single-cell clones did not differ, suggesting that the behavior of these various cell populations was not influenced by their ploidy.

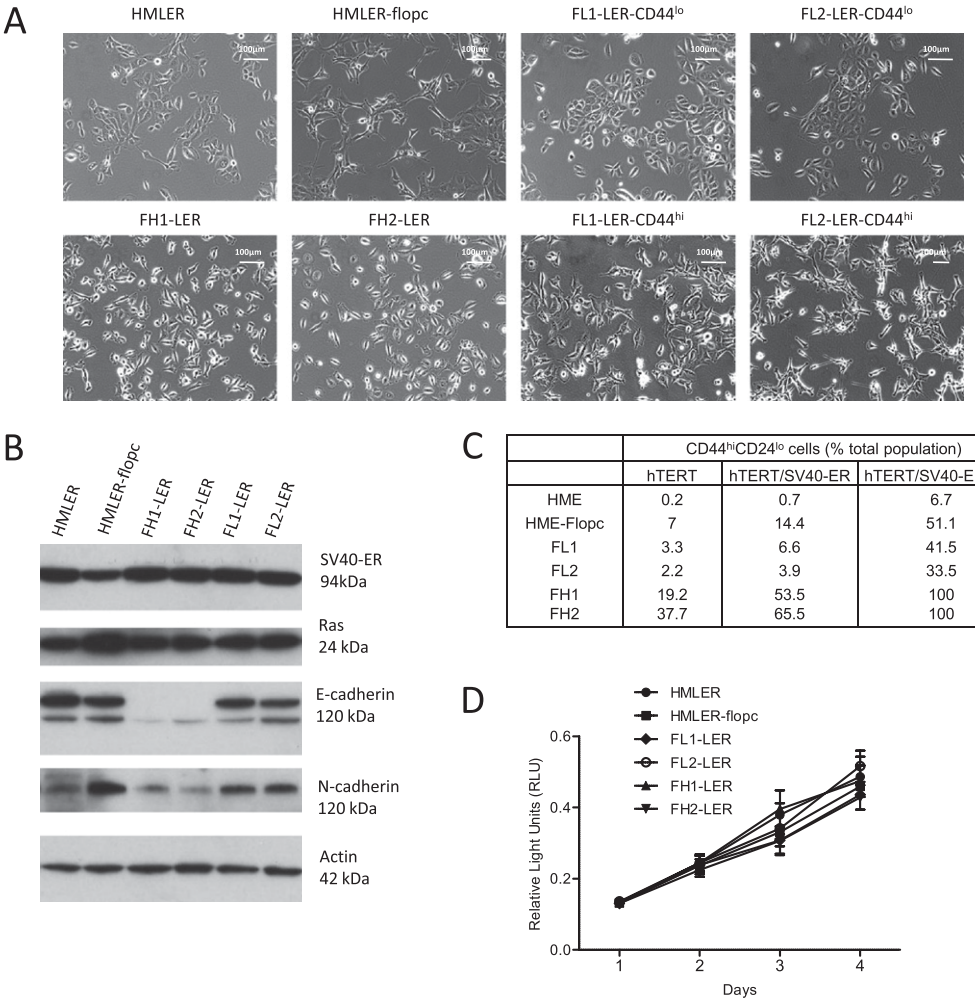
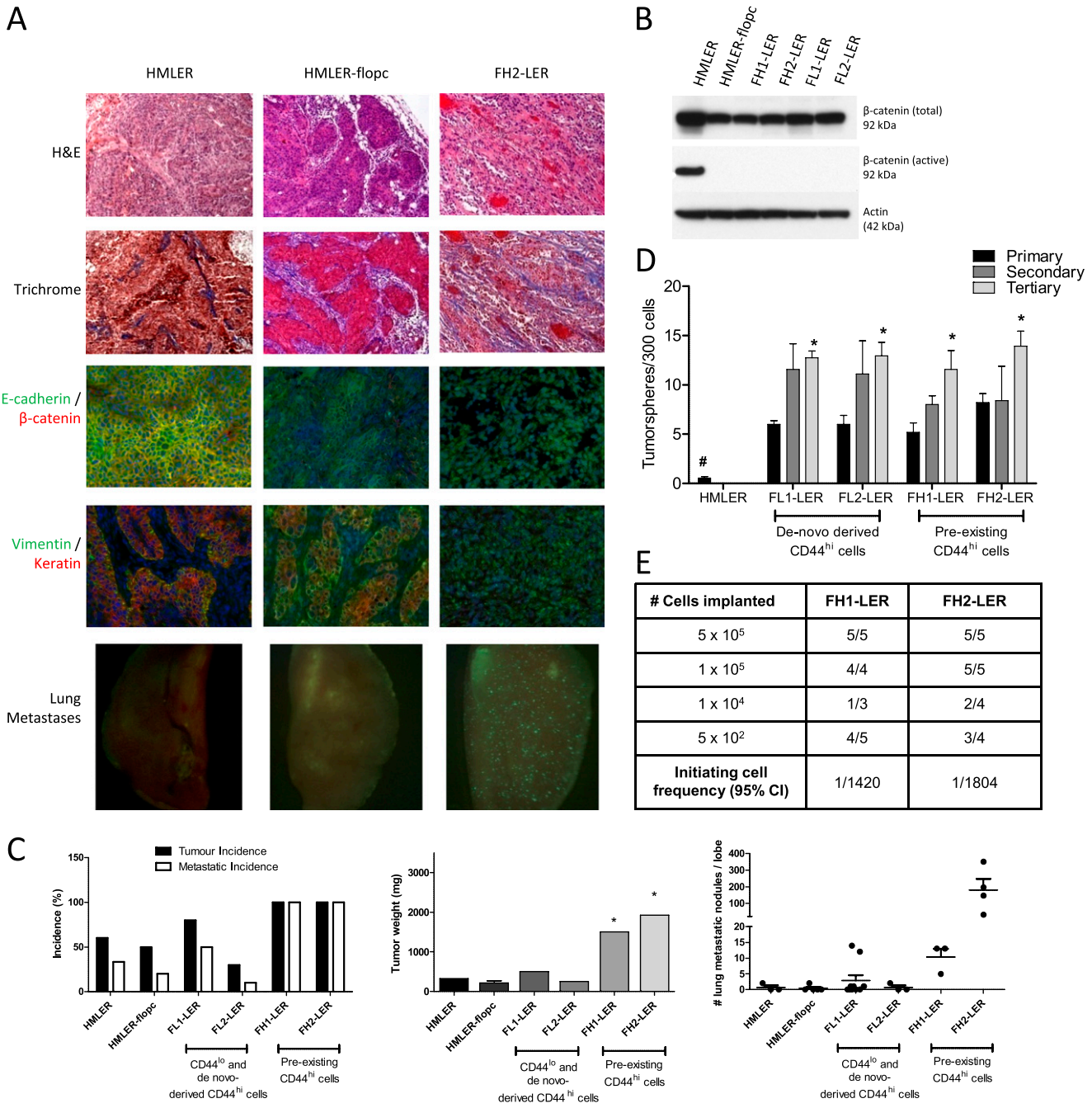


Fig. S4. (*A*) Phase contrast micrographs showing morphology of cultured transformed HME cells and subpopulations. (*B*) Western blot analysis for SV40 early region (SV40-ER) and Ras oncoproteins and epithelial (E-cadherin) and mesenchymal (*N*-cadherin) markers. (*C*) Flow cytometry analysis of $CD44^{hi}CD24^{lo}$ fraction of cells in the various transformed populations of HME cells. (*D*) Growth rates of various cell populations in vitro (determined by MTS assay). Results are mean ± SEM.



# Cells implanted	FH1-LER	FH2-LER
5 x 10^5	5/5	5/5
1 x 10^5	4/4	5/5
1 x 10^4	1/3	2/4
5 x 10^2	4/5	3/4
Initiating cell frequency (95% CI)	1/1420	1/1804

Fig. S5. (*A*) Comparison of tumor histology (hematoxylin and eosin staining and trichrome staining) for HMLER, HMLER-flopc, and FH2-LER ($CD44^{hi}$ single-cell clone); immunofluorescence for E-cadherin, β-catenin, vimentin, pan-cytokeratin (each 20×), and representative pictures of lung metastases formed by the different cell populations. We noted previously (Fig. 1*B*) that HME cells express abundant membranous E-cadherin and β-catenin, whereas HME-flopc cells do not. Similarly, we found that HMLER tumors showed abundant membrane-associated E-cadherin and β-catenin, whereas HMLER-flopc tumors showed dramatically reduced E-cadherin and β-catenin staining. We noted, however, that HMLER tumors expressed vimentin only where tumor cells were in close contact with stroma, whereas tumors derived from HMLER-flopc cells had an even distribution of vimentin-positive cells throughout. (*B*) Western blot analysis showing active β-catenin in HMLER cells only and decreased E-cadherin protein expression in FH1-LER and FH2-LER $CD44^{hi}$ single-cell clones. (*C*) Tumor incidence, weight, and metastatic ability (to lung) of various transformed HME subpopulations following orthotopic injection into NOD/SCID mice (5×10^5 cells/animal, $n = 6–9$ animals/group, $P < 0.05$). Transformed pure HME-flopc-$CD44^{hi}$ single-cell clones (SCCs) (FH1-LER and FH2-LER) exhibit the highest tumorigenicity and metastatic ability compared with HMLER cells and transformed HME-flopc populations (HMLER-flopc, FL1-LER, and FL2-LER). Results are mean ± SEM. (*D*) Serial tumorsphere passage of HMLER cells and purified $CD44^{hi}$ cells from FL1-LER, FL2-LER, FH1-LER, and FH2-LER. Results are mean ± SEM. $^{\#}P < 0.05$, different from all other primary tumorspheres (one-way ANOVA, Newman–Keuls multiple comparison test); $^{*}P < 0.05$, different from primary tumorspheres of the same cell line (Student's *t* test). (*E*) Limiting dilution analysis of two HME-flopc-$CD44^{hi}$ SCCs (FH1-LER and FH2-LER) following s.c. injection in nu/nu mice.

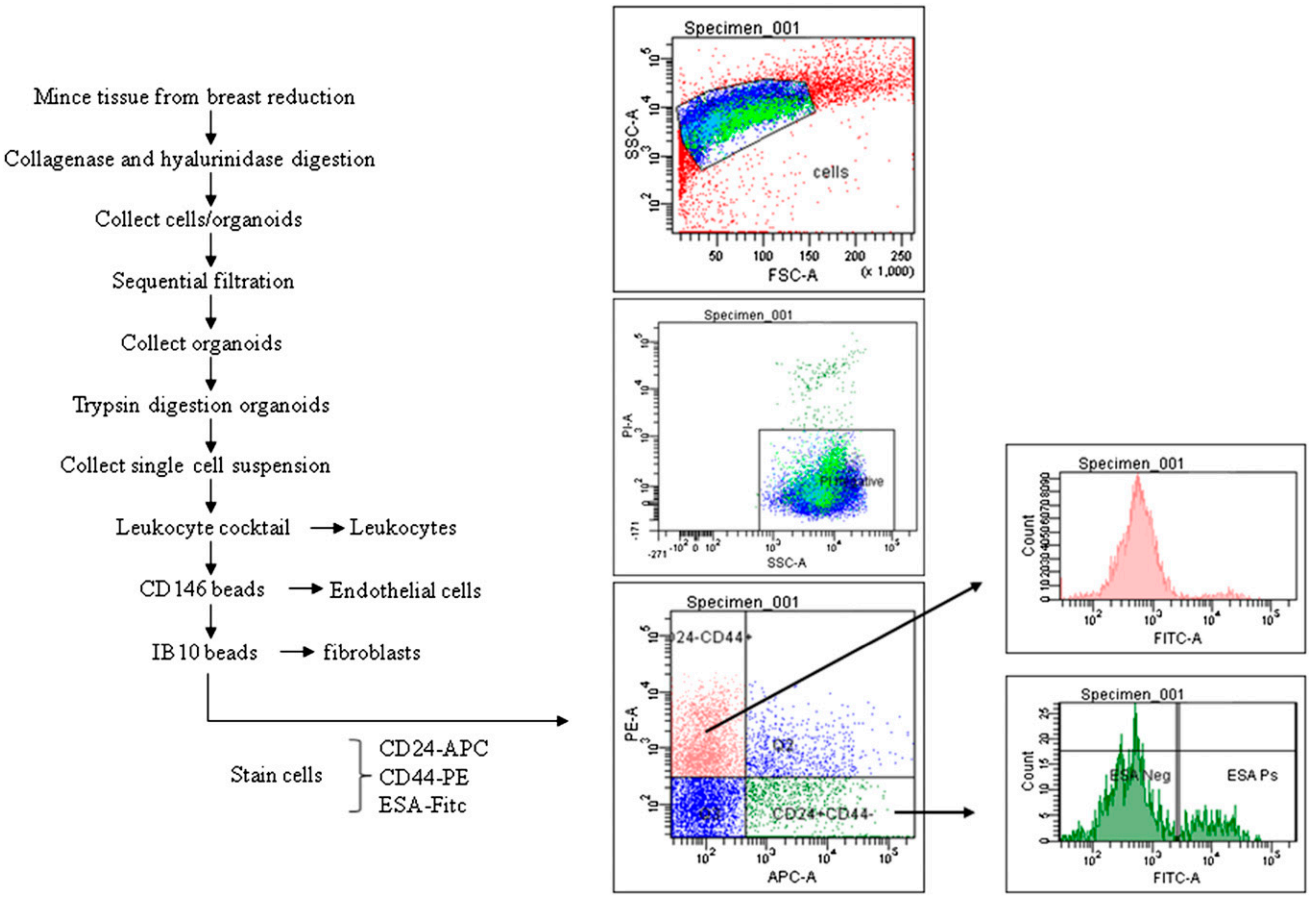


Fig. S6. Schematic outline of primary human mammary epithelial cell enrichment and multicolor FACS procedure.

Table S1. List of antibodies

Antibody	Company	Procedure
Vimentin (V9)	Neomarkers	Immunostaining and Western blot
Cytokeratin-14	Thermo Scientific	
E-cadherin	Cell Signaling	
β-Catenin (total)	BD Biosciences	
ZO-1	Invitrogen	
E-cadherin	BD Biosciences	Western blot
N-cadherin	BD Biosciences	
β-Catenin (active)	BD Biosciences	
β-Actin	Upstate	
Cytokeratin-18	Thermo Scientific	
Cytokeratin-19	Vector Laboratories	
Fibronectin	Santa Cruz Biotechnology	
Large T	Santa Cruz Biotechnology	
Ras	Santa Cruz Biotechnology	
CD10	Stem Cell Technologies	Mammosphere immunofluorescence
Cytokeratin-14	Thermo-Scientific	
Vimentin (V9)	Neomarkers	
MUC1	BD Biosciences	
CD24-FITC	BD Biosciences	Flow cytometry
CD24-PE	BD Biosciences	
CD44-APC	BD Biosciences	
ESA-FITC	GeneTex	