Cell Stem Cell
Resource

# Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process

Afshin Raouf,[1] Yun Zhao,[1] Karen To,[1] John Stingl,[1,7] Allen Delaney,[2] Mary Barbara,[4] Norman Iscove,[4] Steven Jones,[2] Steven McKinney,[3] Joanne Emerman,[5] Samuel Aparicio,[3] Marco Marra,[2,6] and Connie Eaves[1,6,*]
[1]Terry Fox Laboratory
[2]Genome Sciences Centre
[3]Molecular Oncology and Breast Cancer Program
British Columbia Cancer Agency, Vancouver, BC V5Z 1L3, Canada
[4]Division of Stem Cell and Developmental Biology, Ontario Cancer Institute, Toronto, ON M5G 2M9, Canada
[5]Departments of Pathology and Laboratory Medicine, Anatomy and Cell Biology
[6]Department of Medical Genetics
University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[7]Present address: Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK
*Correspondence: ceaves@bccrc.ca
DOI 10.1016/j.stem.2008.05.018

## SUMMARY

**Mature mammary epithelial cells are generated from undifferentiated precursors through a hierarchical process, but the molecular mechanisms involved, particularly in the human mammary gland, are poorly understood. To address this issue, we isolated highly purified subpopulations of primitive bipotent and committed luminal progenitor cells as well as mature luminal and myoepithelial cells from normal human mammary tissue and compared their transcriptomes obtained using three different methods. Elements unique to each subset of mammary cells were identified, and changes that accompany their differentiation in vivo were shown to be recapitulated in vitro. These include a stage-specific change in NOTCH pathway gene expression during the commitment of bipotent progenitors to the luminal lineage. Functional studies further showed NOTCH3 signaling to be critical for this differentiation event to occur in vitro. Taken together, these findings provide an initial foundation for future delineation of mechanisms that perturb primitive human mammary cell growth and differentiation.**

## INTRODUCTION

The normal female mammary gland grows rapidly at puberty to produce an elaborate bilayered tree-like structure composed of an inner layer of luminal cells surrounded by an outer layer of myoepithelial cells. Later cycles of expansion and involution occur during each estrous cycle and even more dramatically with each pregnancy (Howlin et al., 2006; Russo and Russo, 2004). This dynamic activity suggests the lifelong maintenance within the normal mammary gland of a population of self-renewing undifferentiated mammary stem cells. Support for this concept was first provided by vector integration site analysis of serially transplantable outgrowths generated in precleared mammary fat pads of mice transplanted with retrovirally marked tissue fragments (Kordon and Smith, 1998). More recently, a quantitative in vivo assay for the cell of origin of these clonal outgrowths has been identified and used to enable their partial characterization and phenotypic distinction from most other mammary epithelial cells in mice, including those capable of colony formation in vitro (Asselin-Labat et al., 2006; Shackleton et al., 2006; Stingl et al., 2006a).

Several lines of evidence suggest that the mammary gland of normal adult women also contains a population of mammary stem cells. These include in situ studies of X chromosome inactivation patterns in normal human mammary tissue indicating a common origin of cells in adjacent lobules and ducts (Tsai et al., 1996) and in vitro experiments identifying a unique and rare subset of human mammary progenitor cells that generate mixed colonies containing both luminal and myoepithelial cells, as well as other subsets of clonogenic cells that produce only one or the other mature cell type (Stingl et al., 1998, 2001). The properties exploited most effectively for the prospective isolation of these different types of colony-forming cells (CFCs) include their shared expression of an epithelial marker known as epithelial cell adhesion molecule (EPCAM) (Stingl et al., 1998, 2001) and $\alpha_6$-integrin (CD49f), a frequent marker of epithelial progenitors (Stingl et al., 2001). The undifferentiated (bipotent) CFCs have also been further distinguished by their expression of the common acute lymphoblastic leukemia antigen (CALLA, also called CD10), a marker of the more basally situated mature myoepithelial cells (Stingl et al., 1998), whereas the luminal-restricted CFCs selectively express Mucin-1 (MUC1), a specific marker of mature luminal cells in the mammary gland.

We now describe a procedure that utilizes additional markers to subdivide normal human mammary epithelial cells into four fractions, two of which are more highly and exclusively enriched in bipotent CFCs and luminal-restricted CFCs than previously achievable, the other two fractions representing the mature myoepithelial and luminal cells. Long serial analysis of gene expression (LongSAGE) and single-channel expression microarray (Affymetrix) technologies were used to generate global transcript profiles for each of these four subpopulations, and some of the interesting differences were then further examined by quantitative real-time PCR (Q-RT-PCR) measurements. The results

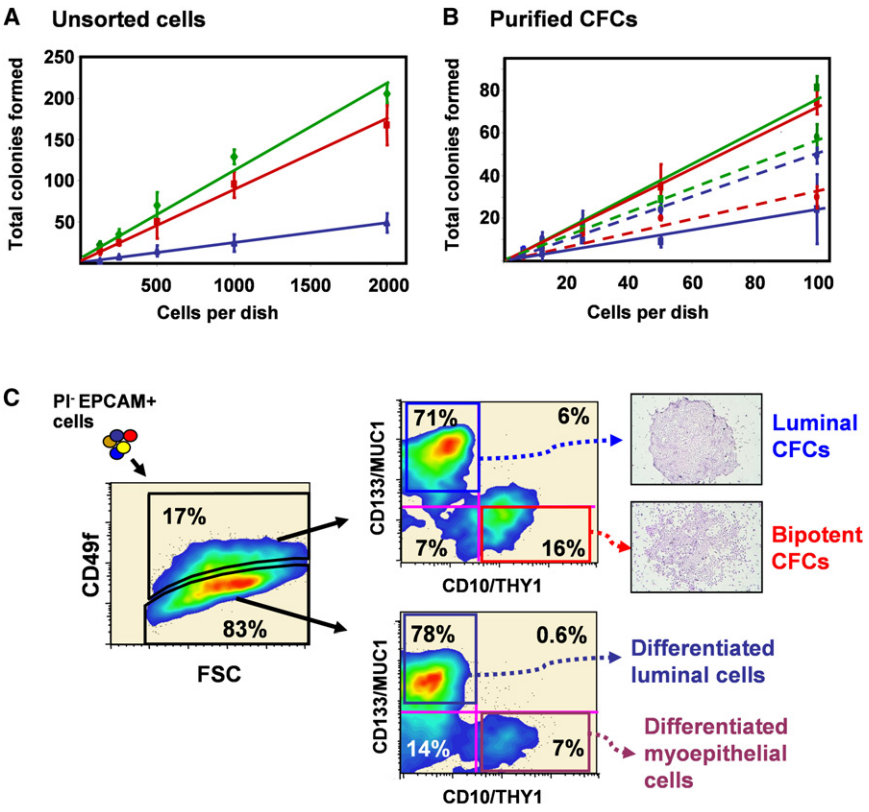


**Figure 1. Detection and Purification of Distinct, Functionally Defined Subsets of Normal Human Mammary Epithelial Cells**

(A) Linearity of colony formation in assays of unseparated EPCAM$^+$ cells from 3-day cultured mammoplasty cells (green, red, and blue lines show data for three different samples, $r \geq 0.95$ in each case). Values shown are the mean ± SEM of three to six replicates in each experiment.

(B) Linearity of colony formation in assays of fractions enriched in bipotent (dashed lines) or luminal-restricted (solid lines) CFCs isolated as shown in (C) from the three samples shown in (A) (and similarly color-coded, $r \geq 0.95$ in each case).

(C) Strategy used to isolate subpopulations of normal human mammary EPCAM$^+$ viable (PI$^-$) epithelial cells selectively enriched in their content of bipotent CFCs, luminal-restricted CFCs, mature myoepithelial cells, or mature luminal cells. A representative FACS plot is shown. Typical mixed and luminal colonies generated in assays of 50 cells from the purified subsets indicated are shown (magnified 32×).

(D) Distribution of bipotent (circles) and luminal-restricted (triangles) CFCs in the subsets indicated (n = 16).

(E) To calculate the yield of bipotent and luminal-restricted CFCs in each fraction, the values for each progenitor type in the EPCAM$^+$CD49f$^+$ fraction were set = 100%.

The error bars represent the standard error of the mean.

identify many gene expression changes that accompany the differentiation of primitive normal human mammary cells both in vivo and in vitro. They also provide new evidence for a previously unrecognized and nonredundant role of NOTCH3 signaling in the commitment of bipotent mammary cells to the luminal lineage.

## RESULTS

### Isolation of Highly Purified Populations of Primitive and Mature Subsets of Human Mammary Epithelial Cells

Human mammary CFCs constitute approximately 1% of freshly dissociated, fibroblast-depleted suspensions of normal reduction mammoplasty samples, and this value can be increased several-fold by culturing the cells in bulk for 3 days (Stingl et al., 2001). Using such "precultured" cells and improved assay conditions, we first showed that these allow CFC progenitor frequencies to be measured independent of the number of cells plated over a wide range (Figures 1A and 1B). Additional experiments showed that highly purified and distinct populations of bipotent and luminal-restricted CFCs were consistently obtained by initial immunomagnetic selection of the EPCAM$^+$ cells present in 3-day cultures of dissociated mammoplasty samples followed by subdividing both the CD49f$^+$ and CD49f$^-$ compartments according to their combined expression of either MUC1 and CD133, or CD10 and THY1 (CD90) (Figure 1C). Using this protocol, we found that the CD49f$^-$ fraction contained very few CFCs (2 ± 1 CFCs per 100 starting CD49f$^-$ cells, n = 4) in contrast to the CD49f$^+$ cells, of which approximately 50% were CFCs and thus contained >95% of the EPCAM$^+$ CFCs. Accordingly, the yield of CFCs present in subsets of CD49f$^+$ cells analyzed according to their expression of MUC1 and CD133, or CD10 and THY1, was used to calculate progenitor yields from different starting populations. The MUC1$^-$CD133$^-$(CD10/THY1)$^+$ subset of the CD49f$^+$ fraction contained the majority (57%) of the bipotent CFCs at a purity of 45% ± 3% with very few luminal-restricted CFCs present (only 3% of this subset). The remaining (MUC1/CD133)$^+$CD10$^-$THY1$^-$CD49f$^+$ cells contained most (96%) of the luminal-restricted CFCs, which were present in this fraction at a purity of 32% ± 3% and were contaminated by the presence

**Table 1. Transcriptome Profiling of Highly Purified Normal Human Mammary Epithelial Progenitor Populations**

| | | Colony Phenotype | | | Transcriptome Profiling | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | LongSAGE Library | | | Affymetrix Analysis |
| Sample | Purified Fraction | Luminal | Mixed | Myoepithelial | Amount of RNA Used | Total Tags | Total Tag Types | Amount of RNA Used |
| 1 | Bipotent CFC-enriched | 0.5 | 51 | 0 | 10 ng | 205,221 | 40,640 | 50 pg |
| | Luminal-restricted CFC-enriched | 28 | 6.5 | 0 | 10 ng | 203,246 | 35,236 | 50 pg |
| | Mature myoepithelial | ND | ND | ND | 10 ng | 210,834 | 38,336 | 50 pg |
| | Mature luminal | ND | ND | ND | 10 ng | 201,037 | 32,632 | 50 pg |
| 2 | Bipotent CFC-enriched | 3.3 | 46 | 1 | | | | 50 pg |
| | Luminal-restricted CFC-enriched | 40.5 | 0.5 | 0 | | | | 50 pg |
| | Mature myoepithelial | ND | ND | ND | | | | 50 pg |
| | Mature luminal | ND | ND | ND | | | | 50 pg |
| 3 | Bipotent CFC-enriched | 5 | 37 | 2 | | | | 50 pg |
| | Luminal-restricted CFC-enriched | 33 | 5 | 0 | | | | 50 pg |
| | Mature myoepithelial | ND | ND | ND | | | | 50 pg |
| | Mature luminal | ND | ND | ND | | | | 50 pg |

Bipotent and luminal-restricted progenitor-enriched fractions were isolated from three different reduction mammoplasty samples. The progenitor frequency in each subfraction was calculated as described in the Experimental Procedures. ND, not determined.

of very few bipotent CFCs (only 5% of this subset) (Figures 1D and 1E).

We next asked whether cells coexpressing cytokeratin 14 and 18 would be found in either of the CD49f$^+$ subsets and, if so, whether their numbers would correlate with the number of CFC present, since coexpression of these two cytokeratins has been associated with a primitive state of mammary epithelial cells (Smalley et al., 1999; Welm et al., 2002). Immunostaining of sorted MUC1$^-$CD133$^-$(CD10/THY1)$^+$ and (Muc1/CD133)$^+$CD10$^-$THY1$^-$ cells from three different reduction mammoplasty samples revealed that both of these fractions consistently contained detectable, albeit variable, frequencies of dually cytokeratin 14$^+$ and 18$^+$ cells. However, relatively high CFC frequencies were seen in some samples where dual cytokeratin 14/18$^+$ cells were rare and vice versa (see Table S1 available online), thus indicating nonidentity of this phenotype with any type of CFC.

## Transcriptional Profiling of Different Purified Subsets of Human Mammary Epithelial Cells

cDNA preparations were generated from each of the four subsets of normal human mammary cells using PCR-amplification methods previously shown to preserve transcript representation in both LongSAGE libraries (Zhao et al., 2007) and array analyses (Iscove et al., 2002). LongSAGE libraries were made from all four fractions obtained from one mammoplasty sample (Table 1), and each library was then sequenced to a depth of ~200,000 tags (Gene Expression Omnibus [GEO] accession number GSE11395). We found that each of these libraries contained approximately 34,000 unique tags when a 99.9% sequence quality cut-off was used. Twelve Affymetrix array hybridizations were also performed, one for each of the four fractions obtained from the same sample used for the LongSAGE libraries and the other eight from the four fractions isolated from two additional mammoplasty samples (Table 1 and GEO accession number GSE11395). An unsupervised hierarchical clustering of the data obtained by Pearson correlation analysis of the LongSAGE library data demonstrated that the population of cells that was most enriched in bipotent CFCs was most closely related to the differentiated myoepithelial cells. Conversely, the population most enriched in luminal-restricted CFCs was most closely related to the differentiated luminal cells (Figure 2A). Similar results were obtained from the larger Affymetrix dataset (Figure 2B).

We then used DiscoverySpace software (Robertson et al., 2007) to identify all tags in the LongSAGE libraries that mapped to a RefSeq transcript at position 3 or lower and were significantly differentially expressed (at a confidence level of 95%) in pairwise comparisons of the four LongSAGE libraries (Tables S2–S5). In terms of cell surface markers, both CD29 and PROCR (endothelial protein C receptor/EPCR) transcripts were found to be upregulated in the bipotent CFC-enriched fraction. Expression of the *CD29* gene is also upregulated in mouse mammary stem cells (Shackleton et al., 2006), and increased expression of *PROCR*/*EPCR* is a feature of hematopoietic and hair-follicle stem cells (Balazs et al., 2006; Blanpain et al., 2004).

From the Affymetrix data, we identified differentially expressed transcripts using a 1.5-fold cut-off and a p value $\leq 0.05$ (Tables S6–S9). To determine whether and how these differences might correlate with functionally distinct cell types, Volcano plots were used to identify the most differentially expressed transcripts from each of the six possible pairwise comparisons, and a K-mean clustering algorithm was then performed using centered Pearson correlation analysis. From this analysis we obtained four gene expression clusters (Figure 2C). One of these clusters was exclusive for the luminal-restricted CFCs and included transcripts for both CD24 and ALDH5A1. CD24 is of interest because it was initially found to be expressed on certain breast cancer stem cells (Al-Hajj et al., 2003) and subsequently was found to be highly expressed on purified human luminal epithelial cells (Jones et al., 2004), on luminal-restricted CFCs in the mouse (Asselin-Labat et al., 2007; Stingl et al., 2006a), and at lower level on mouse mammary stem cells (Asselin-Labat et al., 2007; Shackleton et al., 2006; Stingl et al., 2006a). Consistent with these

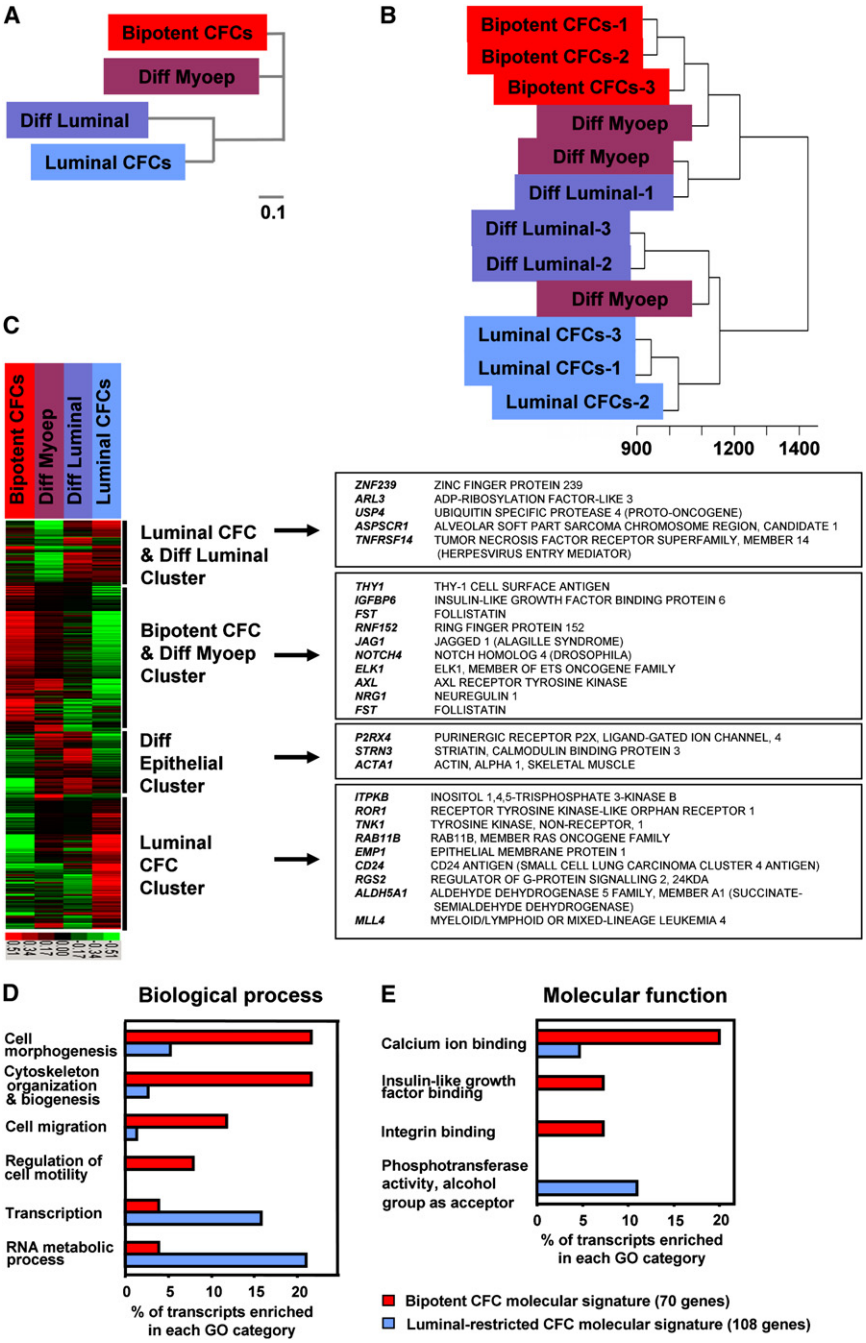


**Figure 2. Relatedness of Undifferentiated, Lineage-Restricted, and Mature Normal Human Mammary Epithelial Subpopulations Revealed by Global Transcriptome Comparisons**

(A) Relatedness of the LongSAGE libraries determined by Pearson correlation analysis.

(B) Relatedness of the transcriptomes identified by Affymetrix analyses of all four subsets of mammary cells (three samples each) derived from an unsupervised hierarchical clustering analysis.

(C) Cluster analysis performed by applying a K-mean clustering algorithm of the most highly differentially expressed transcripts identified in the Affymetrix data comparisons of gene expression in the four subsets of human mammary cells studied here.

(D) Distribution according to the Biological Process (D) and Molecular Function (E) GO categories at level 5 of transcripts found to be overexpressed in the bipotent as compared to the luminal-restricted CFC-enriched subpopulations (70 genes, red bars) and in the luminal-restricted CFC-enriched cells as compared to the bipotent CFC-enriched subpopulations (108 genes, blue bars) from both the LongSAGE and Affymetrix data comparisons based on their assignment.

findings, we determined from FACS analyses that virtually all of the $CD24^+$ $EPCAM^+$ cells coexpressed MUC1 and/or CD133 and accounted for approximately 82%–96% of the $(MUC1/CD133)^+$ subset (Figure S1). High *ALDH5A1* expression in the luminal CFC-enriched fraction is of interest in view of the recent report that primitive normal and malignant human mammary cells exhibit uniquely elevated ALDH activity (Ginestier et al., 2007; Liu et al., 2008). In contrast to the luminal-restricted CFCs, a distinct cluster for the bipotent CFCs was not resolved. Instead, a cluster that included both the bipotent CFCs and the differentiated myoepithelial cells was identified. This cluster contained transcripts for THY1/CD90, which was used to isolate these populations, and also for both JAGGED1 (JAG1) and the NOTCH4 receptor (Figure 2C).

A total of 178 genes were identified as differentially expressed in both the LongSAGE and Affymetrix comparisons of the fractions enriched in bipotent versus luminal-restricted CFCs, 70 of which were more highly expressed in the bipotent CFC fraction and the other 108 of which were more highly expressed in the luminal CFC fraction (Table S10). Further categorization of these differentially expressed genes according to their Gene Ontology Biological Process and Molecular Function groups revealed that the bipotent CFC molecular signature contained a preponderance of transcripts for proteins involved in cell migration, shape control, and morphogenesis, and in calcium, integrin, and insulin signaling (Figures 2D and 2E). In contrast, the luminal-restricted CFC molecular signature was found to contain a preponderance of transcripts for proteins involved in transcription, RNA metabolism, phosphotransferase activity, and retinoic acid synthesis (ALDH1A3) and signaling (CRAPB2 and RBP7). Shared differentially expressed transcripts for other comparisons are listed in Tables S11–S13, and their numbers are shown diagrammatically in Figure S2. Together, these data suggest at least 332 genes that undergo changes in expression during normal human mammary epithelial cell differentiation. Interestingly, CD44, a cell surface adhesion protein

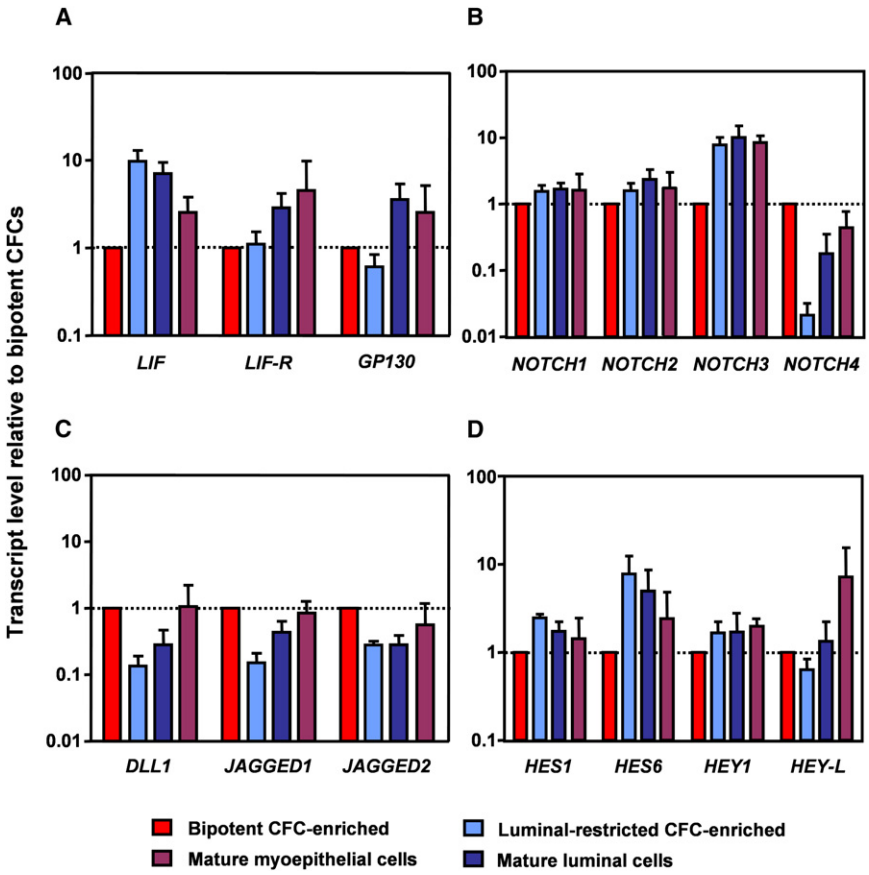


**Figure 3. Comparison of Q-RT-PCR Expression Data for LIF and NOTCH Pathway Genes Expressed in Different Subsets of Human Mammary Cells**

(A–D) The level of expression of each transcript type was assessed by normalizing its level to the level of *GAPDH* transcripts in the same extract and then to the levels of the same test transcript in the bipotent CFC-enriched fraction (set = 1, dotted lines). The error bars represent the standard error of the mean.

whose expression has been reported to allow selective enrichment of breast tumor-initiating cells (Al-Hajj et al., 2003), was not found to be differentially expressed in either CFC-enriched fraction by either transcriptome profiling method. This failure to find CD44 expression to be a discriminating feature of any of the subsets of EPCAM$^+$ cells studied here was further supported by FACS analysis of antibody-stained preparations (Figure S1). These showed 30%–50% of all EPCAM$^+$ cells are CD44$^+$ and that among the CD49f$^+$ cells (which includes both the luminal and bipotent CFCs subtypes, Figure 1), approximately 90% of the cells coexpress CD44.

We then selected 32 genes for more extensive analysis of their expression levels as determined by Q-RT-PCR. These genes included 18 identified as differentially expressed in different mammary subsets either in the LongSAGE libraries or the Affymetrix data sets as well as 14 genes implicated in differentiation and/or breast cancer. Transcript levels for each test gene were normalized to *GAPDH* transcript levels in the same extract, which gave similar results to those obtained when an alternative highly expressed gene (*RPL36*) was used for the normalization (data not shown). The results of the Q-RT-PCR assays demonstrated consistent differential expression by 15 of the 18 genes identified from the global gene expression studies and 19 of all 32 surveyed (Table S14).

Included among these were the estrogen and progesterone receptor genes. These showed high progesterone receptor and low estrogen receptor transcript levels in the bipotent CFC-enriched fraction, which changed to low levels of progesterone receptor transcripts and high levels of estrogen receptor transcripts in the luminal-restricted CFC-enriched fraction (Table S14). Q-RT-PCR also confirmed the LongSAGE data indicating *LIF* expression to be upregulated (9-fold) in the fraction enriched in luminal-restricted CFCs as compared to the fraction enriched in bipotent CFCs and then similarly reduced (~9-fold) in the differentiated luminal cells but remaining unchanged in the mature myoepithelial cell fraction (Figure 3). Q-RT-PCR data further demonstrated transcripts for both components of the LIF receptor (LIFR and GP130) to be present in all four cell fractions examined, albeit at constant levels.

Another regulatory gene that showed a marked but progressive change in expression during luminal cell differentiation (30-fold overall decrease from bipotent progenitors to mature luminal cells) was the Iroquois-class homeobox gene *IRX4* (Table S14). *IRX4* has been associated with the development of the lungs and heart (Bruneau et al., 2001; van Tuyl et al., 2006). Interestingly, transcripts for *IRX2*, another Iroquois-class homeobox gene, were previously found to be preferentially expressed by luminal and not basal epithelial cells in the mouse mammary gland (Lewis et al., 1999).

All profiling methods also showed that transcripts for the NOTCH ligand JAGGED2 were highest in the bipotent CFC-enriched fraction as compared to the luminal-restricted CFC-enriched fraction, with the opposite scenario for the NOTCH signaling receptor, NOTCH3 (Table S10 and Figure 3). Expression of *NOTCH1* and *NOTCH2* was also found to increase during the process of luminal differentiation, although the changes in transcript levels for these genes were more modest. *NOTCH4* expression showed the opposite pattern; transcripts levels were relatively high in the bipotent CFC-enriched fraction and then decreased more than 50-fold during luminal differentiation and approximately 2-fold during myoepithelial cell differentiation. The Q-RT-PCR data also showed higher (10-fold) expression of the NOTCH receptor ligand genes *DLL1*, *JAGGED1*, and *JAGGED 2* in the bipotent CFC-enriched fraction as compared to the luminal-restricted CFC-enriched fraction (Figure 3 and Table S14). Interestingly, transcripts for *HES1*, *HES6*, and *HEY1*, known target genes of NOTCH, were found to be upregulated in parallel with the upregulated expression of the *NOTCH1*, *-2*, and *-3* receptors in the luminal-restricted CFC-enriched fraction (Figure 3).

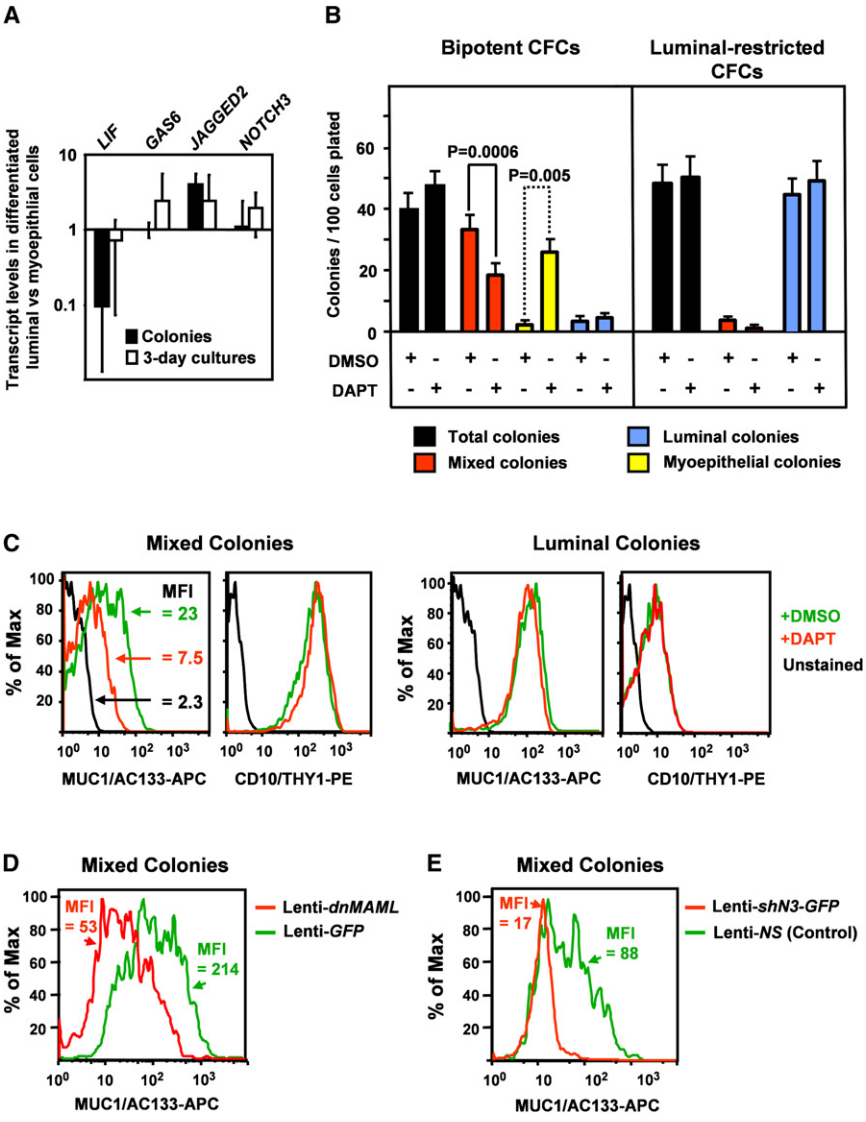


**Figure 4. Changes in Gene Expression Seen during the Differentiation of CFCs In Vitro Are Similar to Those Evident in 3-Day Cultured Cells and Evidence of a Role for Notch in Normal Human Mammary Epithelial Cell Commitment to the Luminal Lineage**

(A) Transcript levels of six genes in 8-day colonies produced by the luminal-restricted CFCs were compared to the levels of the same transcripts in 8-day colonies produced by the bipotent CFCs (in vitro ratio). These ratios are compared to the transcript levels in the mature luminal [EPCAM$^+$CD49f$^-$(MUC1/CD133)$^+$CD10$^-$THY1$^-$] or myoepithelial [EPCAM$^+$CD49f$^-$MUC1$^-$CD133$^-$(CD10/THY1)$^+$] cells isolated from 3-day cultured mammoplasty samples (n = 3).

(B) Bipotent and luminal-restricted CFC-enriched fractions were purified from six different 3-day precultured mammoplasty samples and then assayed for CFC with (10 μM) or without (+ DMSO) DAPT. Colony counts were performed 8 days later.

(C) Representative FACS profiles of the EPCAM$^+$ cells present in the 8-day colonies generated in (B). Staining with antibodies against EPCAM allowed human mammary cells to be discriminated from contaminating mouse fibroblast feeder cells. Mean fluorescence intensities (MFI) were calculated using FlowJo software (Ashland, OR). The arrow indicates the decreased expression of the luminal cell markers (MUC1 and/or CD133) when bipotent CFCs were cultured in the presence of DAPT as compared to the DMSO control.

(D) A representative FACS profile of the cells present in 8-day colonies generated from bipotent CFCs infected with a lenti-*dnMAML-GFP* or a control lenti-GFP virus (from one of two experiments) after staining the harvested cells as in (C). The mean fluorescence (MFI) values show the decreased expression of MUC1 and/or CD133 in the GFP$^+$ cells.

(E) Representative FACS profile of the cells present in 8-day colonies generated from bipotent CFCs or luminal-restricted CFCs infected with either Lenti-*shN3* or control Lenti-*NS* virus (from one of two experiments) after staining the harvested cells as in (C). The mean fluorescence (MFI) values show the decreased expression of MUC1 and/or CD133 in the GFP$^+$ cells.

The error bars represent the standard error of the mean.

## The CFC Assay Recapitulates Human Mammary Cell Differentiation

As a first use of these gene expression data we asked the extent to which the differentiation programs obtained in the in vitro CFC assay replicate those executed largely in vivo. Accordingly, we analyzed the levels of expression of human *LIF*, *GAS6*, *JAGGED2*, and *NOTCH3* in the predominantly myoepithelial and luminal cells present after 8 days in cultures initiated with EPCAM$^+$ cells enriched in either bipotent or luminal-restricted CFCs, respectively, and compared these with the directly isolated mature EPCAM$^+$ myoepithelial and luminal cells (that had been in culture for a maximum of 3 days). The four genes selected were chosen because they had shown large differences between the CFC-enriched fractions and the two populations of mature mammary cells (Table S14). To facilitate comparison of lineage-specific differences in the expression of these four genes in differently derived cells from multiple samples, we examined the ratio of the relative transcript levels in the two mature cell types isolated from each source (Figure 4A). As can be seen the relative levels of expression of the four selected genes in the two sets of culture-derived "mature" cells show a pattern of change that is similar to that characteristic of the in vivo generation of their mature counterparts.

## NOTCH Signaling Regulates the Restriction of Bipotent Mammary Cells to the Luminal Pathway

Given the gene profiling data suggesting that bipotent CFCs undergo marked changes in NOTCH receptor expression when they generate luminal-restricted CFCs, we hypothesized that NOTCH expression may have an important functional role in this lineage restriction process. To investigate this possibility, we cultured purified bipotent and luminal CFCs in the presence

or absence of a γ-secretase inhibitor, *N*-[N-(3, 5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT, 10 μM) and then measured the effect on the numbers and types of colonies obtained. To ensure that any effects on initial cell attachment were avoided, the DAPT (or vehicle control) was not added until after the first 16 hr of culture and then every other day thereafter. We found no effect of this treatment on the size or total number of colonies obtained in assays of either the bipotent or luminal CFC-enriched fractions (Figure 4B and data not shown). However, DAPT consistently and significantly ($p = 0.0006$, Student's t test) reduced the number of colonies that contained a detectable luminal component in the assays of the bipotent CFC-enriched fraction (Figure 4B). At the same time, the number of apparently "pure" myoepithelial colonies was proportionately increased ($p = 0.005$, Student's t test). In contrast, DAPT had no effect on the ability of the luminal-restricted CFCs to form colonies of differentiated luminal cells (Figure 4B), in spite of an equivalent extent of DAPT-mediated suppression of NOTCH signaling as shown by a reduction in *HEY1* transcripts in cultures initiated with either purified bipotent or luminal-restricted CFCs (2- and 6-fold, and 2- and 2-fold, respectively, in two experiments). The differential effects of DAPT on bipotent and luminal-restricted CFCs were confirmed by flow cytometric analysis of the immunophenotype of the cells present in 8-day colonies derived from purified progenitors. These analyses showed a specific loss of $(MUC1/AC133)^+$ cells exclusively among the progeny produced by the DAPT-treated bipotent CFCs and no change in output of $(CD10/THY1)^+$ cells (Figure 4C).

As a second approach to investigating the requirement of bipotent cells for NOTCH signaling to undergo restriction to the luminal pathway, we infected suspensions of purified bipotent CFCs with a lentivirus expressing a dominant-negative (dn) form of the human mastermind-like-1 gene (*MAML*) fused to GFP (Lenti-*dnMAML*) or a Lenti-*GFP* control virus and then assessed the phenotype of the $GFP^+$ cells harvested from bulk CFC assays 8 days later (Figure 4D). MAML is a coactivator of NOTCH signaling, and the dn cDNA used here was previously shown to suppress NOTCH signaling (Maillard et al., 2004; Weng et al., 2003). As shown in Figure 4D, forced expression of the *dnMAML* also decreased the output of $(MUC1/AC133)^+$ cells from the bipotent progenitors.

We next asked whether this restriction step was directed by specific activation of the NOTCH3 receptor using a similar strategy in which purified bipotent CFCs were first transduced with a GFP lentivirus also expressing a short hairpin (sh) *NOTCH3* RNA or a nonsilencing control RNA. As shown in Figure 4E, knockdown of the expression of *NOTCH3* receptor in the bipotent CFCs (90% and 85% in the two experiments undertaken; data not shown) reduced their ability to generate $(MUC1/AC133)^+$ luminal cells. Thus, activation of NOTCH3 appears to be critical for the restriction of most bipotent progenitors to the luminal pathway, and other NOTCH receptors cannot substitute for this activity.

## DISCUSSION

The development of specific assays for cells at distinct stages of differentiation in parallel with strategies for their selective isolation at high purities constitute critical steps in delineating the molecular mechanisms that regulate normal cell populations and are likely targets for malignant transformation. In applying this approach to the normal human mammary gland, we demonstrate here the robustness of the 2D colony assay conditions now available to distinguish undifferentiated (bipotent) and lineage-restricted mammary progenitors of human origin. Second, we describe a cell separation method that allows each of these progenitor types to be isolated routinely at purities of more than 30% and with less than 5% contamination with each other. Attainment of such purities depends, however, on the use of a starting population that has been cultured for an initial period of 3 days to selectively enrich for CFCs while also allowing cells damaged by the initial enzymatic dispersion process to be removed. The CFC purification scheme we describe also takes advantage of the fact that the majority of bipotent progenitors share a number of features with mature myoepithelial cells and lack features of mature luminal cells, whereas the converse is true for luminal-restricted progenitors (Russo and Russo, 2004; Stingl et al., 1998; Tsai et al., 1996), in addition to the previous observation that primitive mammary cells with significant proliferative potential express high levels of the $\alpha_6$-integrin (CD49f) in contrast to the majority of mature luminal and myoepithelial mammary cells (Stingl et al., 2001, 2006b).

Our first application of this technology enabled us to refute the notion that dual expression of cytokeratins 14 and 18 is a feature of undifferentiated or even primitive luminal-restricted progenitors in the normal human mammary gland. A second application was to generate molecular signatures for each of the four different cell populations from their global gene expression profiles. These allowed us to demonstrate that the differentiated cells produced in the colony assays resemble their counterparts isolated directly after only 3 days in vitro. However, the fact that all of the gene expression data were from cells that had been in culture for 3 days would predict that there would be some differences with the corresponding cell types present in vivo. The importance of culture conditions in replicating in vivo transformation events in human mammary cells has recently been documented in an elegant study by Ince et al. (2007).

The gene expression data presented here have also offered some new insights into the molecular characteristics of normal undifferentiated human mammary cells and their immediate luminal-restricted progeny. For example, we found that bipotent CFCs contain transcripts for the progesterone receptor but not the estrogen receptor, with an opposite picture for the luminal-restricted CFCs. Both of these patterns thus differ from the double-negative profile described for normal mouse mammary stem cells (Asselin-Labat et al., 2006; Shackleton et al., 2006) or any of the four categories of human breast cancer defined by expression profiling methods (Hu et al., 2006). We also examined the expression of transcripts for LIF and its receptor elements. This cytokine has been previously implicated in controlling apoptosis in the mammary gland (Kritikou et al., 2003) and is downregulated in some breast cancers (Hu et al., 2004). In addition, Dontu et al. (2003) noted that *LIF* expression was elevated in cultured human mammospheres, which represent mixed populations but contain elevated frequencies of primitive mammary cells. The purity of the discrete populations obtained here now demonstrates that the expression of *LIF* is markedly upregulated specifically at the time of progenitor commitment to the luminal differentiation pathway, at which time transcripts for the receptor

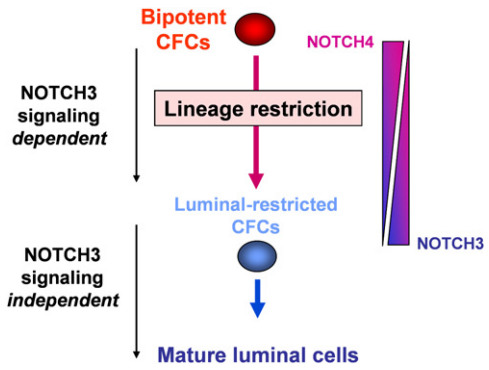


**Figure 5. Model of the Proposed Role of NOTCH in Regulating the Differentiation of Normal Human Mammary Progenitors**

The initial commitment process is uniquely dependent on NOTCH3 signaling, whereas subsequent execution of the luminal differentiation program proceeds independently.

are also present. It is therefore inviting to speculate that these cells may be the most important physiological targets for paracrine or autocrine effects of LIF in the normal mammary epithelium.

A similar stage-specific change in gene expression emerged from a survey of transcripts for genes associated with the evolutionary conserved NOTCH pathway. This pathway is widely involved in cell-fate decisions (Fuchs and Raghavan, 2002; Osborne and Minter, 2007) and has been shown to have a role in mammary gland development and transformation (Callahan and Egan, 2004; Farnie et al., 2007; Leong et al., 2007; Pece et al., 2004; Shi and Harris, 2006). Upon binding of their ligands, the NOTCH receptors (NOTCH1–4) undergo proteolytic cleavage, thereby releasing the NOTCH intracellular domain (NICD), which then moves into the nucleus and activates genes such as *HES1* and *HES6* (Callahan and Egan, 2004; Osborne and Minter, 2007). *Notch4* was first identified as an oncogene (*Int3*) in mice by MMTV insertional activation (Raafat et al., 2004). More recently, higher levels of *NOTCH4* transcripts were found in extracts of whole human mammospheres than in mature human mammary cells, and addition of soluble ligand to these cultures increased the generation of CFCs in the mammospheres (Dontu et al., 2004). We now show that *NOTCH4* gene expression is highest in undifferentiated human clonogenic mammary progenitors and is then markedly downregulated when these cells become committed to the luminal lineage but before they have lost their proliferative activity. We also found an opposite pattern to hold for *NOTCH3* and *HES6*, and to a lesser extent for *NOTCH1* and *NOTCH2*. This is interesting because the NOTCH4 receptor possesses a shorter EGF-like repeat sequence in its extracellular domain than NOTCH3 and also lacks cytokine response sequences in its intracellular domain (Allenspach et al., 2002; Bigas et al., 1998). Importantly, we also found that two different methods of nonspecifically blocking NOTCH signaling in bipotent progenitors (by exposing the cells to a γ-secretase inhibitor or by forced expression in them of a dn form of MAML) selectively prevented them from generating mature luminal progeny without affecting their ability to proliferate and generate mature myoepithelial cells. Nevertheless, the same treatment had no effect on the production of differentiated luminal cells from already committed luminal progenitors. Interestingly, knockdown of *NOTCH3* expression was sufficient to produce this same result, suggesting that NOTCH1 and -2, which appear to be coexpressed at similar levels in these cells, do not share this unique commitment-requiring function of NOTCH3 (Figure 5). The recent demonstration of an increased production during alveogenesis of basal cells in *Rbp-j* or *Pofut1* null mice (Buono et al., 2006) is consistent with this model, as is the original observation that WAP-*Int3* (=*Notch4*) transgenic mice display a block in mammary cell differentiation.

The present analysis thus illustrates a number of key changes in signaling molecules and points to a specific and previously unrecognized role of NOTCH in the process of restriction of bipotent progenitors to the luminal lineage. They also provide a comprehensive description of the gene expression profiles of distinct subsets of normal human mammary epithelial cells that establish an important new framework for further interrogation of pathways involved in the regulation of normal human mammary cell behavior.

## EXPERIMENTAL PROCEDURES

### Preparation and Isolation of Mammary Cell Subsets

Discarded reduction mammoplasty tissue was obtained with appropriate consent; the cells were processed, frozen, and thawed; and single-cell suspensions were prepared as described (Stingl et al., 2005). The cells were then cultured for 3 days on top of irradiated 3T3 mouse fibroblasts ($8 \times 10^3$ per $cm^2$) in EpiCult-B media supplemented with 5% fetal calf serum (FCS) (both from StemCell Technologies Inc.), and the adherent cells were resuspended using trypsin. $EPCAM^+$ cells (92.3% ± 4.7% $EPCAM^+$, n = 7) were isolated immunomagnetically using the Human EPCAM positive selection kit (StemCell Technologies). Subsequently, cells were stained with anti-MUC1 (1:100 dilution, StemCell Technologies), anti-CD133 (AC133, 1:100 dilution, Miltenyi Biotech), anti-α6 integrin (CD49f) conjugated to fluorescein isothiocyanate (FITC) (1:40 dilution, Becton Dickenson PharMingen), anti-CD10 (CALLA, 1:10 dilution, StemCell Technologies) conjugated to R-phycoerythrin (PE), and anti-CD90 (THY-1 = 5E10, 1:125 dilution, from Dr. P. Lansdorp, Terry Fox Laboratory, Vancouver, BC, Canada). A goat anti-mouse antibody conjugated to allophycocyanin (APC, 1:500 dilution, PharMingen) was used to detect cells expressing either MUC1 and/or CD133. IgG antibodies directly conjugated to FITC or PE were used as isotype controls. To distinguish between live and dead cells, propidium iodide (PI, Sigma) was added at 1 μg/ml to each sample (82% ± 4% $PI^-$ cells in 12 experiments). Cells were sorted on a FACS Vantage SE using gates that excluded 99.9% of events present in negatively stained control samples. Events with very high forward and side light scatter profiles were also excluded to improve sort efficiency. This was estimated to be 80%–98% as measured by sorting $10^5$ $MUC1^+$ cells into a tube, and viable cell count was obtained by a hemocytometer. To obtain RNA extracts for gene expression studies, aliquots of cells were sorted directly into Trizol reagent (Invitrogen, http://www.invitrogen.com/) and DNase treated, and the quality of the RNA samples was examined with the Agilent 2100 Bioanalyzer (Agilent Technologies).

### Lentiviral Transduction of Isolated Progenitors

Human *dnMAML-GFP* (Maillard et al., 2004) was obtained from Dr. A. Weng (Terry Fox Laboratory, Vancouver, BC, Canada) and cloned into the KA391 lentiviral vector and virus-containing supernatants generated as previously described (Imren et al., 2004). Bipotent and luminal-restricted CFC-enriched populations were infected in 100 μl of EpiCult-B growth media containing 5% FCS, $5 \times 10^5$ lentiviral particles (either Lenti-*dnMAML-GFP* or Lenti-*GFP*, or Lenti-*shN3-GFP* [Open Biosystems] or the Lenti-*shNS-GFP* [Open Biosystems]), and 5 μg/ml of protamine sulfate for 4 hr at 37°C. Subsequently, cells were washed three times in 2% FCS in Hank's solution and plated at clonal densities in CFC assays (see below).

### CFC Assays

CFC assays were performed as described (Stingl et al., 1998, 2001) with the modification that the plates were first precoated with type I collagen by adding a solution of 70 μg/ml (Vitrogen 100, Collagen Biomaterials) for 1 hr at 37°C

followed by washing with PBS to remove unpolymerized collagen. After 8 days the cultures were fixed with methanol acetone (1:1 ratio) and stained with Wright-Giemsa (Sigma), and colonies containing $\geq$50 cells were scored and typed using a microscope.

### Construction and Analysis of LongSAGE Libraries

RNA from each isolated subpopulation was converted to cDNA and then amplified as described (Zhao et al., 2007). LongSAGE libraries were constructed and sequenced using standard protocols (Saha et al., 2002). The relatedness of the libraries was determined by calculating Pearson correlation values, and the results were displayed as a tree diagram using Phylip software (http://evolution.genetics.washington.edu/phylip.html). The complete data set has been uploaded onto the GEO website (accession number GSE11395).

### Microarray Analyses

RNA from each sample was exponentially amplified as described (Iscove et al., 2002) except that the total number of PCR cycles was restricted to 42 and executed in a single reaction. From each sample, 50 pg aliquots were amplified in each of ten replicate 20 μl reactions, each yielding 2–3 μg of amplified product. The amplified cDNA preparations were then purified using the High Pure PCR Product purification kit (Roche Applied Science), end-labeled with Biotin-N6 ddATP (StemCore Laboratories, OHRI, Ottawa, Ontario, Canada), and hybridized to Affymetrix human X3P GeneChip arrays according to the manufacturer's protocol (Tietjen et al., 2003). The complete data set has been uploaded onto the GEO website (accession number GSE11395).

### Q-RT-PCR Analyses

cDNA was prepared from 1 ng of RNA using the SuperScript III reverse transcriptase enzyme (Invitrogen) according to the manufacturer's protocol and subjected to real-time PCR (7500 Sequence Detection System, Applied Biosystems) using gene-specific primers for each transcript analyzed. To quantify the relative expression of transcripts from the mature luminal and myoepithelial cells generated in vitro, RNA was extracted from the pooled 8-day progeny of cells from purified bipotent and luminal-restricted CFC-enriched fractions cultured at 50 cells per plate and pooled for extraction. Relative levels of expression of each test transcript were calculated by normalizing to the level of *GAPDH* transcripts in the same extract, and to compare results in particular subsets isolated from different samples, values were normalized to those measured in the bipotent CFC-enriched fractions.

### Statistical Analysis

Differentially expressed transcripts from different LongSAGE libraries were identified using a 95% statistical cut-off as determined by DiscoverySpace software (Robertson et al., 2007). Transcripts represented by tags expressed only once in any library (singletons) were excluded from further analysis. Differentially expressed transcripts from the Affymetrix hybridizations were analyzed using ArrayAssist software (http://Stratagene.com/software solutions) and R/BioConductor (Gentleman et al., 2004). The implemented GC-RMA algorithm, which accounts for probe composition, was used to subtract the background and determine the corrected probe intensity for each probe set cluster. Transcripts showing >1.5-fold differences with p values $\leq$ 0.05 (after Bonferroni-Holmes correction for multiple comparisons) in a Student's t test across replicates were deemed to be significant. Student's t test was used for pairwise comparisons to determine statistical significance. Volcano plots were generated using ArrayAssist software to identify the most differentially expressed transcripts. K-mean clustering was carried out using Cluster3.01 software (Eisen et al., 1998) using centered correlations.

## ACCESSION NUMBERS

All raw data are available online under Gene Expression Omnibus (GEO) accession number GSE11395.

## SUPPLEMENTAL DATA

The Supplemental Data include two figures and 14 tables and can be found with this article online at http://www.cellstemcell.com/cgi/content/full/3/1/109/DC1/.

## ACKNOWLEDGMENTS

The authors acknowledge excellent technical contributions from Darcy Wilkinson and Margaret Hale and the staff of the Flow Cytometry Facility of the Terry Fox Laboratory, which is funded in part by the Michael Smith Foundation for Health Research. Mammoplasty tissue was obtained with the assistance of Drs. Jane Sproul, Peter Lennox, Nancy Van Laeken, and Richard Warren. This project was funded by grants from Genome BC/Genome Canada, the Stem Cell Network, The BC/Yukon Division of the Canadian Breast Cancer Foundation (CBCF), and the Canadian Institute of Health Research (CIHR), including the CIHR Centre for Molecular Pathology at the BC Cancer Agency. A.R. was funded by CBCF and CIHR Fellowships. Y.Z. was funded by a Leukemia Research Foundation of Canada Fellowship, and J.S. was funded by a CBCF Fellowship and an Industrial Fellowship cofunded by the Natural Sciences and Engineering Research Council of Canada and StemCell Technologies, Inc.

Received: November 5, 2007
Revised: April 9, 2008
Accepted: May 15, 2008
Published: July 2, 2008

## REFERENCES

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA *100*, 3983–3988.

Allenspach, E.J., Maillard, I., Aster, J.C., and Pear, W.S. (2002). Notch signaling in cancer. Cancer Biol. Ther. *1*, 466–476.

Asselin-Labat, M.L., Shackleton, M., Stingl, J., Vaillant, F., Forrest, N.C., Eaves, C.J., Visvader, J.E., and Lindeman, G.J. (2006). Steroid hormone receptor status of mouse mammary stem cells. J. Natl. Cancer Inst. *98*, 1011–1014.

Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat. Cell Biol. *9*, 201–209.

Balazs, A.B., Fabian, A.J., Esmon, C.T., and Mulligan, R.C. (2006). Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. Blood *107*, 2317–2321.

Bigas, A., Martin, D.I., and Milner, L.A. (1998). Notch1 and Notch2 inhibit myeloid differentiation in response to different cytokines. Mol. Cell. Biol. *18*, 2324–2333.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell *118*, 635–648.

Bruneau, B.G., Bao, Z.Z., Fatkin, D., Xavier-Neto, J., Georgakopoulos, D., Maguire, C.T., Berul, C.I., Kass, D.A., Kuroski-de Bold, M.L., de Bold, A.J., et al. (2001). Cardiomyopathy in Irx4-deficient mice is preceded by abnormal ventricular gene expression. Mol. Cell. Biol. *21*, 1730–1736.

Buono, K.D., Robinson, G.W., Martin, C., Shi, S., Stanley, P., Tanigaki, K., Honjo, T., and Hennighausen, L. (2006). The canonical Notch/RBP-J signaling pathway controls the balance of cell lineages in mammary epithelium during pregnancy. Dev. Biol. *293*, 565–580.

Callahan, R., and Egan, S.E. (2004). Notch signaling in mammary development and oncogenesis. J. Mammary Gland Biol. Neoplasia *9*, 145–163.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Dontu, G., Jackson, K.W., McNicholas, E., Kawamura, M.J., Abdallah, W.M., and Wicha, M.S. (2004). Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. Breast Cancer Res. *6*, R605–R615.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Farnie, G., Clarke, R.B., Spence, K., Pinnock, N., Brennan, K., Anderson, N.G., and Bundred, N.J. (2007). Novel cell culture technique for primary ductal carcinoma in situ: role of Notch and epidermal growth factor receptor signaling pathways. J. Natl. Cancer Inst. *99*, 616–627.

Fuchs, E., and Raghavan, S. (2002). Getting under the skin of epidermal morphogenesis. Nat. Rev. Genet. *3*, 199–209.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555–567.

Howlin, J., McBryan, J., and Martin, F. (2006). Pubertal mammary gland development: insights from mouse models. J. Mammary Gland Biol. Neoplasia *11*, 283–297.

Hu, Y., Sun, H., Drake, J., Kittrell, F., Abba, M.C., Deng, L., Gaddis, S., Sahin, A., Baggerly, K., Medina, D., and Aldaz, C.M. (2004). From mice to humans: identification of commonly deregulated genes in mammary cancer via comparative SAGE studies. Cancer Res. *64*, 7748–7755.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics *7*, 96.

Imren, S., Fabry, M.E., Westerman, K.A., Pawliuk, R., Tang, P., Rosten, P.M., Nagel, R.L., Leboulch, P., Eaves, C.J., and Humphries, R.K. (2004). High-level beta-globin expression and preferred intragenic integration after lentiviral transduction of human cord blood stem cells. J. Clin. Invest. *114*, 953–962.

Ince, T.A., Richardson, A.L., Bell, G.W., Saitoh, M., Godar, S., Karnoub, A.E., Iglehart, J.D., and Weinberg, R.A. (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. Cancer Cell *12*, 160–170.

Iscove, N.N., Barbara, M., Gu, M., Gibson, M., Modi, C., and Winegarden, N. (2002). Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. Nat. Biotechnol. *20*, 940–943.

Jones, C., Mackay, A., Grigoriadis, A., Cossu, A., Reis-Filho, J.S., Fulford, L., Dexter, T., Davies, S., Bulmer, K., Ford, E., et al. (2004). Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. Cancer Res. *64*, 3037–3045.

Kordon, E.C., and Smith, G.H. (1998). An entire functional mammary gland may comprise the progeny from a single cell. Development *125*, 1921–1930.

Kritikou, E.A., Sharkey, A., Abell, K., Came, P.J., Anderson, E., Clarkson, R.W., and Watson, C.J. (2003). A dual, non-redundant, role for LIF as a regulator of development and STAT3-mediated cell death in mammary gland. Development *130*, 3459–3468.

Leong, K.G., Niessen, K., Kulic, I., Raouf, A., Eaves, C., Pollet, I., and Karsan, A. (2007). Jagged1-mediated Notch activation induces epithelial-to-mesenchymal transition through Slug-induced repression of E-cadherin. J. Exp. Med. *204*, 2935–2948.

Lewis, M.T., Ross, S., Strickland, P.A., Snyder, C.J., and Daniel, C.W. (1999). Regulated expression patterns of IRX-2, an Iroquois-class homeobox gene, in the human breast. Cell Tissue Res. *296*, 549–554.

Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C.G., Merajver, S.D., Dontu, G., and Wicha, M.S. (2008). BRCA1 regulates human mammary stem/progenitor cell fate. Proc. Natl. Acad. Sci. USA *105*, 1680–1685.

Maillard, I., Weng, A.P., Carpenter, A.C., Rodriguez, C.G., Sai, H., Xu, L., Allman, D., Aster, J.C., and Pear, W.S. (2004). Mastermind critically regulates Notch-mediated lymphoid cell fate decisions. Blood *104*, 1696–1702.

Osborne, B.A., and Minter, L.M. (2007). Notch signalling during peripheral T-cell activation and differentiation. Nat. Rev. Immunol. *7*, 64–75.

Pece, S., Serresi, M., Santolini, E., Capra, M., Hulleman, E., Galimberti, V., Zurrida, S., Maisonneuve, P., Viale, G., and Di Fiore, P.P. (2004). Loss of negative regulation by Numb over Notch is relevant to human breast carcinogenesis. J. Cell Biol. *167*, 215–221.

Raafat, A., Bargo, S., Anver, M.R., and Callahan, R. (2004). Mammary development and tumorigenesis in mice expressing a truncated human Notch4/Int3 intracellular domain (h-Int3sh). Oncogene *23*, 9401–9407.

Robertson, N., Oveisi-Fordorei, M., Zuyderduyn, S.D., Varhol, R.J., Fjell, C., Marra, M., Jones, S., and Siddiqui, A. (2007). DiscoverySpace: an interactive data analysis application. Genome Biol. *8*, R6.

Russo, J., and Russo, I.H. (2004). Development of the human breast. Maturitas *49*, 2–15.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002). Using the transcriptome to annotate the genome. Nat. Biotechnol. *20*, 508–512.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Shi, W., and Harris, A.L. (2006). Notch signaling in breast cancer and tumor angiogenesis: cross-talk and therapeutic potentials. J. Mammary Gland Biol. Neoplasia *11*, 41–52.

Smalley, M.J., Titley, J., Paterson, H., Perusinghe, N., Clarke, C., and O'Hare, M.J. (1999). Differentiation of separated mouse mammary luminal epithelial and myoepithelial cells cultured on EHS matrix analyzed by indirect immunofluorescence of cytoskeletal antigens. J. Histochem. Cytochem. *47*, 1513–1524.

Stingl, J., Eaves, C.J., Kuusk, U., and Emerman, J.T. (1998). Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation *63*, 201–213.

Stingl, J., Eaves, C.J., Zandieh, I., and Emerman, J.T. (2001). Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. Breast Cancer Res. Treat. *67*, 93–109.

Stingl, J., Raouf, A., Emerman, J.T., and Eaves, C.J. (2005). Epithelial progenitors in the normal human mammary gland. J. Mammary Gland Biol. Neoplasia *10*, 49–59.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006a). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.

Stingl, J., Raouf, A., Eirew, P., and Eaves, C.J. (2006b). Deciphering the mammary epithelial cell hierarchy. Cell Cycle *5*, 1519–1522.

Tietjen, I., Rihel, J.M., Cao, Y., Koentges, G., Zakhary, L., and Dulac, C. (2003). Single-cell transcriptional analysis of neuronal progenitors. Neuron *38*, 161–175.

Tsai, Y.C., Lu, Y., Nichols, P.W., Zlotnikov, G., Jones, P.A., and Smith, H.S. (1996). Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis. Cancer Res. *56*, 402–404.

van Tuyl, M., Liu, J., Groenman, F., Ridsdale, R., Han, R.N., Venkatesh, V., Tibboel, D., and Post, M. (2006). Iroquois genes influence proximo-distal morphogenesis during rat lung development. Am. J. Physiol. Lung Cell. Mol. Physiol. *290*, L777–L789.

Welm, B.E., Tepera, S.B., Venezia, T., Graubert, T.A., Rosen, J.M., and Goodell, M.A. (2002). Sca-1(pos) cells in the mouse mammary gland represent an enriched progenitor cell population. Dev. Biol. *245*, 42–56.

Weng, A.P., Nam, Y., Wolfe, M.S., Pear, W.S., Griffin, J.D., Blacklow, S.C., and Aster, J.C. (2003). Growth suppression of pre-T acute lymphoblastic leukemia cells by inhibition of notch signaling. Mol. Cell. Biol. *23*, 655–664.

Zhao, Y., Raouf, A., Kent, D., Khattra, J., Delaney, A., Schnerch, A., Asano, J., McDonald, H., Chan, C., Jones, S., et al. (2007). A modified polymerase chain reaction-long serial analysis of gene expression protocol identifies novel transcripts in human CD34+ bone marrow cells. Stem Cells *25*, 1681–1689.