AWARD NUMBER:

W81XWH-12-1-0106

TITLE:
Use of a Novel Embryonic Mammary Stem Cell Gene Signature to Improve Human Breast Cancer Diagnostics and Therapeutic Decision Making

PRINCIPAL INVESTIGATOR:
Geoffrey M. Wahl

CONTRACTING ORGANIZATION: The Salk Institute for Biological Studies

La Jolla, CA 92037

REPORT DATE:

December 2015

TYPE OF REPORT:

Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2015	Final	30 Sep 2012 – 29 Sep 2015

4. TITLE AND SUBTITLE

Use of a Novel Embryonic Mammary Stem Cell Gene Signature to Improve Human Breast Cancer Diagnostics and Therapeutic Decision Making

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0106

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Wahl, Geoffrey M., Perou, Charles, Spike, Benjamin, Lasken, Roger

E-Mail: wahl@salk.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road
La Jolla, CA 92037

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Our major goals have remained the same as described in the previous report. In Aim 1, we ascertained Fetal Mammary Stem Cell (fMaSC) signatures to determine whether they chemotherapy response and metastasis in different breast cancer intrinsic subtypes (AIM1). Aim 2 developed single cell sequencing protocols to identify gene expression programs that correlate with entry into and exit from the mammary stem cell state. The data obtained should enable us to 1) better categorize distinct cell types within the fMaSC population, 2) identify biomarkers for prospective stem cell purification and in situ localization, and 3) identify candidate stem cell regulatory pathways that should reveal therapeutic targets and improved prognosticators and response biomarkers. Over the past year we published two papers using the fMaSC signatures to identify expression features correlating with chemotherapeutic response of human breast cancer patients. We obtained improved fMaSC profiles through RNA Sequencing and validated our single cell RNA Sequencing and analytical pipeline over multiple developmental time points. We found shared co-expression of genes in fMaSCs and certain breast cancers. The single cell profiles also provide candidate biomarkers for fMaSC-like cells in tumors. The two publications resulting from this Idea Award Expansion largely complete the proposed Aims and provide the bases for substantial future work funded by other sources.

15. SUBJECT TERMS
Breast Cancer Prognosis, Mammary Stem Cells, Embryonic Development, Single Cell Transcriptomics

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	14	USAMRMC

Table of Contents

Page

Introduction

This Idea Award Expansion proposed two aims to capitalize on two discoveries made under the originally funded Idea Award; 1) identify and transcriptionally profile Fetal Mammary Stem Cells (fMaSCs), and 2) uncover of molecul similarities between fMaSCs and human breast cancers.

Our first Aim derived refined transcriptomic profiles from cell fractions obtained by fluorescence activated cell sorting (FACS). We used both publically available databases of the adult human and mouse luminal, myoepithelial, luminal progenitor, and MaSC-enriched fractions, and the fetal MaSC-enriched and stromal populations. We developed a meta-analysis approach to derive a consensus gene signature for each fraction by using data from all published human studies that isolated the indicated FACS fraction. This approach reduced signature variability generated by technical and biologic variability. Our studies revealed similarities between the normal human cell types and intrinsic breast cancer subtypes. Similar analyses performed with the fetal and adult mouse cell fractions enabled generation of correlations with human intrinsic breast cancer subtypes, and more precise assignment of realtionships of genetically engineered mouse mammary cancer models to both normal cell types and to the human intrinsic subtypes. An important conclusion from these studies is that enrichment for the human luminal progenitor signature, and for one of the features of the fMaSC signature, predicts pathologic complete response to neoadjuvant anthracycline/taxane based chemotherapies across all human cancer subtypes. This correlation pertains even after controlling for intrinsic subtype, proliferation, and clinical variables. On the other hand, another feature of the fMaSC signature predicts for resistance to chemotherapy. These results will be described in greater detail in Dr. Perou's Progress Report, and can be found in a collaborative publication: Pfefferele, Spike, Wahl, and Perou (2015) Breast Cancer Res Treat (pubished online, Jan. 10, 2015; DOI 10.1007/s10549-014-3262-6)

Aim2 comprised the major focus of the Wahl lab. We used single cell RNA-sequencing to deconvolute the fMaSC population into its component cell types. Our long term goal is to identify fetal genes and pathways that can be used for early detection of triple negative breast cancers, to elucidate fetal pathways uniquely used by breast cancers exhibiting enrichment for the fMaSC signature as the basis for developing targeted therapeutic or immunotherapeutic strategies. By studying single cell transcriptional patterns across developmental time, we are also attempting to create precise signatures for the pathways that enable entry into and exit from the fMaSC state.

Keywords

Breast Cancer Prognosis, Mammary Stem Cells, Embryonic Development, Single Cell Transcriptomics

Overall Project Summary

Statement of Work: (W=Wahl Lab; P=Perou Lab; L=Lasken Lab)

Task 1. Embryonic mammary stem cell signature refinement to delineate fMaSC traits in human cancers and to identify new targets for cancer stem cell directed therapies

Task 1 is mainly the work of the Perou lab. Consequently, Dr. Perou will submit under separate cover the summary of his lab's progress to complete the proposed tasks. I note that much of the data reporting this progress is published in Pfefferle et al, 2015 (see references below).

1a. The Perou lab will obtain the gene expression raw data of the fMaSC and fStroma samples previously characterized by the Wahl lab. First, using the current fMaSC signature of 600 genes, a score for each gene will be assigned based on its differential expression across the two groups. Second, using a cross validation approach, a genomic predictor will be created using the smallest gene list possible that can correctly discriminate the fMaSC vs. fStroma samples. Finally, samples used for the identification of the minimum gene list will be re-run onto the Fluidigm BioMark platform, and the optimal classification ability of the new genomic predictor will be re-tested. **P,W** (months l-2).

lb. The latest and most extensive cell line microarray database of the Perou Lab will be used for this analysis. This data set includes 40 breast cancer cell lines, 12 human mammary epithelial and fibroblast cell lines (primary and immortalized), 3 human embryonic stem cell lines and 3 mesenchymal stem cell lines. For each cell line, the genomic Euclidian distance to the fMaSC and fStromal centroids will be calculated, and the ratio of both distances will be the final "enrichment score". **P** (months 2-4)

lc. In addition, to further compare the levels of gene expression of the fMaSC- and fStromal-enriched populations with the Perou lab's cell lines, -12 fMaSC/fStromal samples will be collected, RNA isolated, amplified and hybridized onto the Perou Lab Whole Genome Custom Array Platform, and their gene expression profiles compared to the rest of cell lines using supervised and unsupervised hierarchical clusterings. **P,W** (months 3-8).

1d. The association of the MaSC signature with pathological complete response will be evaluated across multiple data sets with annotated clinical data and where gene expression microarrays had been performed in the pre-treatment samples. Each sample will be assigned an enrichment score as described above. The association of the score with pathCR will be evaluated in all patients and also within each intrinsic molecular subtype as determined by the PAM50-subtype predictor. For those data sets with survival data, association with survival outcomes will also be evaluated using univariate and multivariate Cox-model analyses. Finally, the enrichment scores during and after chemotherapy will be calculated in the samples of the ISPY-1 trial and also in one publicly available data set where pre- and post-treatment samples after single agent docetaxel or endocrine therapy were profiled. **P** (months 5-8).

I e. The association of the MaSC signature with the development of distance metastases will be evaluated using univariate and multivariate analyses in >800 primary tumor samples were the location and time of the first site of distant relapse was documented (within all patients and also each of the intrinsic subtypes). A signature enrichment score will be calculated as described above. A similar approach will be done for those samples of the UNC database where match primary and metastatic lesions were profiled. **P** (months 7-10).

lf. The genes comprising refined fMaSC and fStromal signatures will be prioritized according to their novelty as potential therapeutic targets and tractability for functional testing. Scientific/Clinical literature will be surveyed to determine novelty. Scientific/Clinical and Company literature will be surveyed to determine the availability of reagents for functional testing. Cell line expression profiles from the ATCC breast cancer collection and the lines referenced in task la. will be bioinformatically evaluated for the presence of signatures suggesting relevance of fMaSC and fStromal genes in the prioritized list and these cell lines will be selected for functional analysis. **P,W** (months 2-l 0)

lg. Reagents such as small molecule inhibitors, receptor specific antibodies and gene clones or inhibitory RNA constructs will be collected for the top candidates. Activating and deactivating genes will be

cloned into existing inducible lentiviral vectors. High titer lentivirus will be produced and validated for inducibility and RNA inhibition or protein production will be validated as appropriate. Other targeted reagents will be validated using standard molecular biological approaches where necessary. **W** (Months 8-14)

1h. The ability of viral vectors or other reagents to specifically impact their molecular targets will be examined in breast cancer cell lines through 2D and 3D culturing systems and standard molecular biological approaches (e.g. western blotting, immunofluorescent/cytochemistry, RT-PCR etc. examining receptor phosphorylation, protein localization,.mRNA abundance, etc.). Cellular effects on proliferation, survival and migration will also be assayed. **W** (months 14-18)

Ii. Cell lines exhibiting biological responses to activation or inhibition of fMaSC and fStromal pathways in vitro will be transduced with lentiviral vectors and injected as xenografts into immune compromised mice. Tumor growth and metastasis will be evaluated in real time using luminescent imaging. **W** (months 16-24)

Task 2. Embryonic mammary stem cell signature refinement using RNA-seq and functional validation

2a. The Wahl lab will obtain timed pregnant female mice, obtain embryos from EI8.5, dissect mammary rudiments, isolate fMaSC and fStroma by methods they have developed. The cells will be flow sorted to obtain fMaSC enriched (CD49fhighCD24highNCAM-) and fStromal (CD49f+, CD24-/+) populations. **W** (months 1-2)

>>>This has been accomplished (see data below).

The Lasken lab will receive fMaSC enriched and fStromal cells from the Wahl lab. The fMaSC enriched cells will be micromanipulated to obtain single cells, which will then be lysed to preserve RNA integrity and maximize efficiency for generating eDNA for SOLiD sequencing. **L** (months 1-2).

>>>>This was accomplished (see data below).

2b. cDNA will be generated from each cell and amplified by a PCR method under conditions suitable for subsequent SOLiD DNA sequencing **L** (months 2-3).

2c. Each single cell cDNA preparation will be pre-screened for the fMaSC phenotype based on expression of Kl4 and K8 using qRT-PCR **L,W** (months 2-3).

2d. The highest quality eDNA samples representing 10 additional K14+K8+, 5 K14+K8- and 5 Kl4-K8+ cells will be RNA sequenced by the SOLiD sequencing method at a level generating about 40 million sequence reads/cell. **L** (months 3-6).

>>>>Tasks 2a-2d were initially done manually using a variety of methods. However, approximately 18 months ago, a Fluidigm C1 single cell microfluidic instrument capable of lysing 96 single cells in situ, and then preparing cDNA samples from them was obtained by the California Institute for Regenerative Medicine (CIRM), which is next to the Salk. We were granted approval to use this instrument. We isolated approximately 100 individual cells from e18.5 in

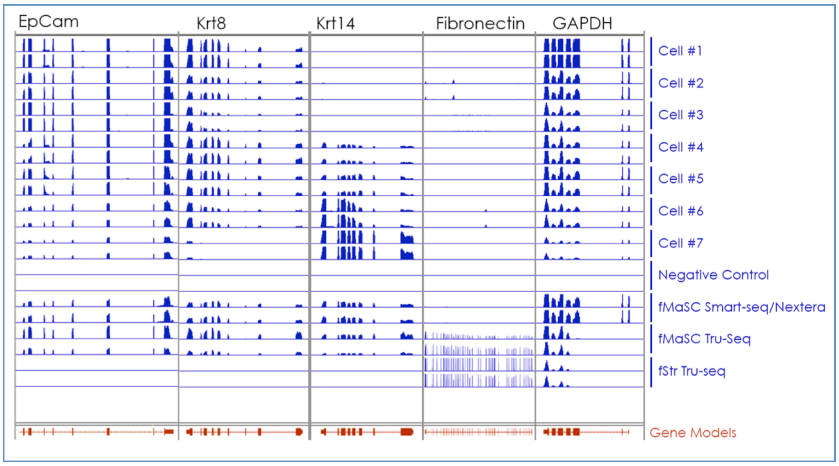


Figure 1. Genomic sequence alignments of RNA derived sequences from several individual fMaSC cells and control samples. The alignments show high concordance with annotated exon structures (Gene Models). The data also show high technical reproducibility between replicate sequencing experiments for each sample. Controls are comprised of pools of fMaSC cells processed using the same biochemistry as the C1 protocol (fMaSC Smart-seq/Nextera) or lacking reverse transcriptase (Negative control). An fMaSC pool and a Stromal pool (fStr) processed by an alternative approach (Tru-Seq) that works on bulk samples are also provided

several independent experiments and loaded them onto the C1 to obtain cDNA. An example of the data obtained is shown in Figure 1. Of note, this instrument enables one to evaluate the cells microscopically using a live-dead cell dye to restrict all RNA-sequencing data to live cells. As can be seen, we detected cells co-expressing K14 and K8 RNA (Figure 1, Cells #4-6), and the RNA seq data follow the gene models very well.

2e. RNA-Seq data will be analyzed to discover additional genes and gene clusters associated with the fMaSC cells. These data will be combined with the analysis of 10 KI4+K8+ cells that are currently be sequenced with funding from the JCVI and a Salk Cancer Center Starter award to yield a total of 20 KI4+K8+ cell analyzed. The sequence and data analyses will be conducted jointly by the Lasken and Wahl labs. **L, W** (months 7-13).

2f. A list of markers will be generated through bioinformatics analysis of single cell RNA-Seq data to identify markers associated with distinct cell types. The literature will be surveyed for the availability of reagents for the prospective isolation of the distinct cell types using the identified markers. Reagents will be acquired. Cells will be isolated based on these markers and the fidelity of separation of the individual cell types and per based analysis resorting will be carried out to assess purity of sorting. **P, L, W** (months 12-15)

>>>>Our initial evaluation of the gene expression profiles within the fMaSC population at single cell resolution showed the cells to be heterogeneous with no clear subpopulation likely to correspond to a distinct stem cell subpopulation. However, through the use of the C1 instrument, we were able to broaden our research approach to include additional developmental states that could be used to delineate gene expression changes that define the gain and loss of the stem cell phenotype over the course of development. That is, instead of focusing just on E18 cells, we decided to obtain cells from throughout development so that we could have a data set that would position us to identify the pathways that are altered in going from the pre-stem cell state at e16, to the stem cell state at e18, and then into the differentiated myoepithelial and luminal lineages associated with adult mammary development. We have now sequenced hundreds of cells from E18, P0, P4 and adult mice, and have clustered the data using multiple bioinformatic methods to assign cellular phenotypes. As one example, we used the Monocle strategy to infer lineages based on generation of minimum spanning trees of transcriptional relatedness (Figure 2 A,B). We then derived an independent approach that does not use the Monocle assumption of direct lineage relationships between cells of close transcriptional relatedness, which resulted in a very similar outcome (Figure 2 C-E)
These methods proved robust

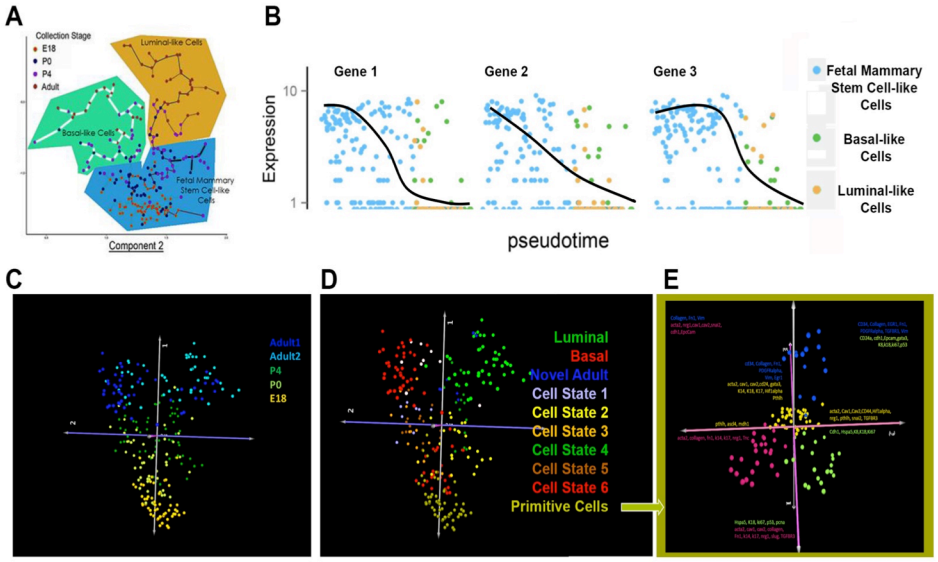


Figure 2. Unsupervised identification of cell types and candidate regulatory genes from single cell data. A) Monocle plot of single cell relatedness (i.e. proximity in the graph), minimum spanning tree model of differentiation ('pseudotime') and identification of three putative cell types (Luminal-like, Basal-like and fMaSC-like). B) Expression levels of three highly fMaSC associated genes in single cells that have been reorganized according to their position along the pseudotime, minimum spanning tree. Note: The majority of adult cells are plotted on the x-axis for these three genes as they are rarely expressed in adult cells (i.e. y=0). C-D) An alternative approach yields similar but more refined results. C) Using our alternative approach cells are found to be distributed along a continuum from E18 to adult (vertical axis). Clustering of these cells according to gene expression ranks, identifies known Luminal and Basal adult cell types, a novel adult cell type and several cell states along the continuum from the most primitive cells to the adult. D) Fine scale analysis of the most primitive cells identifies genes associated with the earliest differentiation events as bi-phenotypic cells become more luminal (green), more basal (pink) or more niche related (blue).

for separating cells according to the developmental time at which they were obtained, and to the lineages to which they are most related. Further analysis based on these approaches has enabled us to define a subset of cells from E18 that appear to be uncommitted to either the luminal or myoepithelial lineages, and yet to express genes indicative of both. For example, these cells express both luminal and myoepithelial cytokeratins, as well as lineage specification genes including Sox10, GATA3, and Elf5. The methods and examples of data obtained from these analyses were presented in the Progress Report submitted last year. A manuscript describing these studies is currently being prepared and is only awaiting data obtained from cells in the "pre-" stem cell state at E15-16.
We have experienced significant technical problems using the C1 instrument to analyze cells derived from E15-16. Recently, we have returned to working with the Lasken lab to sort cells from E15-16 cells directly into wells of a 384 well plate, and then used the SMART-seq 2 protocol to prepare cDNA libraries. Sequencing of these libraries is currently in progress.

2g. Cells will be sorted using population specific markers. In vitro colony growth, serial replating ability and immunofluorescent analysis ofbipotent progeny will be evaluated for each candidate marker. **W** (month 13-18)

2h. Markers yielding stem cell phenotypes in vitro will be used to sort fetal mammary cells. Cells will be transplanted at limiting dilution to reconstitute murine mammary glands additionally single cells will be transplanted to reconstitute murine mammary glands. **W** (months 18-24)

2i. Additional RNA-seq will be performed to refine the data obtained from validated fMaSCs. A second SOLiD sequencing run will be carried out on eDNA from ten single cells to refine and test conclusions obtained in the first year of the grant. **L,W** (Months 14-18).

>>>>We have accomplished the major goals of Aims 2g-i, and provide the following as one important example of the value and relevance of the data obtained. The results, summarized briefly below, identified the cell state regulator SOX10 as a developmental control gene able to identify and highly enrich for fMaSCs that is also required for the fMaSC state (Dravis et al (2015), Cell Reports, v. 12, pgs 2035-2048).

Our expression profiling identified SOX10 as one of the most differentially regulated genes in fMaSCs using both microarrays and single cell RNA sequencing (Dravis et al, Figure 1A, pg 2036). We obtained a mouse expressing an H2B-Venus transgene under the control of the SOX10 endogenous promoter. The mammary epithelial cells in the embryonic rudiments were brilliantly labeled, while there was little if any staining in the surrounding stroma (Dravis et al, Figure 2A, pg. 2038). We used FACS to obtain various cell fractions on the basis of their expression of different levels of EpCAM or SOX10 (i.e., venus fluorescence). We found that only those cells that were EpCAM+ and SOX10-high exhibited all of the properties expected of fMaSCs: generation of polarized organoids in vitro, ability to generate full functional mammary outgrowths from limited numbers of cells transplanted into de-epithelialized fat pads, and ability to "self-renew" as assayed by multiple rounds of transplantation, or dissociation and re-formation of spheres in vitro (Dravis et al, Figure 2E, F, G, H, I, pg 2038). We also obtained mice with floxed SOX10 genes, deleted the SOX10 genes from fMaSCs in vitro, and found that fMaSCs lacking SOX10 expression no longer formed organoids in vitro or transplanted in vivo (Dravis et al, Figure 4A-E, pg. 2041). Finally we showed that over-expressing SOX10 led to two very important phenotypes. First, after short periods of expression, we found the fMaSCs could form secondary organoids with much higher efficiency than if they did not express high SOX10 levels. Second, if SOX10 expression was maintained at high levels, the fMaSCs lost expression of epithelial markers, no longer expressed luminal or basal cytokeratins, gained expression of vimentin, and became motile but non-proliferative (Dravis et al, Figure 5A-F, pg. 2042). In other words, they acquired many characteristics of cells that had undergone an epithelial-mesenchymal transition. Importantly, reducing SOX10 expression in the cells that had moved away from the organoids to set up solitary "satellites" resulted in reversion of the cells to an epithelial state, re-entry into the cell cycle, and restoration of their ability to generate both luminal and myoepithelial descendants. In other words, the stem state was readily reversed depending on SOX10 levels.
We have begun to search for in vivo conditions that regulate SOX10 in the mammary gland and that could be relevant to fMaSC genesis and breast cancer biology. We found that FGF10 specifically induces SOX10 transcription, and that either leaving SOX10 out of the culture medium, or using an pan-FGF receptor inhibitor, prevents SOX10 transcriptional activation, and prevents fMaSCs from undergoing an EMT (Dravis et al, 2015, Figures 1B-E, pg 2036). Interestingly, FGF10 is one of the factors produced during wound healing. We speculate that as wound signatures have been correlated with initiation and progression of breast cancer, that

exposure of the fMaSC-like cells we have documented to be present in basal-like breast cancers may enable them to acquire motility, depart the local tumor environment, and metastasize to distant sites at which, if exposed to a lower FGF environment, then may reverse their phenotype, become more stem-like, and produce a heterogeneous cellular mass at an ectopic location.

2j. Identification of gene signatures corresponding to fMaSC from bioinformatic analysis (task 2e,f) and bioinformatic refinement/reduction of the signature. Selection of candidate markers for analysis of fMaSC contribution to archival tumor samples and tissue analysis **P,L,W** (months 12-24)

>>>>We found that elevated Sox10 expression is found in Basal-like and some Claudin-low human breast cancers (Dravis et al, 2015, Figure 1F, pg 2036).

2k. Immuno-histochemical and in situ analysis of archival tumor tissue. **P,W** (months 12-24).

>>>>We are now developing the collaborations we need to obtain relevant samples from UCSD, and we continue to work with Dr. Perou to analyze his human and mouse tissue samples as we derive additional informative signatures. Unfortunately, we have found no Sox10 antibodies suitable for IHC or IF analyses.

Key Research Accomplishments:

Aim 1

1. Development of a meta-analysis approach to derive more precise signatures for normal mammary cell luminal, progenitor, myoepithelial, and stem cell populations from human and mouse systems. This method proved more robust than using single studies for analysis, and sets a precedent for use of such meta-analysis-derived signatures in future studies.
2. Application of refined signatures based on normal mammary cell types to analysis of human breast cancers and mouse cancer models to determine which normal cell types correspond most closely to cancers in each species.
3. Use of single sample classifiers revealed diversity of cellular relationships among each GEMM and human breast cancer intrinsic subtype.
4. Demonstration that the human luminal progenitor and one feature of the mouse fMaSC signature correlate with pCR across all human breast cancer subtypes, and retains significance in multi-variable analyses including proliferation, subtype, and clinical parameters. Importantly, one feature of the fMaSC profile associated with luminal attributes predicted for poor response to anthracycline/taxane based chemotherapy for patients whose tumors display enrichment for this profile.

Aim 2

1. Obtained transcriptomes from hundreds of individual cells across four developmental time points critical for understanding mechanisms of acquisition and loss of the stem cell state during mouse mammary development.
2. Use of transcriptomic data to identify candidate transcriptional regulators relevant to acquisition of mammary stemness. Identification of SOX10 as one such gene.
3. Demonstrated fetal mammary cells expressing SOX10 uniquely identify the stem cell population. This discovery facilitated purification of the most pure fMaSC population to date, which enabled obtaining more precise transcriptomic data.
4. Genetic strategies were employed to show that SOX10 is required for fMaSC function in vitro and in vivo.
5. Developed a genetic system to enable analysis of the effects of SOX10 overexpression. These studies showed that persistent SOX10 expression preserves fMaSC multipotentiality, but long term high SOX10 expression causes fMaSCs to undergo a mesenchymal transistion that does not correlate with elevated levels of Slug, Snail, Zeb1, or Twist as reported for other systems. The mesenchymal transition was reversed upon reducing SOX10 levels.
6. Gene expression and functional studies revealed a positive feedback loop between FGF signaling and SOX10. Elevated SOX10 led to upregulation of potentiators of FGF signaling, and down regulation of FGF signaling antagonists.

Conclusion

Our new data are consistent with our previous studies showing that fMaSC signatures contain unique combinations of expressed genes with relevance to human breast cancer biology, including the response of breast cancers of all intrinsic subtypes to chemotherapy. We have thus developed a potentially useful metric for clinical decision-making. We continue to improve methods for doing single cell RNA-seq, and for bioinformatically analyzing the results. These studies revealed the potential relevance of SOX10 to fMaSC biology, which we established using a combination of in vitro and in vivo approaches.

Publications:

1. Pfefferle, A.D., Spike, B.T., Wahl, G.M., and Perou, C.M. (2015) Breast Cancer Res. Treat. 149:425-43. DOI 10.1007/s1059-014-3262-6. PMCID: PMC4308649

2. Dravis, C., Spike, B.T., Harrell, C.J., Johns, C., Trejo, C., Southard-Smith, M.E., Perou, C.M. and Wahl, G.M. (2015) Sox10 regulates stem cell and mesenchymal states in mammary epithelial cells. Cell Reports, *pii: S2211-1247(15)00925-0.* PMCID:PMC4591253

Abstracts and Presentations:

Geoff Wahl, Ph. D.
1/1/15: Speaker, Moores Cancer Center's Head & Neck Cancer Retreat, UCSD
4/10/15: Keynote Lecture, University of Arizona Cancer Center
5/15/15: Speaker, University of Chicago Seminar Series, University of Chicago

Chris Dravis, Ph.D.
4/22/15 AACR Annual Meeting, Philadelphia, PA (Poster)

Benjamin T. Spike, Ph.D.
1/12/15 University of Utah, Huntsman Cancer Institute
4/2/15 University of California, San Diego, Department of Pathology
4/30/15 University of Colorado, Denver, Department of Pathology

"Inventions, Patents and Licenses"
None.

Reportable Outcomes
- fMaSC gene signatures correlated to chemotherapeutic response
- Sequencing and analytical pipeline for single cell RNA Sequencing
- SOX10 as a marker and functionally relevant transcription factor contributing to the mammary stem cell state
- Involvement of FGF signaling in SOX10 induction, and regulation of stem and mesenchymal states in the mammary gland

Other Achievements
G.M. Wahl
- Successfully competed for Outstanding Investigator Award funding: 7 years, ~$600,00 per year

B.T. Spike
- Obtained Assistant Professor position at University of Utah

References
None

Appendices
None

Breast Cancer Res Treat (2015) 149:425–437
DOI 10.1007/s10549-014-3262-6

PRECLINICAL STUDY

Luminal progenitor and fetal mammary stem cell expression features predict breast tumor response to neoadjuvant chemotherapy

Adam D. Pfefferle · Benjamin T. Spike · Geoff M. Wahl · Charles M. Perou

Received: 22 December 2014 / Accepted: 23 December 2014 / Published online: 10 January 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Mammary gland morphology and physiology are supported by an underlying cellular differentiation hierarchy. Molecular features associated with particular cell types along this hierarchy may contribute to the biological and clinical heterogeneity observed in human breast carcinomas. Investigating the normal cellular developmental phenotypes in breast tumors may provide new prognostic paradigms, identify new targetable pathways, and explain breast cancer subtype etiology. We used transcriptomic profiles coming from fluorescence-activated cell sorted (FACS) normal mammary epithelial cell types from several independent human and murine studies. Using a meta-analysis approach, we derived consensus gene signatures for both species and used these to relate tumors to normal mammary epithelial cell phenotypes. We then compiled a dataset of breast cancer patients treated with neoadjuvant anthracycline and taxane chemotherapy regimens to determine if normal cellular traits predict the likelihood of a pathological complete response (pCR) in a multivariate logistic regression analysis with clinical markers and genomic features such as cell proliferation. Most human and murine tumor subtypes shared some, but not all, features with a specific FACS-purified normal cell type; thus for most tumors a potential distinct cell type of 'origin' could be assigned. We found that both human luminal progenitor and mouse fetal mammary stem cell features predicted pCR sensitivity across all breast cancer patients even after controlling for intrinsic subtype, proliferation, and clinical variables. This work identifies new clinically relevant gene signatures and highlights the value of a developmental biology perspective for uncovering relationships between tumor subtypes and their potential normal cellular counterparts.

Keywords Breast cancer · Comparative genomics · Genetically engineered mouse models · Genomic signatures · Neoadjuvant chemotherapy · Normal mammary tissue

Electronic supplementary material The online version of this article (doi:10.1007/s10549-014-3262-6) contains supplementary material, which is available to authorized users.

A. D. Pfefferle · C. M. Perou
Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: adamp@email.unc.edu

A. D. Pfefferle · C. M. Perou (✉)
Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: cperou@med.unc.edu

B. T. Spike · G. M. Wahl
Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92130, USA
e-mail: bspike@salk.edu

G. M. Wahl
e-mail: wahl@salk.edu

C. M. Perou
Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Introduction

The mammalian breast is a dynamic organ, with major morphological changes occurring during organogenesis, puberty, pregnancy, lactation, and involution [1]. Underlying these mammary gland changes is a complex cell hierarchy that supports these processes [2–4]. The simplest model places the multipotent mammary stem cell (MaSC) at the base of this

hierarchy, having extensive, self-regenerative potential [5]. During mammary development, the MaSC has been proposed to divide asymmetrically to produce basal/myoepithelial cells as well as luminal progenitors (LumProg), which have more restricted proliferative and differentiation capabilities [5]. LumProg cells are capable of further differentiation into mature luminal (MatureLum) cells, such as estrogen receptor (ER)-positive ductal epithelium, which have an even more limited proliferative potential and some of which are terminally differentiated [5].

Breast tumors may originate from several, if not all, of the cell types within this complex mammary hierarchy. These various cellular foundations for tumor initiation may help explain the heterogeneous nature of human breast tumors [6], which consist of multiple histological and genomic subtypes; these genomic groups, which are defined by their gene expression profiles, have become known as the intrinsic subtypes of breast cancer and are referred to as basal-like, claudin-low, HER2-enriched, luminal A, and luminal B [7–10]. A simple etiological explanation for these different subtypes involves a one-to-one relationship between each intrinsic subtype and a distinct cell type of origin that largely maintains its phenotypic identity after oncogenic transformation; however, both normal and neoplastic non-stem cells can acquire stem-like properties, suggesting that the normal cell hierarchy model could also include an element of reversibility [11]. This also raises the possibility that molecular features defining tumor subtypes, may be acquired during tumorigenesis [12].

Genetically engineered mouse models (GEMMs) of breast carcinoma develop heterogeneous tumors [13, 14], but the extent to which they represent human disease is an area of active investigation. We previously showed that murine mammary tumors comprise at least 17 distinct intrinsic subtypes/classes, with eight classes being identified as strong human subtype counterparts by gene expression similarity [14]. As with human breast cancer, the degree to which murine models reflect normal mammary epithelial subpopulations requires further analysis. Characterization of the cellular features of these murine classes is also needed to better determine their preclinical utility, to shed light on trans-species associations [14], and to help interpret preclinical study observations [15–18].

Several studies have independently profiled fluorescence-activated cell sorted (FACS) purified normal mammary cell types from both human [19–21] and murine [22, 23] mammary tissues. Here, we use a meta-analysis approach to compare the transcriptomic profiles from FACS-enriched mammary cell populations with each other and with primary tumors. These data not only identify a number of clinically relevant biomarkers that may be useful for predicting chemotherapy benefit, but also suggest a cell type of origin for many tumor subtypes.

Methods

Detailed methods can be found in Supplemental File 1.

Mammary cell subpopulation gene signatures

Gene expression measurements from FACS-enriched mammary subpopulations were obtained from three human and two murine published studies: GSE16997 [19], GSE19446 [22], GSE27027 [23], GSE35399 [20], and GSE50470 [21]. Using a meta-analysis approach, a consensus 'enriched' gene signature was produced for each mammary subpopulation. 'Enriched' signatures comprised genes that were identified as being uniquely and highly expressed (false discovery rate (FDR) < 5 %) within a given subpopulation as determined using a two-class (subpopulation X versus all others) significance analysis of microarrays (SAM) analysis [9, 24]. Each 'enriched' signature was further refined by supervised clustering using the human UNC308 breast tumor dataset [9] to identify subpopulation 'features', which were defined as having at least ten genes with a Pearson correlation greater than 0.5 across all tumors [15, 25]. Expression scores for gene signatures were determined by calculating the mean expression of the signature within each tumor; all gene signature lists are provided in Supplemental Table 1.

Mammary cell subpopulation centroids

Mammary cell subpopulation centroids were created using the union of the 'enriched' epithelial gene signatures. Distance weighted discrimination (DWD) single sample predictor [26] was used to calculate the shortest Euclidean distance between each tumor and each epithelial cell-enriched centroid. Samples with a positive silhouette width were considered to have a strong association with a given subpopulation [27].

Chemotherapy response

A combined breast cancer gene expression dataset of patients treated with neoadjuvant anthracycline and taxane chemotherapy regimens was created from three public datasets: GSE25066 [28], GSE32646 [29], and GSE41998 [30]. Univariate (UVA) and multivariate (MVA) logistic regression analyses were used to determine if gene signatures derived from normal cell populations were capable of predicting pathological complete response (pCR).

Results

Comparison of human mammary subpopulation transcriptomic datasets

Several groups have independently obtained transcriptomic profiles of normal human breast cells and compared the genomic biology of these different cell types with human tumors [19–21]. In these studies, normal mammary tissues obtained from female donors were FAC sorted using cell surface markers to enrich for specific mammary subpopulations before microarray analysis (Table 1; Fig. 1). While these initial studies were important, the datasets themselves were relatively small ($n = 12$ for Lim et al. [19], $n = 72$ for Shehata et al. [20], $n = 18$ for Prat et al. [21]), and few if any comparisons across studies were performed. Importantly, FACS-based cell fractionation can only enrich for specific subpopulations. Therefore, transcriptomic profiles reflect features of other contaminating cell types to varying degrees. As such, study-specific biases may be present in any single dataset; therefore, we used consensus information from all three FACS-enriched human transcriptomic datasets to reduce technical and study-specific biases.

Following DWD normalization [26], an unsupervised cluster of the most variably expressed genes was performed using Gene Cluster v3.0 by selecting all genes with an absolute $\log_2$ expression value greater than three in at least four samples (212 genes) (Fig. 2a). In general, the four major array dendrogram nodes correspond to the four FACS-enriched mammary subpopulations, indicating that the most highly and variably expressed genes are similarly expressed across the different studies. Even when using all genes in the dataset, there is a high Pearson correlation within a given subpopulation across studies and low correlations to other subpopulations (Fig. 2b).

On a per-sample basis, the first principle component separated the stroma and adult mammary stem cell (aMaSC) samples from the LumProg and MatureLum samples (Fig. 2c). The second principle component separated the stroma and aMaSC samples into distinct groups, while the third principle component separated the LumProg and MatureLum samples into distinct groups. The aMaSC subpopulation displayed the highest level of variation, which is likely attributable to varying degrees of contamination by other cell types.

Human mammary cell subpopulation enriched gene signatures

As shown in Fig. 2, there is a natural degree of variation between samples of a given subpopulation. We therefore developed gene signatures for each human mammary subpopulation by integrating consensus information across all three datasets (Table 1) to identify the highest confidence subpopulation-specific genes. First, genes highly expressed (FDR < 5 %) within each mammary subpopulation were found using a two-class (subpopulation X versus all others) SAM analysis [24] within each dataset [19–21]. Second, the overlap of genes highly expressed within a particular subpopulation across studies was determined. Lastly, as it is possible in the above analysis to have the same gene in the signature of more than one subpopulation, genes that were identified to be significantly associated with more than one subpopulation were also removed. This resulted in a single, consensus *Homo sapiens*-enriched (HsEnriched) signature per subpopulation (Fig. 3a). The average Euclidean distance was

Table 1 Human FACS-enriched normal mammary cell subpopulation studies

Enriched population	FACS markers	Species	Source	Abbreviation	Reference
Stroma	CD49fneg, EpCAMneg	Human	Adult	aStr-Lim09	Lim et al. [19]
	CD49fneg, EpCAMneg	Human	Adult	aStr-Shehata	Shehata et al. [20]
	CD49fneg, EpCAMneg	Human	Adult	aStr-Prat	Prat et al. [21]
Stem cell	CD49fpos, EpCAMneg	Human	Adult	aMaSC-Lim09	Lim et al. [19]
	CD49fpos, EpCAMneg	Human	Adult	aMaSC-Shehata	Shehata et al. [20]
	CD49fpos, EpCAMneg	Human	Adult	aMaSC-Prat	Prat et al. [21]
Luminal progenitor	CD49fpos, EpCAMpos	Human	Adult	LumProg-Lim09	Lim et al. [19]
	CD49fpos, EpCAMpos	Human	Adult	LumProg-Shehata	Shehata et al. [20]
	CD49fpos, EpCAMpos	Human	Adult	LumProg-Prat	Prat et al. [21]
Mature luminal	CD49fneg, EpCAMpos	Human	Adult	MatureLum-Lim09	Lim et al. [19]
	CD49fneg, EpCAMpos	Human	Adult	MatureLum-Shehata	Shehata et al. [20]
	CD49fneg, EpCAMpos	Human	Adult	MatureLum-Prat	Prat et al. [21]

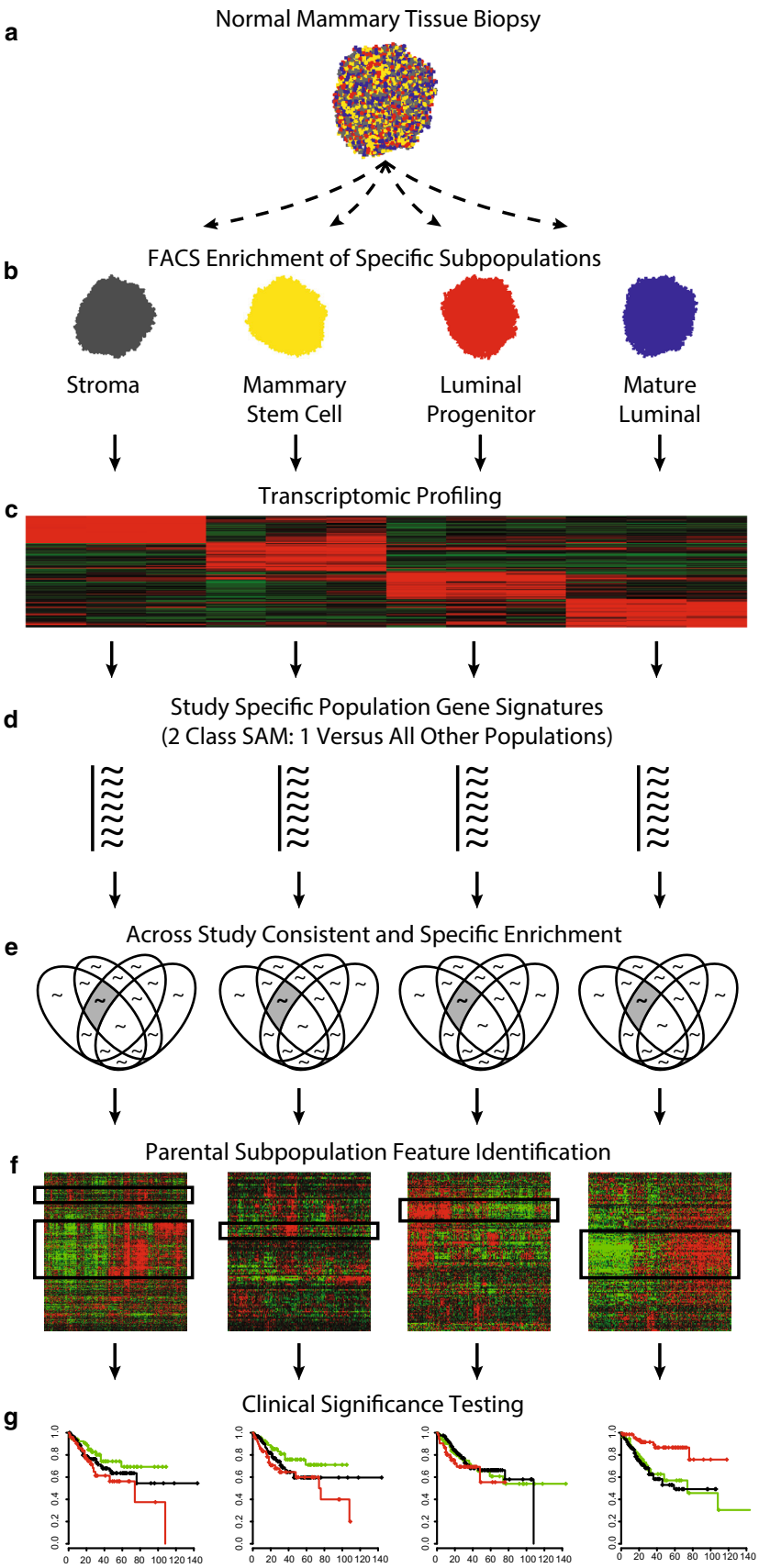


Fig. 1 Flowchart of analysis. Normal mammary tissue biopsies were taken from female patients (**a**) and FACS-enriched into distinct mammary cell subpopulations (**b**). Transcriptome profiling was performed on each subpopulation using gene expression microarrays by three different studies (**c**). Within each study, genes highly expressed within each subpopulation were determined using a two-class SAM (**d**). Genes commonly and specifically enriched within each subpopulation across studies were determined to identify 'enriched' gene signatures (**e**). Each 'enriched' signature was refined by supervised hierarchical clustering to identify gene 'features' highly correlated across a diverse set of human breast tumors (**f**). These gene signatures were then used for clinical testing (**g**)

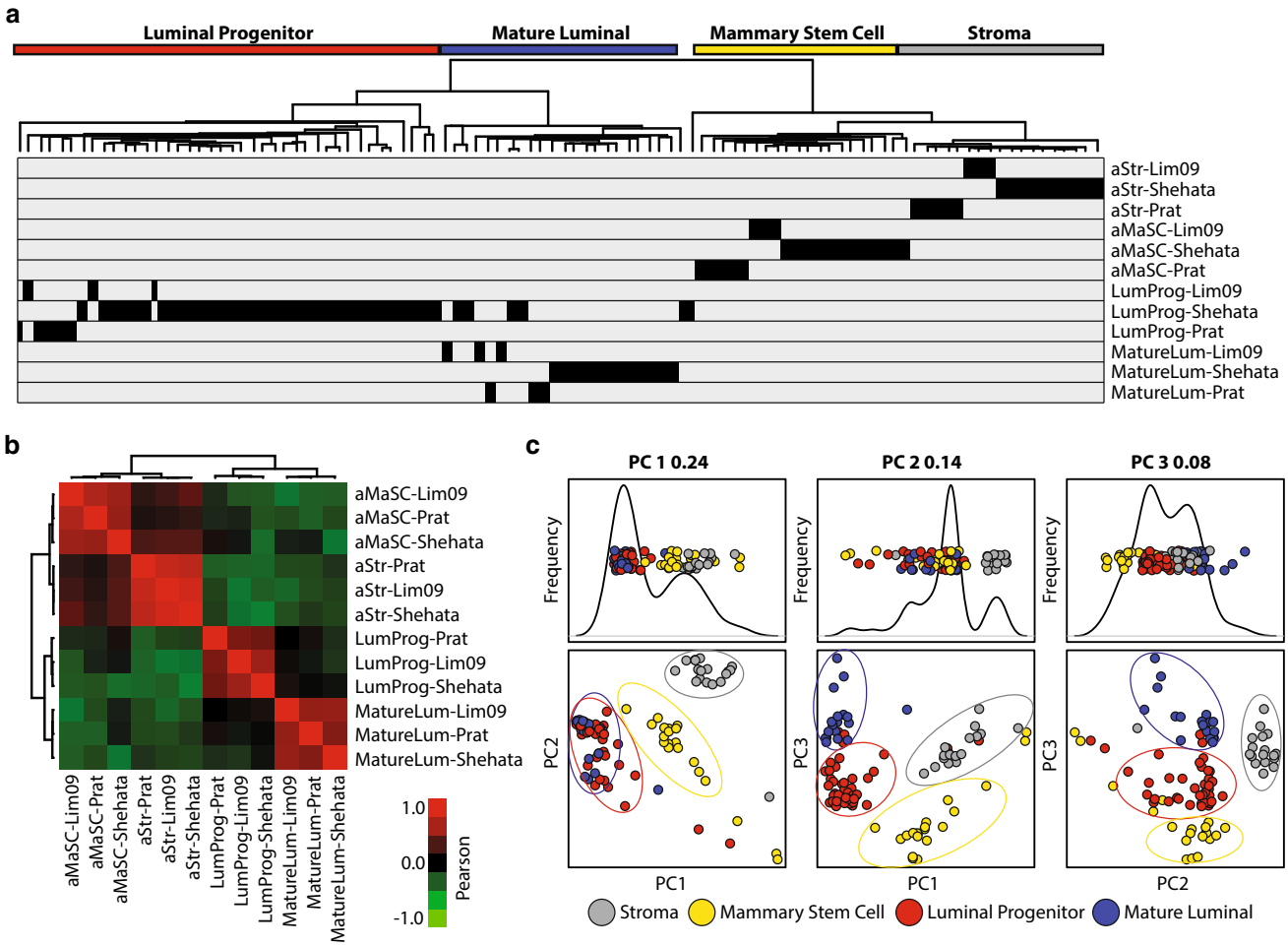


Fig. 2 Comparison of mammary subpopulations across studies. **a** Unsupervised hierarchical clustering was performed with the normal human mammary subpopulation dataset using any gene that had a $\log_2$ absolute expression value greater than three in at least four samples. **b** Pearson correlations were determined between the average expressions of each study's subpopulations using all genes. **c** The first three principle components were determined across the human mammary subpopulation dataset

determined using a 10-fold cross validation for each normal mammary subpopulation sample to centroids created using either the HsEnriched-derived gene signatures or to centroids created using the gene signatures derived separately from each human study (Supplemental Fig. 1). The HsEnriched centroids had a significantly reduced Euclidean distance ($\sim$70 %) to each mammary subpopulation (t test $p < 0.0001$), indicating greater specificity for the consensus HsEnriched signatures when compared with any individual dataset's subpopulation signature.

We next evaluated the utility of these signatures for distinguishing human tumor subtypes. Figure 3b displays the standardized average expression of each HsEnriched signature across the human intrinsic breast tumor subtypes [7, 9] using over 3,000 tumors [9, 31, 32]. The aStr-HsEnriched signature was highest in claudin-low and normal-like tumors. Interestingly, claudin-low tumors also highly express the aMaSC-HsEnriched signature. High expression of the aMaSC-HsEnriched signature in claudin-low tumors is unlikely an artifact of stromal cells in these tumors since the Pearson correlation between the aStr-HsEnriched and aMaSC-HsEnriched signatures was −0.19 across the normal human mammary samples. The LumProg and MatureLum-HsEnriched signatures were most highly expressed in basal-like and luminal subtype tumors, respectively (Fig. 3b).

We noted a considerable degree of signature variation within a subtype, indicating that it is not necessarily the case that all tumors of a given subtype share features with the same normal cell type. A nearest centroid predictor with a 10-fold cross validation error rate of 4.8 % was created to individually determine which normal mammary epithelial subpopulation is most similar to each tumor. Samples with positive silhouette widths [27] were considered to have a strong association with their particular subpopulation, with all other tumors being categorized as

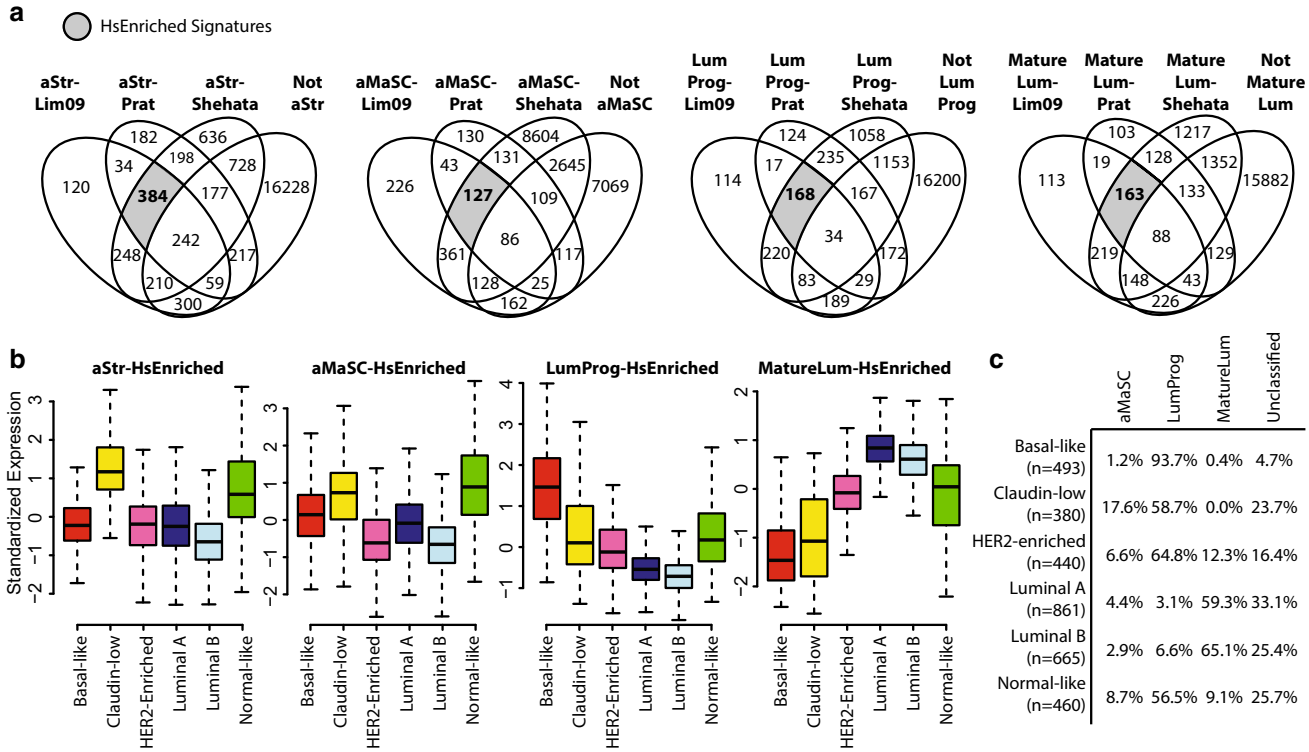


Fig. 3 *Homo sapiens*-enriched gene signatures. **a** HsEnriched gene signatures were identified for each mammary subpopulation. First, the overlap of genes highly expressed within each subpopulation across studies was determined. This overlapping gene set was further filtered to remove genes also identified as enriched in another subpopulation to limit the signature to genes specific to an individual subpopulation. The remaining genes comprised the HsEnriched gene signature for that subpopulation, as indicated by the *shaded box*. **b** The standardized average expression of the four HsEnriched gene signatures was calculated across three human datasets and displayed by intrinsic tumor subtype. **c** A nearest centroid predictor using the HsEnriched gene signatures was used to determine which epithelial features each tumor most represented. To reduce spurious findings, any tumor with a negative silhouette width was considered to have a weak association and was labeled as 'unclassified'

'unclassified' [33] (Fig. 3c). Specifically, 94 % of basal-like tumors had LumProg expression profiles. The claudin-low subtype had the highest percentage of tumors classified as aMaSC (18 %), although most claudin-low tumors were classified as having LumProg features (59 %). The HER2-enriched subtype was predominantly classified as having LumProg expression features. The luminal A and B subtypes were most similar to the MatureLum subpopulation.

Murine mammary cell subpopulation enriched gene signatures

Several groups have also profiled normal murine mammary cell subpopulation expression features using FACS [22, 23] (Table 2). In addition to highlighting conserved expression features across species [22], murine studies are uniquely positioned to enable comparisons with developmental states not easily accessed in humans, including early fetal development [23]. We were particularly interested in fetal mammary stem cells (fMaSC) [23], which is a distinct cell population not captured in any human study performed thus far (Table 3). Using the same approach that we used to derive the HsEnriched signatures, we created *Mus musculus*-enriched (MmEnriched) signatures for each murine mammary subpopulation (Fig. 4a) [22, 23].

We calculated the standardized average expression of each MmEnriched signature across the murine intrinsic subtypes/classes (Fig. 4b) [14]. As in human tumors, the Str-MmEnriched signature was most highly expressed in Normal-like^{Ex} and Claudin-low^{Ex}; this common feature was anticipated given the high similarity of these two classes to their human subtype counterparts and their known enrichment for stroma-associated genes [14, 23]. The aMaSC-MmEnriched signature was most highly expressed in Class14^{Ex} and to a slightly lesser extent in Wnt1-Late^{Ex}, Wnt1-Early^{Ex}, p53null-Basal^{Ex}, and Squamous-like^{Ex}. The fMaSC-MmEnriched signature was most highly expressed in WapINT3^{Ex}, which is consistent with the finding that *Int3* (*Notch4*) inhibits mammary cell differentiation [34, 35]. The LumProg-MmEnriched signature was highest in PyMT^{Ex} and Neu^{Ex}. This finding was unexpected given that these two mouse classes have been shown to resemble

Table 2 Murine FACS-enriched normal mammary cell subpopulation studies

Enriched population	FACS markers	Species	Source	Abbreviation	Reference
Stroma	Cd24neg/low/med	Mouse	Fetal	fStr-Spike	Spike et al. [23]
	Cd29neg, Cd24neg	Mouse	Adult	aStr-Lim10	Lim et al. [22]
Stem cell	Cd49fhi, Cd24hi	Mouse	Fetal	fMaSC-Spike	Spike et al. [23]
	Cd49fhi, Cd24med	Mouse	Adult	aMaSC-Spike	Spike et al. [23]
	Cd29pos, Cd24pos, Cd61pos	Mouse	Adult	aMaSC-Lim10	Lim et al. [22]
Luminal progenitor	Cd29neg, Cd24pos, Cd61pos	Mouse	Adult	LumProg-Lim10	Lim et al. [22]
Mature luminal	Cd29neg, Cd24pos, Cd61neg	Mouse	Adult	MatureLum-Lim10	Lim et al. [22]

Table 3 Gene set analysis of human and murine cell subpopulations

Murine subpopulation	Human subpopulation			
	Str	aMaSC	LumProg	MatureLum
Str	**0.044**	–	–	–
fMaSC	–	–	0.4395	0.4395
aMaSC	–	**0.044**	–	–
LumProg	–	–	**0.042**	0.386
MatureLum	–	0.464	0.306	**0.004**

A comparative analysis of each human subpopulation versus each murine subpopulation was performed using GSA. The FDR is displayed for all comparisons with a positive association. Statistically significant associations (FDR < 0.05) are bolded

luminal human tumors [13, 14]. Lastly, the MatureLum-MmEnriched signature was most highly expressed in Stat1^{Ex} and Class14^{Ex}. Both the *Stat1*$^{-/-}$ and *Pik3ca*-H1047R mouse models, which define these two classes respectively, are often ER positive [36, 37], and these data suggest that they have MatureLum features. Class14^{Ex} also exhibited significant expression of the aMaSC-MmEnriched signature, indicating that these tumors contain a mixture or share features of multiple cell types.

Consistent with Fig. 4b, 91 % of WapINT3^{Ex} tumors were classified as having fMaSC features in a nearest centroid predictor analysis. Mouse luminal classes of breast carcinoma (Erbb2-like^{Ex}, Myc^{Ex}, PyMT^{Ex}, and Neu^{Ex}) were most similar to LumProg cells, which again were unexpected but consistent with previous findings [22, 38]. Wnt1-Early^{Ex}, p53null-Basal^{Ex}, and Squamous-like^{Ex} tumors had primarily aMaSC features. Interestingly, Claudin-low^{Ex} and to a lesser extent C3-Tag^{Ex} tumors also had aMaSC features. All Stat1^{Ex} tumors had MatureLum features, consistent with being ER positive [36].

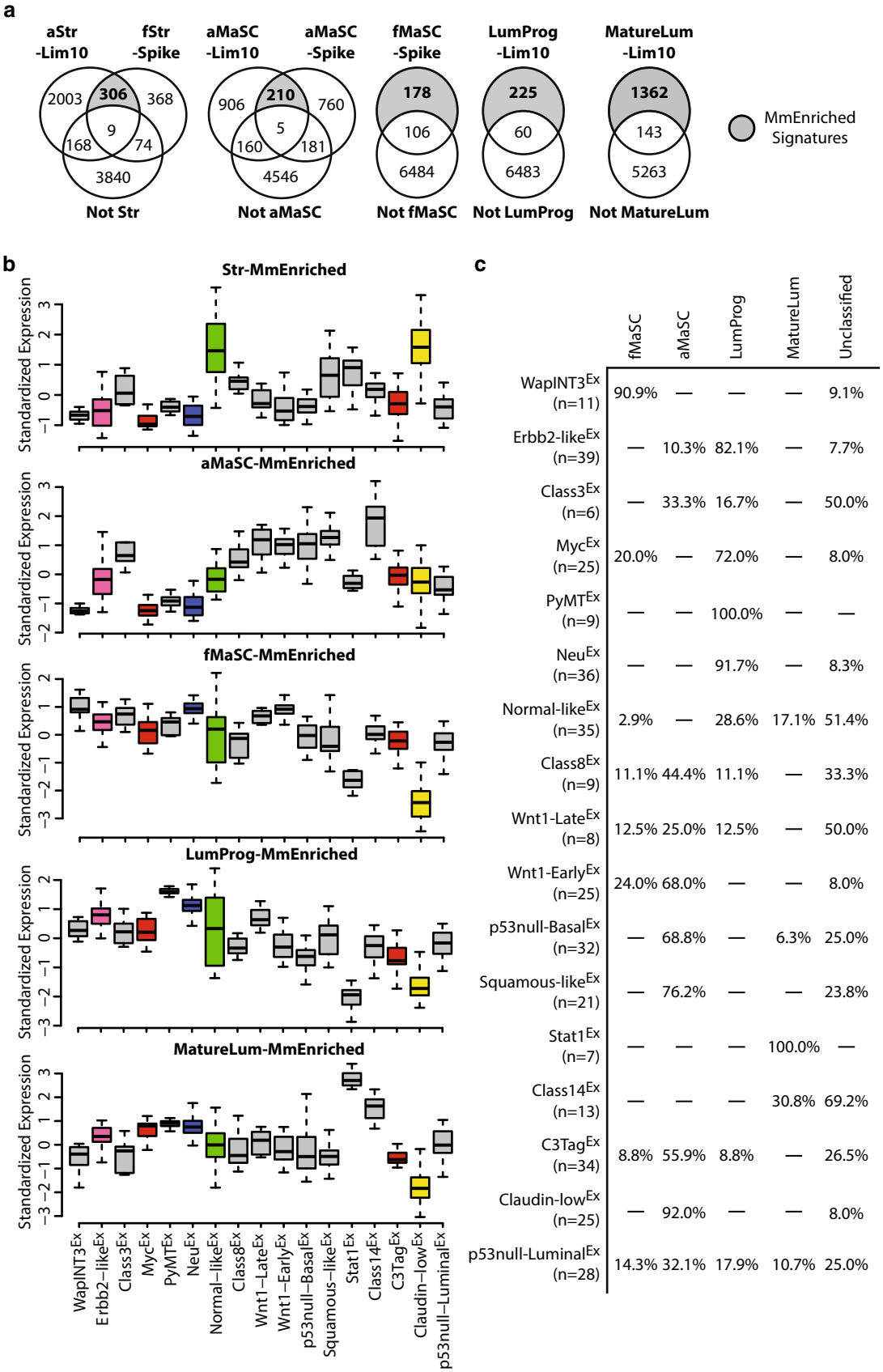
LumProg and fMaSC features predict neoadjuvant chemotherapy response

Breast tumors respond heterogeneously to neoadjuvant chemotherapy treatment [15]. We hypothesized that cellular features of normal mammary subpopulations may identify tumors most likely to respond to neoadjuvant chemotherapy. To test this, we compiled a dataset of 702 neoadjuvant anthracycline and taxane chemotherapy-treated patients (Supplemental Table 2).

Although genes within each 'enriched signature' are highly correlated within their respective normal cell subpopulation, it does not necessarily follow that all genes within a given normal cell signature would be as coordinately regulated in tumors. Therefore, we subdivided each signature into smaller features (feature1, feature2, etc.) that are coordinately expressed in tumors, reasoning that such refined 'features' may be more clinically robust. All 'enriched' and refined 'features' were tested for their ability to predict pCR to neoadjuvant chemotherapy in a UVA (Supplemental Table 3). UVA significant signatures ($p < 0.05$) were then considered in a MVA with age, ER status, PR status, HER2 status, tumor stage, PAM50 subtype [39], and PAM50 proliferation score [39] to determine if any mammary subpopulation 'features' added novel information for predicting pCR (Supplemental Table 4).

Six normal mammary gene signatures were UVA and MVA significant (Supplemental Tables 3 and 4), with the 95 % UVA odds ratio of these six signatures and all other 'enriched signatures' displayed in Fig. 5a. Interestingly, the LumProg-HsEnriched and LumProg-HsEnriched-feature1 signatures, both of which were highly correlated (Fig. 5b), were significant in the UVA and MVA analyses, indicating that tumors with LumProg features are more likely to respond to neoadjuvant treatment. Importantly, this response was independent of proliferation, as highlighted by their low correlation to the PAM50-Proliferation gene signature (Fig. 5b).

Interestingly, the fMaSC-MmEnriched signature refined into two distinctly opposite, highly significant signatures in both the UVA and MVA (Supplemental Table 3, 4; Fig. 5b, c). While the fMaSC-MmEnriched signature was highest in basal-like tumors, the refined signatures varied,


a
aStr -Lim10
fStr -Spike
aMaSC -Lim10
aMaSC -Spike
fMaSC -Spike
LumProg -Lim10
MatureLum -Lim10
2003
306
368
9
168
74
3840
Not Str
906
210
760
5
160
181
4546
Not aMaSC
178
106
6484
Not fMaSC
225
60
6483
Not LumProg
1362
143
5263
Not MatureLum
MmEnriched Signatures
b
Str-MmEnriched
aMaSC-MmEnriched
fMaSC-MmEnriched
LumProg-MmEnriched
MatureLum-MmEnriched
Standardized Expression
WapINT3Ex
Erbb2−likeEx
Class3Ex
MycEx
PyMTEx
NeuEx
Normal−likeEx
Class8Ex
Wnt1−LateEx
Wnt1−EarlyEx
p53null−BasalEx
Squamous−likeEx
Stat1Ex
Class14Ex
C3TagEx
Claudin−lowEx
p53null−LuminalEx
c
fMaSC
aMaSC
LumProg
MatureLum
Unclassified
WapINT3Ex (n=11) 90.9% — — — 9.1%
Erbb2-likeEx (n=39) — 10.3% 82.1% — 7.7%
Class3Ex (n=6) — 33.3% 16.7% — 50.0%
MycEx (n=25) 20.0% — 72.0% — 8.0%
PyMTEx (n=9) — — 100.0% — —
NeuEx (n=36) — — 91.7% — 8.3%
Normal-likeEx (n=35) 2.9% — 28.6% 17.1% 51.4%
Class8Ex (n=9) 11.1% 44.4% 11.1% — 33.3%
Wnt1-LateEx (n=8) 12.5% 25.0% 12.5% — 50.0%
Wnt1-EarlyEx (n=25) 24.0% 68.0% — — 8.0%
p53null-BasalEx (n=32) — 68.8% — 6.3% 25.0%
Squamous-likeEx (n=21) — 76.2% — — 23.8%
Stat1Ex (n=7) — — — 100.0% —
Class14Ex (n=13) — — — 30.8% 69.2%
C3TagEx (n=34) 8.8% 55.9% 8.8% — 26.5%
Claudin-lowEx (n=25) — 92.0% — — 8.0%
p53null-LuminalEx (n=28) 14.3% 32.1% 17.9% 10.7% 25.0%

◀**Fig. 4** *Mus musculus*-enriched gene signatures. **a** MmEnriched gene signatures were identified for each mammary subpopulation. First, the overlap of genes highly expressed within each subpopulation across studies was determined. This overlapping gene set was further filtered to remove genes also identified as enriched in another subpopulation to limit the signature to genes specific to an individual subpopulation. The remaining genes comprised the MmEnriched gene signature for that subpopulation, as indicated by the *shaded box*. **b** The standardized average expression of the five MmEnriched gene signatures was calculated across a murine dataset and displayed by intrinsic tumor class. **c** A nearest centroid predictor using the MmEnriched gene signatures was used to determine which epithelial features each tumor most represented. To reduce spurious findings, any tumor with a negative silhouette width was considered to have a weak association and was labeled as 'unclassified'

with fMaSC-MmEnriched-feature1 (Fig. 5d) being highest in basal-like tumors and fMaSC-MmEnriched-feature2 (Fig. 5e) expressed in luminal tumors. Tumors with fMaSC-MmEnriched-feature1 expression were more likely to respond to neoadjuvant chemotherapy, while those tumors with fMaSC-MmEnriched-feature2 were more resistant. The fMaSC-MmEnriched-feature1 signature was very highly correlated with the LumProg-HsEnriched signatures (Fig. 5b), sharing four genes in common (Fig. 5d). These results support the hypothesis that subsets of genes within the larger 'enriched signature' are likely regulated by different biological mechanisms.

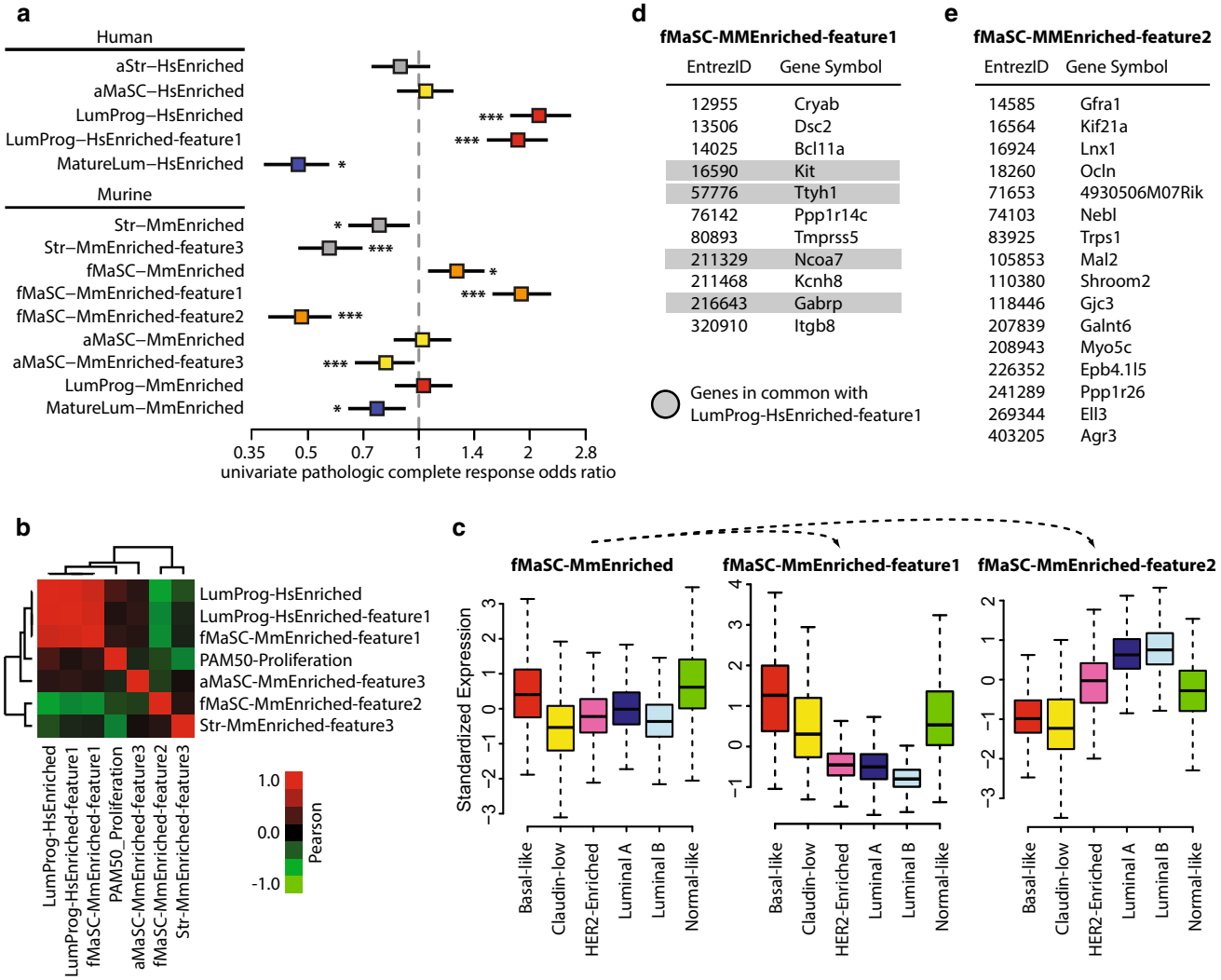


Fig. 5 fMaSC-enriched gene signatures. **a** The univariate logistic regression odds ratio predicting pathologic complete response to neoadjuvant anthracycline and taxane chemotherapy was determined using a 702 patient dataset, with the 95 % confidence interval shown as a forest plot. A single '*' indicates that the signature was univariate significant, while '***' indicates that the signature was both univariate and multivariate significant ($p < 0.05$). **b** Pearson correlations of multivariate significant gene signatures and proliferation were determined. **c** The standardized average expression of the fMaSC-MmEnriched signature and its two refined signatures were calculated across three human datasets and displayed by intrinsic tumor subtype. **d** Genes in the fMaSC-MmEnriched-refined1 signature. **e** Genes in the fMaSC-MmEnriched-refined2 signature

Discussion

Normal mammary gland physiology is supported by an underlying, complex cell hierarchy [2–5]. The simplest model treats differentiation from mammary stem cells to progenitor cells to mature cells as unidirectional, but recent observations indicate that bidirectional processes are also possible for normal and neoplastic cells [11]. This differentiation plasticity may allow tumors to acquire cell features foreign to the initial cell-of-origin or to lose native features through the accumulation of specific genetic aberrations [40].

Regardless of how different cellular traits are acquired, it is critical to identify the 'current' normal cellular features within a tumor, and therefore, we first analyzed the expression profiles of normal human and mouse mammary epithelial cell subpopulations [19–23]. We chose to use nomenclature that maintains continuity with the literature. However, these terms should be considered provisional as the complete biological profiles of these FACS fractions are investigated [4]. Recent work by Prater et al. [41] found that mouse 'LumProg' cells ($CD49f^{+}$, $EpCAM^{+}$) have complete mammary gland repopulating potential, indicating that 'LumProg' may be a misnomer. Importantly, even if our understanding and naming of these cell subpopulations change, only the retrospective interpretation of the data presented here will be affected, not the data itself.

Using a meta-analysis approach, FACS-purified mammary epithelial cell subpopulation 'enriched' gene signatures were derived and a nearest centroid predictor was developed to identify which normal mammary subpopulation each human and mouse tumor most represented using over three thousand human patients and 27 mouse models of mammary carcinoma [14]. While these analyses imply a cell-of-origin for a given tumor, additional experiments (e.g., lineage tracing) will be required to unequivocally determine this. Nevertheless, these associations at the very least identify which normal mammary subpopulation a given tumor most represents in its current state.

With this in mind, several associations between both the human and mouse intrinsic subtypes and specific normal cell subpopulations were observed. First, human basal-like tumors have been referred to as 'undifferentiated', which is consistent with their exhibiting LumProg [19] and fetal MaSC features [23]. Three mouse classes have been identified to be human basal-like counterparts: Myc^{Ex}, $p53null\text{-}Basal^{Ex}$, and $C3\text{-}Tag^{Ex}$ [14]. Myc^{Ex} tumors were the most similar to the LumProg cell profile. By contrast, both $p53null\text{-}Basal^{Ex}$ and $C3\text{-}Tag^{Ex}$ tumors had adult MaSC features. These results indicate that Myc^{Ex} tumors share similar cell features as their human basal-like counterpart, making it an attractive mouse model for studying basal-like tumors with aberrant Myc signaling [10, 42]. Interestingly, neither $p53null\text{-}Basal^{Ex}$ nor $C3\text{-}Tag^{Ex}$ tumors had strong LumProgs features, indicating that their association with human basal-like tumors is more likely driven by their underlying genetics [10].

Human claudin-low tumors had heterogeneous normal cell features. While most were similar to LumProg cells, the claudin-low subtype also had the largest percentage of tumors classified as adult MaSC. Given that claudin-low tumors are enriched with epithelial-to-mesenchymal transition features [9, 43, 44], our results suggest that these tumors may originate from the LumProg population prior to acquiring adult MaSC and/or mesenchymal features. Similarly, mouse $Claudin\text{-}low^{Ex}$ tumors were also strongly associated with the adult MaSC population, indicating that such tumors may be the closest analogs of the subset of human claudin-low tumors with adult MaSC features.

Human HER2-enriched tumors were the most similar to the LumProg subpopulation. This is a novel finding and may explain why both human basal-like and HER2-enriched subtype tumors show high TP53 mutation frequencies (>70 %) and widespread chromosomal instability [10]. These data could suggest that the normal LumProg cell is somehow extremely dependent on TP53 function. The murine $Erbb2\text{-}like^{Ex}$ class has been identified as a mouse counterpart for human HER2-enriched tumors [14] and was shown here to also have LumProg features.

When analyzing the human luminal A and B subtypes, a clear association with normal MatureLum cells was observed. The murine Neu^{Ex} class is a proposed counterpart for human luminal A tumors [14], yet these mouse tumors were most similar to normal mouse LumProg cells. The Myc^{Ex} class was also identified to resemble human luminal B tumors [14]. As discussed, Myc^{Ex} tumors have LumProg features; therefore, most mouse luminal A/B tumor models do not share the same normal cell features as their human tumor counterparts. These differences may reflect limitations of model system design, as tumors within these mouse classes are primarily driven by either the WAP or MMTV promoter. These differences in cell features, however, indicate that the trans-species associations observed previously [14] are possibly driven by the genetics of each mouse model. Nevertheless, broad molecular features are conserved between these human–murine counterparts [14]. Therefore, we propose that these mouse models retain significant preclinical utility provided that shared versus distinct molecular features are taken into account.

Neoadjuvant chemotherapy is a common approach for treating breast tumors, but only a relatively low percentage of patients have a pCR (~20 % overall). We tested the clinical significance of normal cellular features for predicting pCR using a combination of UVA and MVA logistic regression analyses. Human LumProg and mouse

fetal MaSC expression features were identified as predictive of pCR sensitivity across all breast cancer patients. More specifically, LumProg-HsEnriched-feature1 and fMaSC-MmEnriched-feature1 were highly expressed in basal-like tumors. This is consistent with the clinical observation that basal-like tumors have better neoadjuvant chemotherapy response rates since higher expression of these normal cell signatures was associated with a higher likelihood of pCR. Distinct from these signatures, tumors with high expression of fMaSC-MmEnriched-feature2 were more resistant to neoadjuvant chemotherapy. Not surprisingly, this signature was most highly expressed in luminal A and B tumors, consistent with the clinical observation that these subtypes have lower chemotherapy response rates. Importantly, these signatures remained significant even after controlling for intrinsic subtype, proliferation, and clinical variables in the MVA analysis; thus these normal cell signatures add information even when tumor subtype and clinical features are known. It is presently unknown whether tumors with these features arise from a LumProg or fetal MaSC cell-of-origin or acquire these features during tumorigenesis. Whether these features are acquired or inherent, the 'current' cellular traits of a tumor are likely most important as these appear to be a major determinant of chemotherapy sensitivity. The biological explanation for why LumProg and fetal MaSC expression features predict tumor responsiveness to neoadjuvant chemotherapy will need to be explored further, but it is likely linked to the common genetic features of TP53 loss [45], RB-pathway loss [46], and high proliferation status [47], as well as other inherent characteristics of these cellular states. This work highlights the efficacy of studying the normal mammary gland cell hierarchy and development to provide insights into human tumor therapy responsiveness.

Acknowledgments We would like to thank J.S. Parker, J.C. Harrell, and the Perou lab for helpful suggestions. This study was supported by funds from the following sources: NCI P50-CA58223 Breast SPORE program (CMP), RO1-CA138255 (CMP), RO1-CA148761 (CMP), Department of Defense W81XWH-12-1-0106 and W81XWH-12-1-0107 (CMP and GMW), NIEHS T32-ES007017-35S1 (ADP), and the Breast Cancer Research Foundation (CMP and GMW).

Conflict of interest C. M. P is an equity stock holder of BioClassifier LLC and University Genomics, and has filed a patent on the PAM50 subtyping assay.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Gjorevski N, Nelson CM (2011) Integrated morphodynamic signalling of the mammary gland. Nat Rev Mol Cell Biol 12:581–593
2. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C (2011) Distinct stem cells contribute to mammary gland development and maintenance. Nature 479:189–193
3. Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, Harrell JC, McNamara G, Schwede M, Culhane AC, Kindelberger D, Rodig S, Richardson A, Schnitt SJ, Tamimi RM, Ince TA (2014) Taxonomy of breast cancer based on normal cell phenotype predicts outcome. J Clin Invest 124:859–870
4. Visvader JE, Stingl J (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev 28:1143–1158
5. Visvader JE (2009) Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev 23:2563–2577
6. Visvader JE (2011) Cells of origin in cancer. Nature 469:314–322
7. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406:747–752
8. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. PNAS 98:10869–10874
9. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 12:R68
10. TCGA (2012) Comprehensive molecular portraits of human breast tumours. Nature 490:61–70
11. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, Brooks M, Reinhardt F, Su Y, Polyak K, Arendt LM, Kuperwasser C, Bierie B, Weinberg RA (2011) Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. PNAS 108:7950–7955
12. Spike BT, Wahl GM (2011) p53, stem cells, and reprogramming: tumor suppression beyond guarding the genome. Genes Cancer 2:404–419
13. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 8:R76
14. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, Rosen JM, Perou CM (2013) Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol 14:R125
15. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L, Karginova O, Jordan J, Combest A, Bridges A, Prat A, Cheang MC, Herschkowitz JI, Rosen JM, Zamboni W, Sharpless NE, Perou CM (2013) Predicting drug responsiveness in human cancers

using genetically engineered mice. Clin Cancer Res 19: 4889–4899
16. Roberts PJ, Usary JE, Darr DB, Dillon PM, Pfefferle AD, Whittle MC, Duncan JS, Johnson SM, Combest AJ, Jin J, Zamboni WC, Johnson GL, Perou CM, Sharpless NE (2012) Combined PI3 K/mTOR and MEK inhibition provides broad antitumor activity in faithful murine cancer models. Clin Cancer Res 18:5290–5303
17. Bennett CN, Tomlinson CC, Michalowski AM, Chu IM, Luger D, Mittereder LR, Aprelikova O, Shou J, Piwinica-Worms H, Caplen NJ, Hollingshead MG, Green JE (2012) Cross-species genomic and functional analyses identify a combination therapy using a CHK1 inhibitor and a ribonucleotide reductase inhibitor to treat triple-negative breast cancer. Breast Cancer Res 14:R109
18. Roberts PJ, Bisi JE, Strum JC, Combest AJ, Darr DB, Usary JE, Zamboni WC, Wong KK, Perou CM, Sharpless NE (2012) Multiple roles of cyclin-dependent kinase 4/6 inhibitors in cancer therapy. J Natl Cancer Inst 104:476–487
19. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, Thorne HJ, Fox SB, Yan M, French JD, Brown MA, Smyth GK, Visvader JE, Lindeman GJ (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med 15:907–913
20. Shehata M, Teschendorff A, Sharp G, Novcic N, Russell A, Avril S, Prater M, Eirew P, Caldas C, Watson CJ, Stingl J (2012) Phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland. Breast Cancer Res 14:R134
21. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, Perou CM (2013) Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast Cancer Res Treat 142:237–255
22. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK, Visvader JE (2010) Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. Breast Cancer Res 12:R21
23. Spike BT, Engle DD, Lin JC, Cheung SK, La J, Wahl GM (2012) A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. Cell Stem Cell 10:183–197
24. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. PNAS 98:5116–5121
25. Hoadley KA, Weigman VJ, Fan C, Sawyer LR, He X, Troester MA, Sartor CI, Rieger-House T, Bernard PS, Carey LA, Perou CM (2007) EGFR associated expression profiles vary with breast tumor subtype. BMC Genomics 8:258
26. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, Marron JS (2004) Adjustment of systematic microarray data biases. Bioinformatics 20:105–114
27. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster-analysis. J Comput Appl Math 20:53–65
28. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, Vidaurre T, Holmes F, Souchon E, Wang H, Martin M, Cotrina J, Gomez H, Hubbard R, Chacon JI, Ferrer-Lozano J, Dyer R, Buxton M, Gong Y, Wu Y, Ibrahim N, Andreopoulou E, Ueno NT, Hunt K, Yang W, Nazario A, DeMichele A, O'Shaughnessy J, Hortobagyi GN, Symmans WF (2011) A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA 305:1873–1881
29. Miyake T, Nakayama T, Naoi Y, Yamamoto N, Otani Y, Kim SJ, Shimazu K, Shimomura A, Maruyama N, Tamaki Y, Noguchi S (2012) GSTP1 expression predicts poor pathological complete response to neoadjuvant chemotherapy in ER-negative breast cancer. Cancer Sci 103:913–920
30. Horak CE, Pusztai L, Xing G, Trifan OC, Saura C, Tseng LM, Chan S, Welcher R, Liu D (2013) Biomarker analysis of neoadjuvant doxorubicin/cyclophosphamide followed by ixabepilone or paclitaxel in early-stage breast cancer. Clin Cancer Res 19:1587–1595
31. Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, Anders C, Ewend M, Perou CM (2012) Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse. Breast Cancer Res Treat 132:523–535
32. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Group M, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346–352
33. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN, Cancer Genome Atlas Research N (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17:98–110
34. Gallahan D, Jhappan C, Robinson G, Hennighausen L, Sharp R, Kordon E, Callahan R, Merlino G, Smith GH (1996) Expression of a truncated Int3 gene in developing secretory mammary epithelium specifically retards lobular differentiation resulting in tumorigenesis. Cancer Res 56:1775–1785
35. Smith GH, Gallahan D, Diella F, Jhappan C, Merlino G, Callahan R (1995) Constitutive expression of a truncated INT3 gene in mouse mammary epithelium impairs differentiation and functional development. Cell Growth Differ 6:563–577
36. Chan SR, Vermi W, Luo J, Lucini L, Rickert C, Fowler AM, Lonardi S, Arthur C, Young LJ, Levy DE, Welch MJ, Cardiff RD, Schreiber RD (2012) STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas. Breast Cancer Res 14:R16
37. Adams JR, Xu K, Liu JC, Agamez NM, Loch AJ, Wong RG, Wang W, Wright KL, Lane TF, Zacksenhaus E, Egan SE (2011) Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation. Cancer Res 71:2706–2717
38. Li Z, Tognon CE, Godinho FJ, Yasaitis L, Hock H, Herschkowitz JI, Lannon CL, Cho E, Kim SJ, Bronson RT, Perou CM, Sorensen PH, Orkin SH (2007) ETV6-NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex. Cancer Cell 12:542–558
39. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27:1160–1167
40. Meacham CE, Morrison SJ (2013) Tumour heterogeneity and cancer cell plasticity. Nature 501:328–337
41. Prater MD, Petit V, Russell IA, Giraddi RR, Shehata M, Menon S, Schulte R, Kalajzic I, Rath N, Olson MF, Metzger D, Faraldo MM, Deugnier MA, Glukhova MA, Sting J (2014) Mammary stem cells have myoepithelial cell properties. Nat Cell Biol 16:942–950

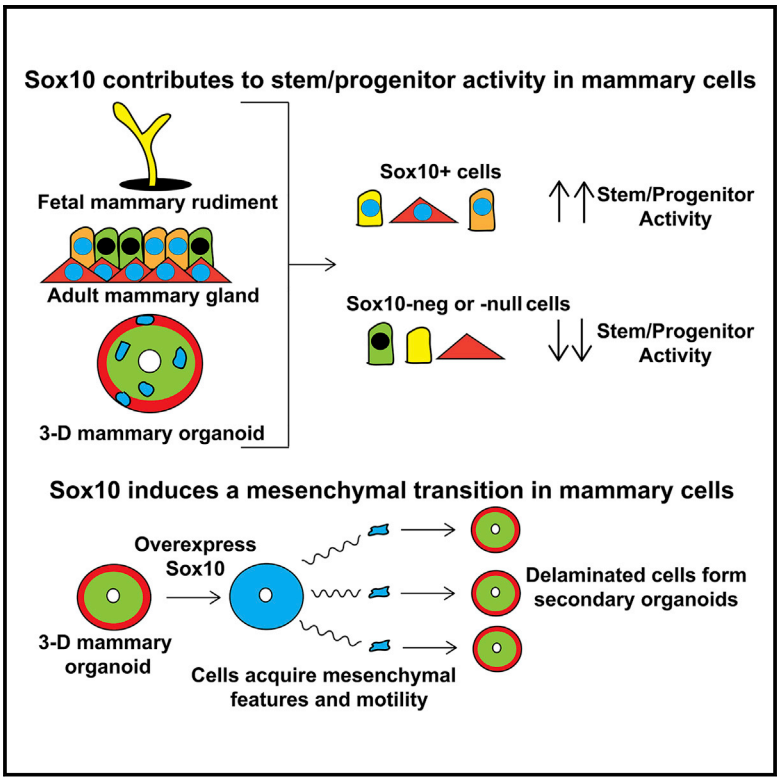
42. Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, Cole MD (2009) A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. PLoS One 4(8):e6693
43. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, Hollier BG, Ram PT, Lander ES, Rosen JM, Weinberg RA, Mani SA (2010) Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. PNAS 107:15449–15454
44. Morel AP, Hinkal GW, Thomas C, Fauvet F, Courtois-Cox S, Wierinckx A, Devouassoux-Shisheboran M, Treilleux I, Tissier A, Gras B, Pourchet J, Puisieux I, Browne GJ, Spicer DB, Lachuer J, Ansieau S, Puisieux A (2012) EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice. PLoS Genet 8:e1002723
45. Gluck S, Ross JS, Royce M, McKenna EF Jr, Perou CM, Avisar E, Wu L (2012) TP53 genomics predict higher clinical and pathologic tumor response in operable early-stage breast cancer treated with docetaxel-capecitabine ± trastuzumab. Breast Cancer Res Treat 132:781–791
46. Herschkowitz JI, He X, Fan C, Perou CM (2008) The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. Breast Cancer Res 10:R75
47. Prat A, Lluch A, Albanell J, Barry WT, Fan C, Chacon JI, Parker JS, Calvo L, Plazaola A, Arcusa A, Segui-Palmer MA, Burgues O, Ribelles N, Rodriguez-Lescure A, Guerrero A, Ruiz-Borrego M, Munarriz B, Lopez JA, Adamo B, Cheang MC, Li Y, Hu Z, Gulley ML, Vidal MJ, Pitcher BN, Liu MC, Citron ML, Ellis MJ, Mardis E, Vickery T, Hudis CA, Winer EP, Carey LA, Caballero R, Carrasco E, Martin M, Perou CM, Alba E (2014) Predicting response and survival in chemotherapy-treated triple-negative breast cancer. Br J Cancer 111:1532–1541

Article

Cell Reports

Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells

Graphical Abstract



Authors

Christopher Dravis, Benjamin T. Spike, J. Chuck Harrell, ..., E. Michelle Southard- Smith, Charles M. Perou, Geoffrey M. Wahl

Correspondence

wahl@salk.edu

In Brief

Dravis et al. report that Sox10 is specifically expressed in either stem-cell- or mesenchymal-like human breast cancers. The authors then demonstrate Sox10 can separately promote both stem/progenitor and mesenchymal-like states in mammary cells. They further identify FGF signaling as a key extrinsic mediator of Sox10 expression and function.

Highlights

- Sox10 labels and contributes to stem/progenitor activity in mammary cells
- Sox10 drives EMT and delamination of clonogenic mammary cells at high levels
- Sox10 expression and functional output are regulated by FGF signaling
- Stem- and EMT-like breast cancers show the highest expression levels of Sox10

Accession Numbers

GSE71300

Dravis et al., 2015, Cell Reports *12*, 2035–2048
September 29, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.08.040

Cell Reports
Article

OPEN ACCESS
CellPress

Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells

Christopher Dravis,[1] Benjamin T. Spike,[1] J. Chuck Harrell,[2] Claire Johns,[1] Christy L. Trejo,[1] E. Michelle Southard-Smith,[3] Charles M. Perou,[4] and Geoffrey M. Wahl[1,*]
[1]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[2]Department of Pathology, Massey Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA
[3]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[4]Departments of Genetics and Pathology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
*Correspondence: wahl@salk.edu
http://dx.doi.org/10.1016/j.celrep.2015.08.040
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

SUMMARY

To discover mechanisms that mediate plasticity in mammary cells, we characterized signaling networks that are present in the mammary stem cells responsible for fetal and adult mammary development. These analyses identified a signaling axis between FGF signaling and the transcription factor Sox10. Here, we show that Sox10 is specifically expressed in mammary cells exhibiting the highest levels of stem/progenitor activity. This includes fetal and adult mammary cells in vivo and mammary organoids in vitro. Sox10 is functionally relevant, as its deletion reduces stem/progenitor competence whereas its overexpression increases stem/progenitor activity. Intriguingly, we also show that Sox10 overexpression causes mammary cells to undergo a mesenchymal transition. Consistent with these findings, Sox10 is preferentially expressed in stem- and mesenchymal-like breast cancers. These results demonstrate a signaling mechanism through which stem and mesenchymal states are acquired in mammary cells and suggest therapeutic avenues in breast cancers for which targeted therapies are currently unavailable.

INTRODUCTION

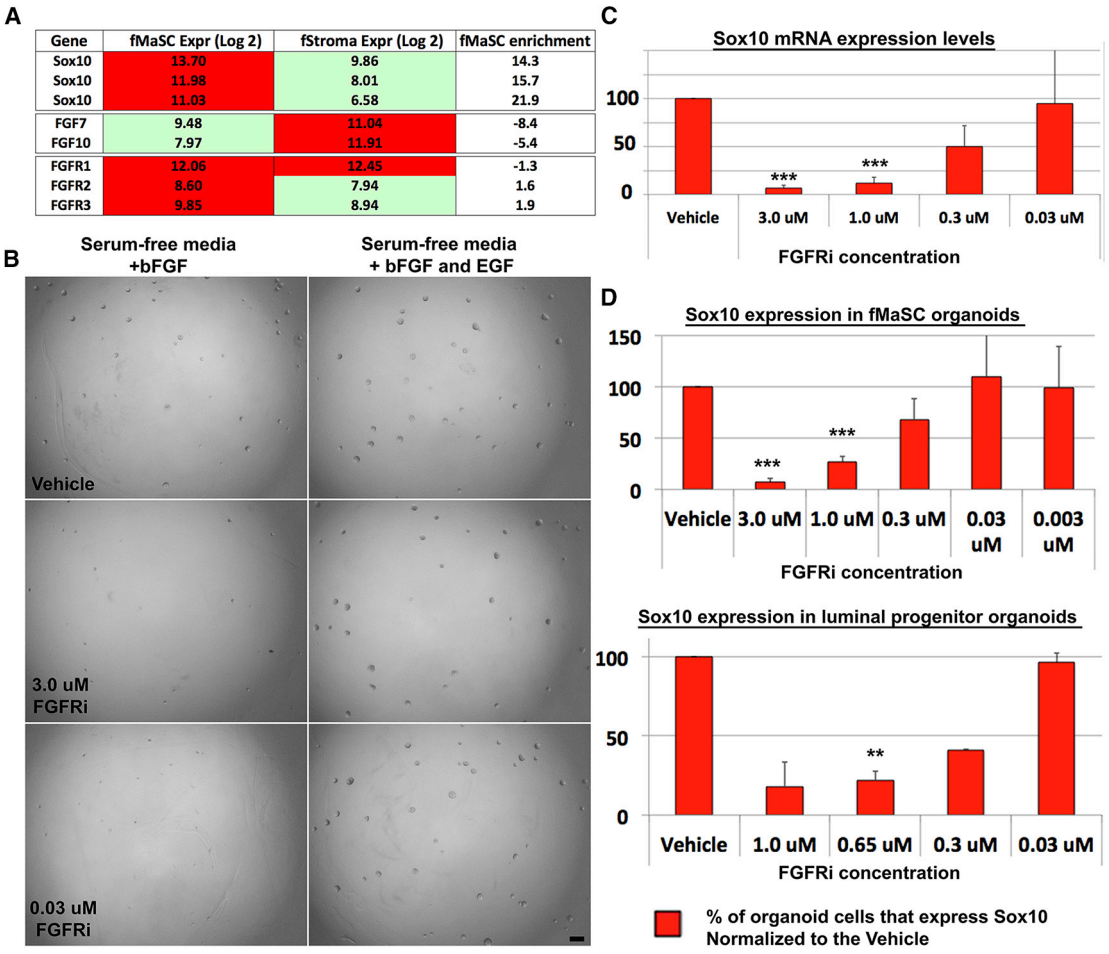
The capacity to reprogram differentiated cells in vivo and ex vivo indicates that the differentiated state is not as fixed as once thought (Takahashi and Yamanaka, 2006; Tata et al., 2013). This plasticity has important implications for cancer, where the dysregulation of stem and mesenchymal states appears to be critical in disease initiation and progression. Phenotypic lability may endow some types of cancer cells, often termed "cancer stem cells" (CSC), with a greater capacity to propagate the disease when assayed in a transplant setting (Al-Hajj et al., 2003; Bonnet and Dick, 1997). In contrast to CSCs, which typically exhibit mesenchymal characteristics, transcriptome analyses have revealed another class of tumorigenic cancer cells whose gene expression profiles resemble those of cells with known stem or progenitor cell functions. Tumors with these distinct "stem-like" cancer cells tend to appear less differentiated and behave more aggressively, while eliminating such cells can attenuate tumor progression (Chen et al., 2012; Eppert et al., 2011; Merlos-Suárez et al., 2011; Schepers et al., 2012). Stem-like cancer cells may arise either by cell of origin, in which the tumor originates in a stem/progenitor cell and retains those properties through tumorigenesis, or through reprogramming of differentiated cells into a stem-like state (Barker et al., 2009; Schwitalla et al., 2013). Because a significant fraction of triple-negative breast cancers contain stem-like cancer cells, we have focused on elucidating the molecular mechanisms that specify the mammary stem cell (MaSC) state, assuming that such knowledge will deepen our understanding of how such breast cancers initiate and progress.

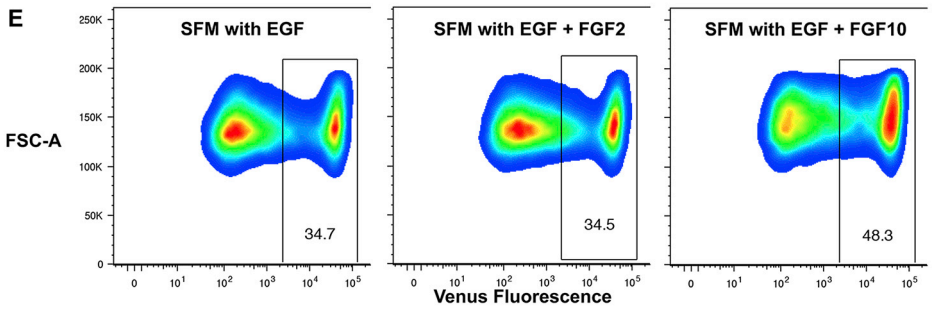
The mammary gland contains at least two populations of cells with stem or progenitor qualities (Shackleton et al., 2006; Stingl et al., 2006). Luminal progenitors comprise a heterogeneous population of cells in the luminal fraction of the gland that possess clonogenic properties in vitro (Shehata et al., 2012). This population may contain the cell of origin for stem-like basal-like breast cancers (Lim et al., 2009). Transplantation studies also demonstrate that the basal fraction of the gland contains cells capable of generating an entire mammary gland. These MaSCs are inferred to possess extensive proliferative, invasive, and multi-lineage differentiation potential, as a single MaSC can regenerate a functional gland (Shackleton et al., 2006).

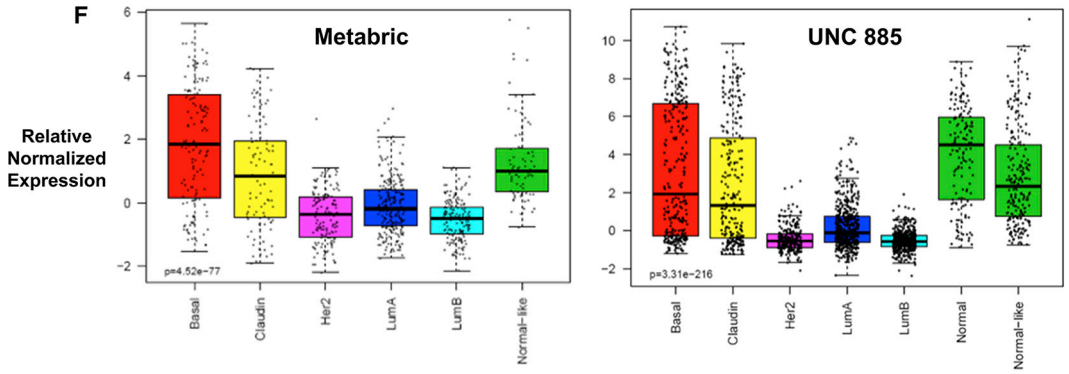
Several fundamental aspects of MaSC biology remain to be elucidated. There is no consensus on the number of MaSCs within the gland, which has hindered analyses of the origin of breast tumors (Tomasetti and Vogelstein, 2015). There are also conflicting data about the topographical location of MaSCs in the gland and the developmental timeframe during which these cells retain multi-lineage potential (Rios et al., 2014; Van Keymeulen et al., 2011). Both of these problems might be resolved by availability of markers enabling prospective MaSC identification. The mechanisms by which mammary cells enter and exit

A

Gene	fMaSC Expr (Log 2)	fStroma Expr (Log 2)	fMaSC enrichment
Sox10	13.70	9.86	14.3
Sox10	11.98	8.01	15.7
Sox10	11.03	6.58	21.9
FGF7	9.48	11.04	-8.4
FGF10	7.97	11.91	-5.4
FGFR1	12.06	12.45	-1.3
FGFR2	8.60	7.94	1.6
FGFR3	9.85	8.94	1.9







(legend on next page)

from the MaSC state also remain to be defined, and resolving this problem may present solutions to those concerning MaSC identification. One recent advance on this topic involves the demonstration that Sox9 and Slug act together to convert mammary epithelial cells into cells with MaSC-like properties (Guo et al., 2012). However, the degree to which this mechanism is utilized in the gland is not clear because the distribution and function of Sox9 or Sox9/Slug cells in unperturbed in vivo contexts remain to be defined. Moreover, mice that are deficient for *Slug* do form a complete native mammary gland, which suggests that Slug is not an essential determinant of the MaSC state (Nassour et al., 2012). Clearly, a better understanding of the transcriptional programs and extrinsic signaling mechanisms that regulate the MaSC state is required.

To investigate the biology of MaSCs and MaSC-like cells in cancer, our research has focused on the stem cells present during fetal mammary development. During mid-late embryogenesis, mammary cells are highly proliferative and invasive and likely experience conditions such as hypoxia and growth-oriented metabolism that resemble those encountered by tumor cells (Masson and Ratcliffe, 2014). Fetal MaSCs (fMaSCs) may therefore most resemble the MaSC-like cancer cells in breast tumors. Indeed, we previously showed that fMaSCs exhibit both the organoid-forming and mammary-repopulating properties found in luminal progenitors and adult MaSCs, respectively (Spike et al., 2012). Transcriptome profiling of fMaSCs and adult MaSCs revealed that the fMaSC signature gene list is uniquely enriched in basal-like breast tumors, indicating the presence of fMaSC-like cells in such tumors. This shared biology suggests that fetal mammary development and fMaSCs can be utilized to identify molecular mechanisms that govern important functions in breast cancer.

Here, we describe how analysis of fMaSCs revealed an important function for Sox10 in mammary cells. Sox family transcription factors have well-defined roles in regulating cell-fate decisions in different tissues and at different stages of development (Sarkar and Hochedlinger, 2013). Sox factors generally induce preferential differentiation down one cell lineage path over another, often by antagonizing the activity of other lineage-specifying factors. This phenomenon has best been described with Sox2 and the elucidation of roles for Sox2 in multiple different cell-fate decisions, each of which occurs in concert with other transcription factors (Sarkar and Hochedlinger, 2013). However, when Sox expression or activity is balanced or kept at lower levels in the cell by other key factors, differentiation is forestalled and stem and progenitor functions arise (Kopp et al., 2008). This is consistent with an emerging model of stem cell specification through the balance of lineage specifiers (Loh and Lim, 2011). Sox factors can thus be mediators and markers of both differentiation and stemness, depending on expression level and cellular context.

Here, we report that Sox10 plays important regulatory roles in promoting both stem- and epithelial to mesenchymal transition (EMT)-like properties in mammary stem cells. Critically, these stem and mesenchymal states are acquired independently of one another; this clear distinction prevents potential conflation of stem cell and mesenchymal properties, and demonstrates how these distinct states can be related by a single factor such as Sox10. We further present evidence that these functions may be conserved in certain types of aggressive breast cancers, and demonstrate the importance of FGF10 in a paracrine signaling mechanism that regulates Sox10.

RESULTS

Sox10 Is an fMaSC- and Tumor-Associated Transcription Factor Regulated by Fibroblast Growth Factor Signaling

To identify molecular mechanisms that specify stem/progenitor cell functions in mammary cells, we analyzed transcriptome profiles of fMaSCs and their surrounding fetal stroma (fStr) (Spike et al., 2012). We prioritized both transcription factors that are differentially expressed in the fMaSC-enriched population and inferred signaling axes between fMaSCs and fStr that could regulate their expression. These analyses identified Sox10 as one of the most prominent transcription factors associated with the fMaSC population (Figure 1A). This was of immediate interest, as Sox family transcription factors play important roles in pluripotent or tissue-specific stem cell states (Sarkar and Hochedlinger, 2013). Further, Sox10 in particular has been shown to be a critical transcription factor in reprogramming differentiated cells into multipotent stem/progenitor states (Hornig et al., 2013; Kim et al., 2014; Najm et al., 2013; Yang et al., 2013).

These analyses also revealed high relative expression of FGF7 and FGF10 in the fStr and expression of multiple fibroblast growth factor receptor (FGFR) family members in the fMaSC population (Figure 1A). Fibroblast growth factor (FGF) signaling plays a critical role in fetal mammary development, and we previously showed that fMaSCs could utilize FGF signaling to promote multipotent growth in vitro (Lu et al., 2008; Mailleux et al., 2002; Spike et al., 2012). Furthermore, FGF signaling has been shown to regulate the expression and function of different Sox family transcription factors in

Figure 1. Sox10 Is an fMaSC- and Tumor-Associated Transcription Factor Regulated by FGF Signaling

(A) Log2 microarray expression values for Sox10 and FGF signaling molecules in E18 fMaSCs and fStroma.

(B) E18 fMaSCs grown in 3D culture conditions for 5–7 days with the indicated media. Scale bar, 150 μm.

(C) Sox10 mRNA expression levels in fMaSC-derived organoids grown with FGFRi for 7 days. Y axis represents Sox10 mRNA levels normalized to the vehicle.

(D) FACS-based quantification of Venus+ cells in 7-day-old FGFRi-treated organoids grown from *Sox10*-H2BVenus fMaSCs or adult mammary luminal progenitors. Y axis represents the # of Venus+ cells as a % of the total # of cells in the primary organoids, normalized to the vehicle.

(E) FACS-based quantification of Venus+ cells in 8-day-old organoids grown from E18 *Sox10*-H2BVenus fMaSCs in defined growth factors. x axis is Venus fluorescence, and the number in the box is % gated Sox10+ cells.

(F) Whisker plots for Sox10 expression from the Metabric and UNC885 breast tumor databases across multiple subtypes. Each dot is a Sox10 expression value from a particular tumor.

Error bars represent SD.

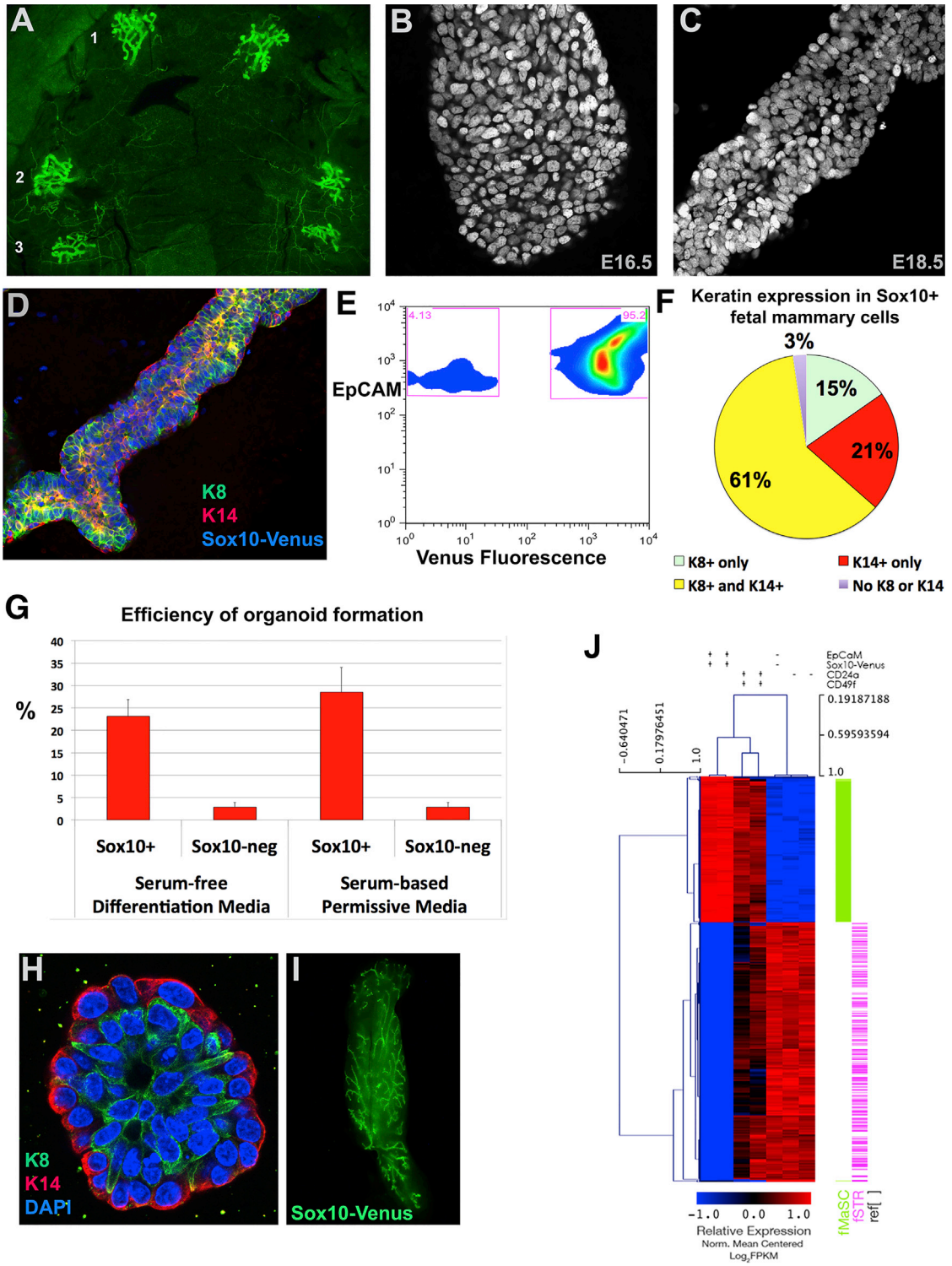


Figure 2. Sox10 Is a Fetal Mammary Stem Cell Marker that Improves fMaSC Purification

(A) Whole-mount view of the one to three mammary rudiment pairs in an E18 *Sox10*-H2BVenus embryo.

(B and C) Venus fluorescence in E16 and E18 *Sox10*-H2BVenus mammary rudiments whole mounts.

(D) Whole-mount mammary rudiment from E18 *Sox10*-H2BVenus embryo immunostained with luminal (K8) and basal (K14) markers.

(E) FACS of E18 *Sox10*-H2BVenus fetal mammary cells (pre-gated for EpCAM+ cells).

(F) Keratin immunostain of single E18 *Sox10*$^{flox\text{-}GFP}$ EpCAM+ fetal mammary cells.

(legend continued on next page)

multiple developing tissues through a feedback loop of unknown mechanism (Chen et al., 2014; Seymour et al., 2012). These observations led us to hypothesize that an FGF signaling axis may regulate Sox10 expression in mammary stem/progenitor cells.

To address this, we grew fMaSCs in 3D culture conditions in the presence of the pan-FGFR inhibitor, JNJ-42756493 (FGFRi). With vehicle only, fMaSCs form organoids when either epidermal growth factor (EGF) or basic FGF (FGF2) is added to the media but fail to form organoids if neither growth factor is present (Figure 1B; Figure S1). The addition of FGFRi blocks organoid formation if FGF is the only available growth factor. However, organoid formation is rescued upon adding EGF to media containing FGFRi (Figure 1B). As the number of dead cells does not increase in FGFRi-treated organoids (data not shown), these data demonstrate that fMaSC-derived organoids can utilize FGF signaling and indicate that FGFRi blocks FGF signaling without eliciting overt cytotoxicity.

To determine if FGF signaling regulates Sox10 expression in mammary cells, we measured Sox10 expression levels in fMaSC-derived organoids plated with vehicle or increasing concentrations of FGFRi. Organoid exposure to FGFRi resulted in significant dose-dependent decreases in Sox10 mRNA expression levels (Figure 1C). Similarly, by using a *Sox10*-H2BVenus bacterial artificial chromosome (BAC) transgenic mouse line (in which H2B-Venus is expressed under Sox10 transcriptional regulatory elements) to quantify the Sox10+ cells through Venus fluorescence, we found that FGFRi exposure significantly reduced the number of Sox10+ mammary organoid cells (Figure 1D). This effect was observed in a serum-based medium or in a serum-free medium (SFM) containing defined growth factors (Figure 1D; Figure S1). Organoids that were generated from adult luminal progenitors also showed a reduction in Sox10+ cells following FGFRi exposure (Figure 1D). fMaSCs grown in the presence of SFM with EGF + FGF10 developed into organoids with increased numbers of Sox10+ cells compared to fMaSCs grown only in SFM with EGF (Figure 1E). This effect was not seen in fMaSCs grown with SFM containing EGF + FGF2, indicating a specific role for FGF10 signaling through its cognate receptor, FGFR2b. No significant differences in Sox10 levels were observed in fMaSCs grown ± EGF (Figure S2). These data indicate that FGF signaling specifically regulates Sox10 expression levels in mammary cells.

To determine whether elevated Sox10 expression was a feature common to fMaSCs and their associated human cancer counterparts, we next analyzed the expression of Sox10 across a panel of tumor samples representing two distinct breast cancer datasets. This analysis revealed that basal-like and claudin-low breast cancers tend to express significantly higher levels of Sox10 than the other subtypes of the disease (Figure 1F), in accordance with two recent studies of Sox10 in breast cancer (Cimino-Mathews et al., 2013; Ivanov et al., 2013). These two subtypes comprise the bulk of triple-negative breast cancers, and both are frequently metastatic and aggressive. However, they differ in that basal-like breast cancers are weakly differentiated and the most fMaSC-like of the breast cancer subtypes, while claudin-low breast cancers possess the most EMT-like morphology and transcriptome among the breast cancer subtypes (Prat et al., 2010; Spike et al., 2012). These findings suggest that Sox10 expression may correlate with distinct stem and mesenchymal properties in human breast cancers.

Collectively, these data identify Sox10 as an FGF-responsive, mammary stem cell-associated transcription factor with likely roles in normal and transformed mammary cells.

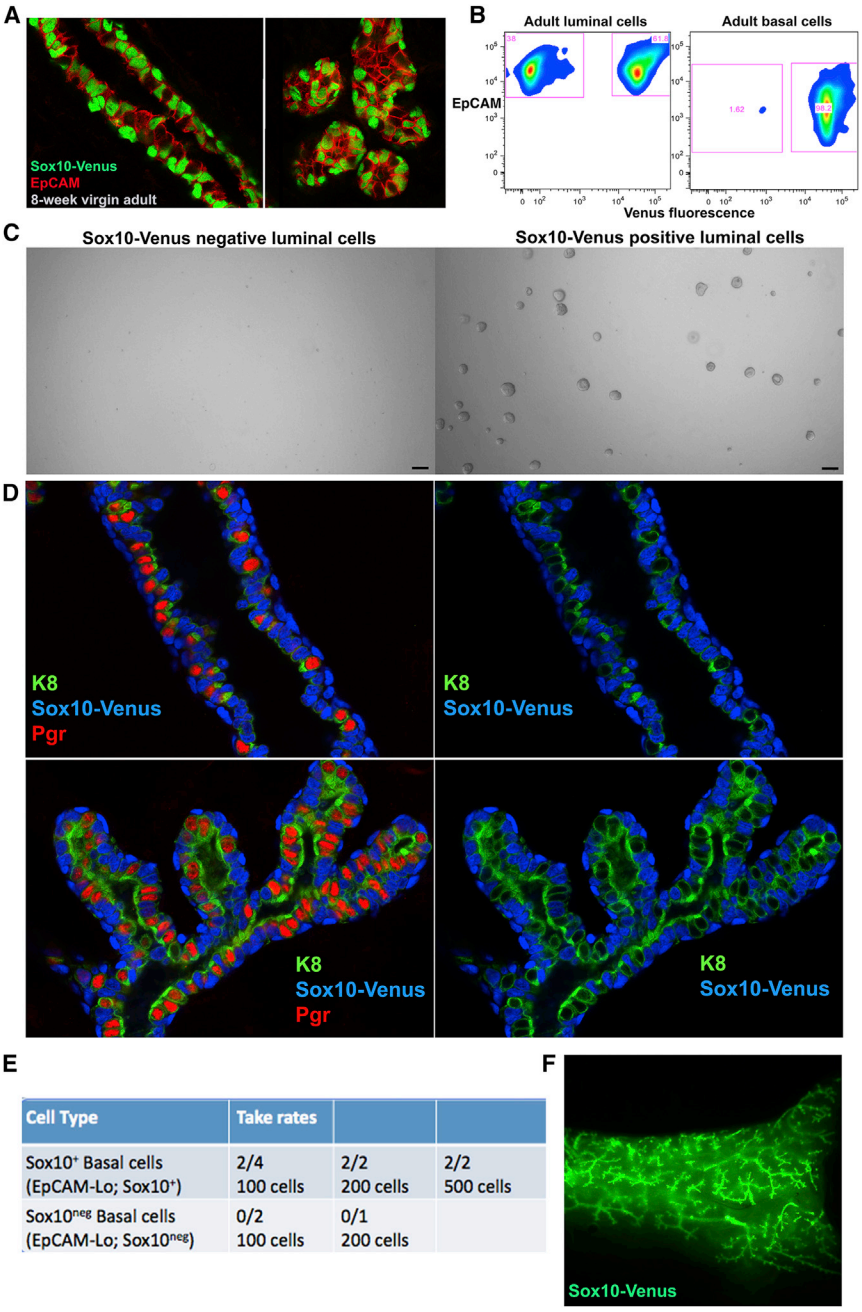
Sox10 Is a Fetal Mammary Stem Cell Marker that Improves fMaSC Purification

To elucidate the role of Sox10 in mammary cells, the *Sox10*-H2BVenus BAC transgenic mouse line was used to visualize Sox10+ cells. Consistent with the fMaSC transcriptome data, Sox10 was robustly expressed in all five fetal mammary rudiment pairs (Figures 2A–2C). The rudiments at these stages appear to be very primitive, as there is amorphous structure at embryonic day 16 (E16), while at E18, the lumen has not yet formed and there is no clear segregation of the luminal marker keratin-8 (K8) and the basal marker keratin-14 (K14) (Figure 2D).

Sox10+ fetal mammary cells were recovered using flow cytometry for more detailed molecular characterization. As cells in the rudiment can be distinguished from surrounding stromal cells by the epithelial cell adhesion marker (EpCAM), fetal Sox10+ mammary cells were isolated as Sox10+;EpCAM+. Consistent with Figure 2C, nearly all cells appear to be Sox10+ within the rudiment by fluorescence-activated cell sorting (FACS) analysis (Figure 2E). It is possible that the stability of the H2B-Venus fusion protein may yield cells that no longer express Sox10 but still retain the Venus fluorescence and thus overrepresent Sox10 expression. To address this, a *Sox10*$^{flox\text{-}GFP}$ mouse line in which a less stable GFP reporter is expressed from native Sox10 transcripts was also analyzed, and we confirmed that the majority of fetal mammary cells are Sox10+ (Figure S3). Consistent with the *Sox10*-H2BVenus whole-mount images, most single *Sox10*$^{flox\text{-}GFP}$ cells also co-expressed K8 and K14, suggesting that they may be bipotent progenitors or stem cells (Figure 2F).

Stem/progenitor cell function in these Sox10+ fetal cells was next analyzed using in vitro and in vivo stem/progenitor cell assays. Single fMaSCs grown in 3D culture conditions will clonally expand to generate bi-lineage organoids that resemble the architecture of the mammary gland with inner

(G) Efficiency of organoid formation from E18 *Sox10*-H2BVenus female mammary rudiments in two different media. y axis is number of organoids per 100 cells plated.
(H) A bi-lineage organoid derived from fMaSCs.
(I) A reconstituted mammary gland following transplantation of Sox10+ fetal cells visualized by *Sox10*-H2BVenus reporter.
(J) *Sox10*-H2BVenus-derived fMaSCs (columns 1 and 2), CD24/CD49f-derived fMaSCs (columns 3 and 4), and fStroma (columns 5–7) were RNA sequenced and clustered (SAM: FDR < 0.01%) using previously indicated differentially expressed genes between fMaSC (green) and fStroma (pink).
Error bars represent SD.



Cell Type	Take rates		
Sox10^{+} Basal cells (EpCAM-Lo; Sox10^{+})	2/4 100 cells	2/2 200 cells	2/2 500 cells
Sox10^{neg} Basal cells (EpCAM-Lo; Sox10^{neg})	0/2 100 cells	0/1 200 cells	

Figure 3. Sox10 Labels Cells with Stem/Progenitor Features in Adult Mammary Tissues

(A) Immunostain for EpCAM in an adult *Sox10*-H2BVenus mammary gland.

(B) FACS of Venus fluorescence (x axis) in adult *Sox10*-H2BVenus luminal and basal populations (y axis is EpCAM). Displayed are luminal cells that were pre-gated as EpCAM^{hi};CD49f$^{low\text{-}med}$, and basal cells as EpCAM$^{low\text{-}med}$;CD49f^{hi}.

(C) Venus(−) or Venus(+) luminal cells from an adult *Sox10*-H2BVenus mammary gland cultured in 3D for 6 days. Scale bar, 65 μm.

(D) Whole-mount immunofluorescence for K8 and progesterone receptor (Pgr) from adult *Sox10*-H2BVenus mammary glands; right image lacks Pgr for easier visualization.

(E) Transplantation take rates for Venus(−) and Venus(+) basal cells from an adult *Sox10*-H2BVenus mammary gland.

(F) A reconstituted mammary gland following transplantation of Sox10+ adult basal cells visualized by the *Sox10*-H2BVenus reporter.

K8+ luminal cells and external K14+ basal cells (Spike et al., 2012). When E18 Sox10+ fetal cells were plated as single cells into 3D culture conditions, they robustly formed bi-lineage organoids (Figures 2G and 2H; Figure S3). This demonstrates that the Sox10+ E18 population contains bipotent cells that generate both luminal- and basal-like cells. By contrast, the more rare Sox10^{neg} fetal mammary cells formed spheres at significantly reduced efficiency. As an in vivo metric of stem cell function, E18 Sox10+ fetal cells were also transplanted into cleared fat pads of immune-compromised mice. As few as five Sox10+ fetal cells were sufficient to generate a full mammary gland, further indicating that Sox10 positivity strongly correlates with fMaSC activity (Figure 2I; Figure S3). Collectively, the data demonstrate that Sox10 expression labels cells in the fetal mammary rudiment that possess bipotent stem/progenitor features.

Notably, the organoid-forming efficiency for fetal cells recovered with the *Sox10*-Venus and EpCAM markers represents a >3-fold improvement over the original CD24 and CD49f fMaSC marker strategy we previously employed. We isolated and RNA-sequenced E17 Sox10+;EpCAM+ fMaSCs and their surrounding fetal stromal cells (Table S1). In parallel, we RNA-sequenced E17 fMaSCs isolated by sorting for CD24^{hi};CD49f^{+} cells to assess the purification afforded by Sox10 and EpCAM. Comparison of these transcriptome profiles revealed that numerous stromal-associated genes were removed from the E17 fMaSC profile by using Sox10 expression to purify fMaSCs (Figure 2J). Taken together, our data show that using Sox10 as a marker produces an fMaSC population significantly purer than obtained previously.

Sox10 Labels Cells with Stem/Progenitor Features in Adult Mammary Tissues

We next analyzed Sox10 expression in the adult mammary gland. Immunofluorescence against positional markers such as EpCAM (high in luminal cells, low in basal cells) indicated that Sox10 expression was more restricted in the adult gland compared to the fetal mammary rudiment (Figure 3A). To

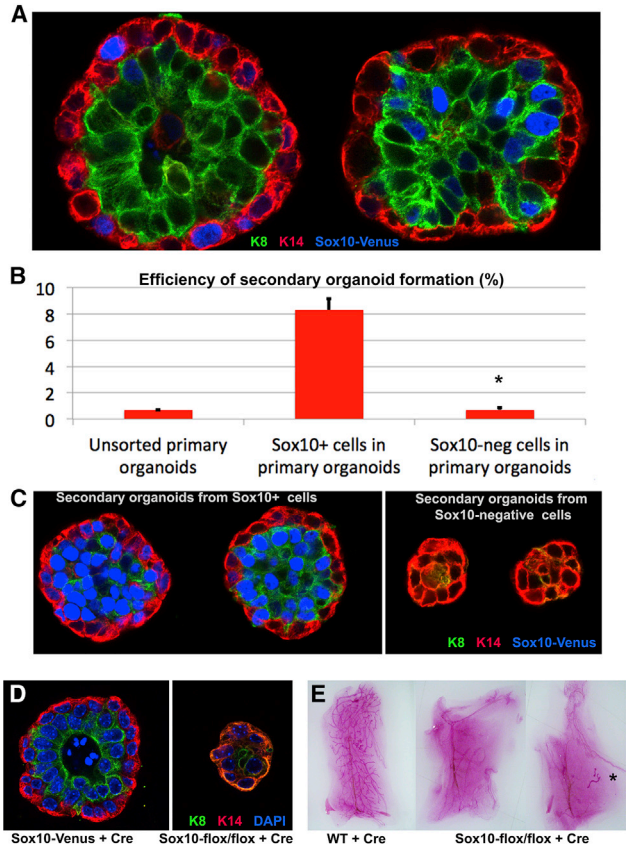


Figure 4. Sox10 Functionally Contributes to Stem/Progenitor Activity in Mammary Cells

(A) Organoids from *Sox10*-H2BVenus fMaSCs contain Venus(+) and Venus(−) cells.

(B and C) Efficiency of secondary organoid formation for Venus(+) and Venus(−) cells taken from primary *Sox10*-H2BVenus fMaSC organoids grown in SFM. y axis is number of secondary organoids per 100 cells plated.

(D) Representative organoid formation following 3D culture of Cre-infected *Sox10*$^{wild\text{-}type}$ or *Sox10*$^{flox/flox}$ fMaSCs.

(E) Carmine staining of transplanted Cre-infected *Sox10*$^{wild\text{-}type}$ or *Sox10*$^{flox/flox}$ fMaSCs into cleared fat pads. Transplants were considered takes if greater than half the fat pad was reconstituted; * marks a partial aborted outgrowth. Error bars represent SD.

quantify the expression of Sox10 by cell type, *Sox10*-H2BVenus and *Sox10*$^{flox\text{-}GFP}$ adult glands were FACS sorted into basal and luminal fractions using EpCAM/CD49f, and the percentage of Sox10+ cells in each fraction was then determined. These analyses revealed that nearly all basal cells express Sox10, whereas ~50% of luminal cells express Sox10 (Figure 3B; Figure S4).

Mammary stem/progenitor cell assays were performed on these Sox10+ basal and luminal cells to better understand their function in the gland. Sox10+ and Sox10^{neg} luminal cells were isolated by FACS and plated into 3D culture conditions. While Sox10+ luminal cells demonstrated sphere-forming potential with luminal characteristics (18.0 ± 2.1%), Sox10^{neg} luminal cells did not form spheres (0.3 ± 0.3%; Figure 3C; Figure S4). This suggests that Sox10+ luminal cells demarcate the colony-forming luminal progenitor cells in the luminal fraction of the mammary gland. Consistent with this, Sox10+ cells do not express progesterone receptor, a mature luminal cell marker, which is instead exclusively expressed in Sox10^{neg} luminal cells (Figure 3D). In the basal cell fraction, both Sox10+ and less common Sox10^{neg} basal cells were transplanted into cleared fat pads to determine MaSC function in an in vivo context. Sox10+ basal cells exhibited robust repopulation potential, whereas no successful transplantation was observed with Sox10^{neg} basal cells (Figures 3E and 3F). Sox10+ luminal cells also failed to exhibit successful transplantation, further indicating that these are lineage restricted progenitor cells.

These data indicate that populations with known mammary stem/progenitor cell properties—fMaSCs in the fetal rudiment, repopulating MaSCs in the adult basal fraction, and luminal progenitors in the luminal layer of the mammary gland—all appear to express Sox10.

Sox10 Labels Cultured Mammary Cells with Stem/Progenitor Characteristics In Vitro

The correlation of Sox10 expression with mammary stem/progenitor populations in vivo led us to next investigate if Sox10 also labels cells with these properties in organoids grown from fMaSCs in vitro. To address this, *Sox10*-H2BVenus fMaSCs were grown into bi-lineage organoids in 3D culture conditions. Intriguingly, these structures exhibited mosaic Sox10 expression in which Sox10+ and Sox10^{neg} cells were clearly evident (Figure 4A). To determine if these cells differ in stem/progenitor functionality, these populations were isolated and replated into identical organoid-forming conditions to generate secondary organoids in a classic surrogate assay of self-renewal for stem cells. Notably, Sox10+ cells from primary organoids had significantly greater potential to form secondary organoids than Sox10^{neg} cells (Figure 4B). Further, the secondary structures from Sox10+ cells were larger and yielded clear bi-lineage differentiation with both luminal and basal cell types present (Figure 4C). The rare secondary outgrowths derived from Sox10^{neg} cells were by contrast smaller and appeared to lack the bi-lineage structure observed in primary and Sox10+ secondary organoids (Figure 4C). These secondary organoids appeared to show more luminal-restricted Sox10 expression compared to primary organoids, which may reflect the restriction in stem/progenitor competence that occurs in this differentiation medium, and may mimic native mammary cell hierarchy. These data indicate that in addition to mammary cells in vivo, Sox10 labels populations with enhanced stem/progenitor functions in cultured mammary organoids in vitro.

Sox10 Functionally Contributes to Stem/Progenitor Activity in Mammary Cells

We next determined if Sox10 actively contributes to fMaSC function by performing stem/progenitor assays on cells in which Sox10 expression was ablated by deletion. We infected *Sox10*$^{flox/flox}$ and *Sox10*$^{wild\text{-}type}$ fMaSCs with Cre-expressing lentivirus to delete Sox10 from the *Sox10*^{flox} cells. While Cre-infected *Sox10*$^{wild\text{-}type}$ fMaSCs generated typical organoids with luminal and basal architecture resembling the mammary gland,

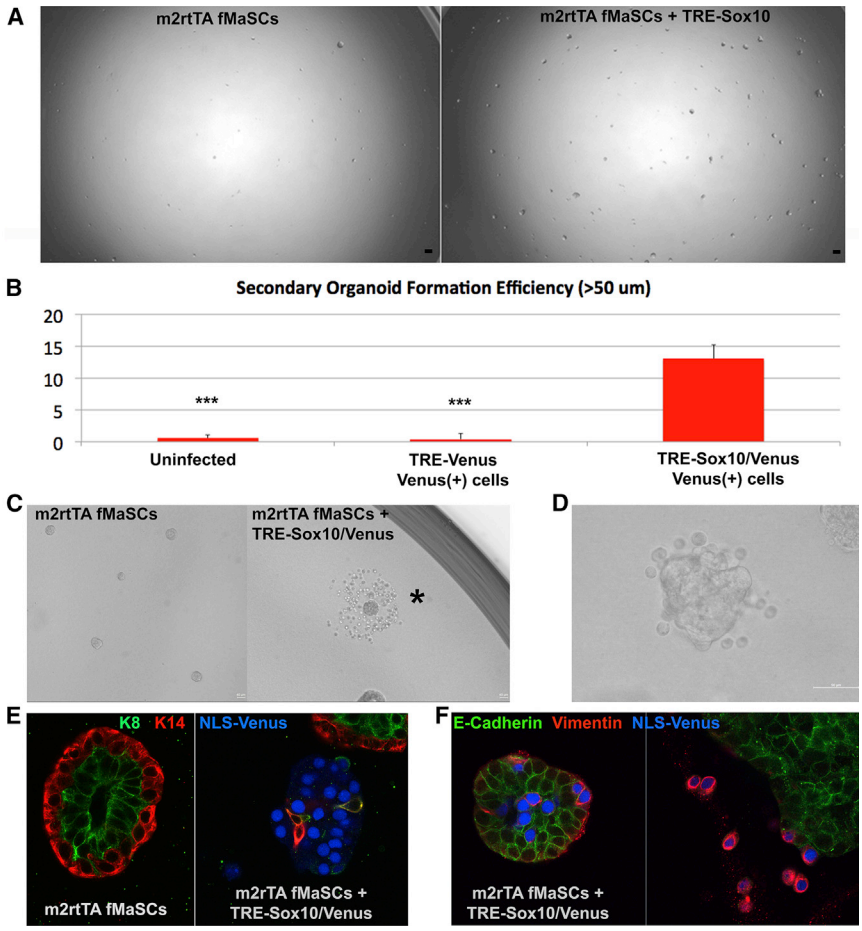


Figure 5. Ectopic Sox10 Expression Expands Stem/Progenitor Activity and Drives Acquisition of Mesenchymal Features

(A) Primary (1°) organoids from control (uninfected) or Sox10^{OE} m2rtTA fMaSCs were dissociated and replated into 3D culture to form secondary (2°) organoids. Shown is 2° organoid growth after 7 days. Scale bar, 75 μm.

(B) Quantification of 2° organoid-forming potential for Sox10^{OE} cells compared to uninfected or Venus-only-infected cells. y axis is # of >50 μm 2° organoids per 100 cells plated.

(C) Sox10^{OE} fMaSCs present with satellite single cell structures surrounding the 1° organoid (*). Scale bar, 40 μm.

(D) Active delamination of cells from a Sox10^{OE} organoid.

(E) Immunostains of control or Sox10^{OE} fMaSC organoids demonstrate the loss of keratin expression (red or green) in Sox10^{OE} cells (blue). Scale bar, 50 μm.

(F) Immunostains of Sox10^{OE} fMaSC organoids reveal upregulation of vimentin and loss of E-cadherin in Sox10^{OE} cells (blue).

Error bars represent SD.

the Cre-infected *Sox10*$^{flox/flox}$ fMaSCs generated fewer organoids, and the structures that did form were typically smaller and failed to develop the morphological features of multi-lineage organoids (Figure 4D; Figure S5).

We also performed transplantation assays with Cre-infected *Sox10*$^{flox/flox}$ fMaSCs or *Sox10*$^{flox/flox}$ adult basal cells to determine if cells were capable of generating full outgrowths following Sox10 deletion. No full outgrowths following transplantation were observed in the Sox10^{null} MaSCs, whereas equivalent numbers of control cells exhibited successful transplantation (Figure 4E; Figure S5). Together, these data indicate that Sox10 is required for full stem/progenitor cell functionality.

To determine if overexpression of Sox10 can increase stem/progenitor function in mammary cells, the Tet-on system was used to drive expression of human Sox10 in fMaSCs. fMaSCs isolated from a mouse strain that ubiquitously expresses the m2rtTA reverse tetracycline transactivator were infected with either LV-TRE-hSox10-2A-NLSVenus (doxycyline [dox] induces expression of Sox10 and Venus) or LV-TRE-NLSVenus (dox induces expression only of Venus) and allowed to form primary organoids. No apparent increase in primary organoid formation was observed with Sox10 overexpression (Sox10^{OE}). These primary organoids were then dissociated to single cells, replated into identical culture conditions, and scored for their ability to generate secondary organoids as a metric for increased persistence of stem/progenitor function. While fMaSCs that did not overexpress Sox10 showed low ability to form secondary organoids in differentiation medium (Figure 5A), Sox10^{OE} fMaSCs now demonstrated robust secondary organoid formation (Figures 5A and 5B). These data indicate that ectopic expression of Sox10 is able to increase or sustain stem/progenitor competence in cultured fetal mammary cells.

Ectopic Sox10 Expression Drives an EMT-like Response in fMaSC-Derived Organoids

While measuring the stem/progenitor function of Sox10^{OE} cells, we discovered that primary organoids with Sox10^{OE} cells demonstrated a novel morphology in which the primary organoid was surrounded by individual cells (Figure 5C). Video microscopy showed that the satellite cells originate from the delamination and extrusion of Sox10^{OE} cells from the primary organoid (Figure 5D; Movies S1 and S2). We found that Sox10^{OE} (Venus+) cells no longer expressed keratin markers, suggesting that the mobility of the cells might result from Sox10^{OE}-induced EMT (Figure 5E; Figure S6). Sox10^{OE} cells also presented with additional EMT markers, including downregulated expression of E-cadherin and upregulated expression of vimentin (Figure 5F; Figure S6). No such changes were observed in organoids not exposed to dox. These data demonstrate that Sox10 can directly mediate an EMT-like response when forcibly expressed at high levels in fMaSC-derived organoids.

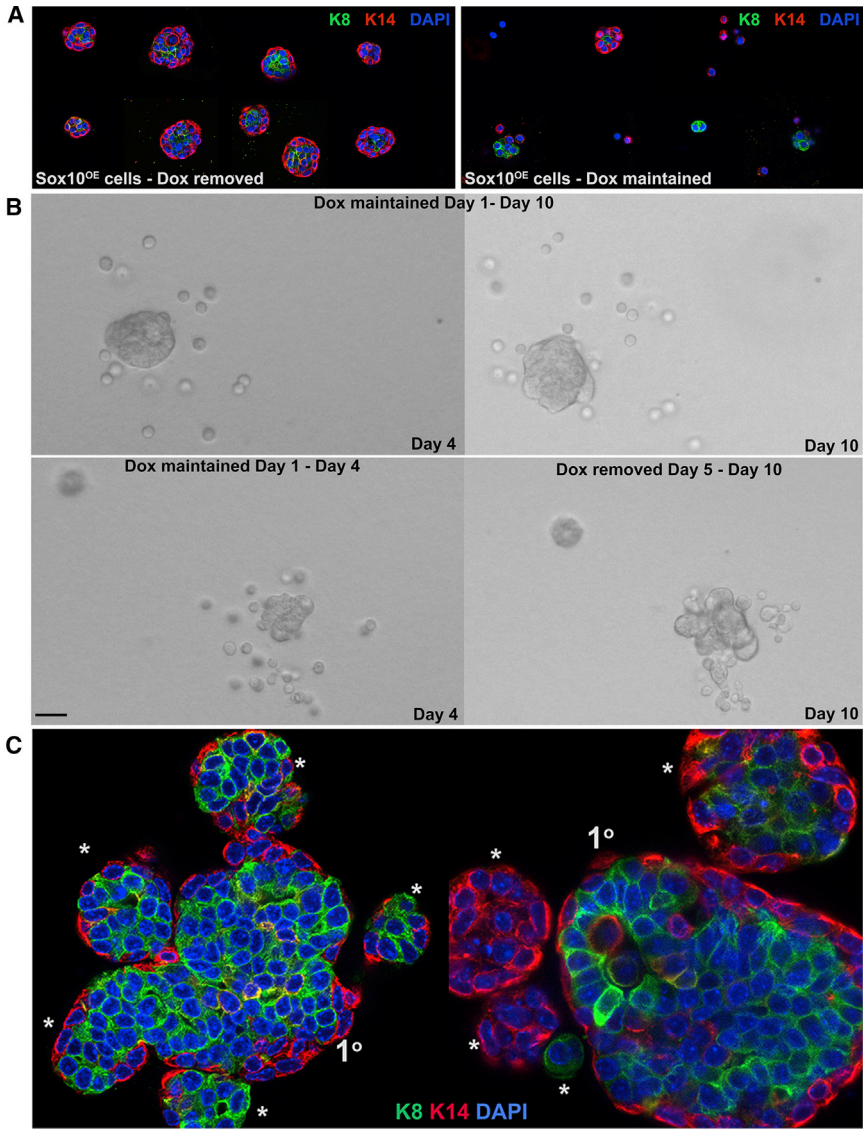


Figure 6. Reversal of Transient Sox10 Overexpression Restores Epithelial Features and Promotes Stem/Progenitor Activity

(A) Sox10^{OE} cells were isolated from 7-day-old fMaSC-derived primary (1°) organoids and replated in 3D culture ± dox. Secondary outgrowths from these cells were immunostained for keratin markers after 7 days.

(B) Sox10^{OE} satellite cells form secondary (2°) organoids surrounding the 1° organoid at greater efficiency if dox is removed from the media after 4 days. Left/right are the same organoids over 10 days of culture. Scale bar, 20 μm.

(C) Sox10^{OE} cells were allowed to form 1° organoids in 3D culture for 7 days, then dox was washed out of the media to ease Sox10 expression. 3–4 days after washout, the delaminated satellite cells initiated 2° organoid formation (*) around the 1° organoid.

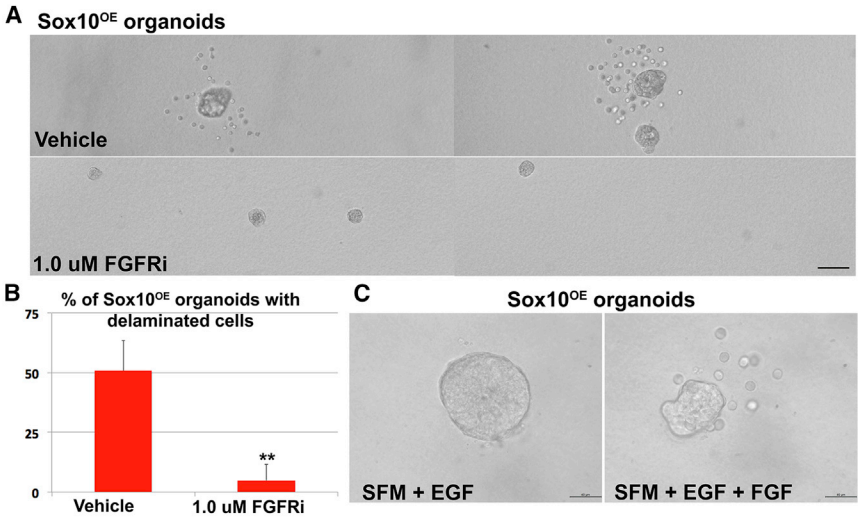
We next determined if the EMT state could be reversed in Sox10^{OE} mammary cells and if they retained or could regain bipotential stem/progenitor function. Sox10^{OE} mammary cells were isolated from primary organoid cultures and replated into 3D culture conditions with or without dox. The Sox10^{OE} mammary cells that were plated into dox, and thus maintained high Sox10 expression, often persisted as single cells and did not organize into secondary organoids (Figure 6A). However, when these same cells were plated into dox-free media, and Sox10 levels were reduced to baseline (Figure S7), the cells now favored the formation of bi-lineage secondary organoids (Figure 6A).

The same phenomenon was observed when Sox10^{OE} organoids that had undergone EMT and cell delamination were subjected to a protocol that removed dox from the media and lowered Sox10 expression to basal levels. While organoids continuously exposed to dox and high Sox10 levels showed mostly persistent single-cell satellite structures, the satellite cells in the dox-withdrawn organoids now initiated the formation of localized secondary organoids (Figure 6B). These secondary organoids exhibited the same bi-lineage features of primary fMaSC organoids, indicating that these single Sox10^{OE} cells have the potential to produce both luminal- and basal-like cells (Figure 6C). Notably, this robust secondary organoid formation occurred in the same strong differentiation media in which cells with retained stem/progenitor qualities are rare (Figure 4B), indicating the downstream effects of Sox10 serve to counterbalance these pro-differentiation factors.

These data reveal that at high levels of expression, Sox10 induces a mesenchymal transition that enables cell migration away from primary organoids. These cells are then capable of undergoing a mesenchymal-epithelial transition (MET) that mediates the formation of secondary organoids, which appears to be favored when Sox10 expression levels are reduced.

FGF Signaling Is Required for Sox10-Induced Cell Motility

We next attempted to identify mechanisms through which Sox10 evokes stem/progenitor and EMT/motility functions in mammary cells. The feedback loop between Sox transcription factors and FGF signaling that appears to involve Sox10 and FGF10 in mammary cells (Figure 1) suggests that these Sox10-mediated cell functions could involve FGF signaling. To test this, fMaSCs were manipulated to overexpress Sox10 as before, but this time in the presence of FGFRi. As expected, fMaSCs that were given vehicle formed primary organoids



D

	GO Terms	Genes	p-value
GO Terms associated with genes upregulated by Sox10-overexpression	Transport	221	4.90E-09
	Regulation of neuron projection development	32	1.20E-06
	Intracellular signal transduction	121	2.52E-06
	Cell projection morphogenesis	48	7.31E-06
	Regulation of cell communication	128	1.12E-05
GO Terms associated with genes downregulated by Sox10-overexpression	Cell adhesion	127	7.16E-25
	Cell differentiation	256	2.17E-23
	Regulation of cell proliferation	143	7.77E-22
	Regulation of programmed cell death	140	3.57E-19
	Tissue development	173	4.74E-31
	Organ development	256	1.71E-30

		Control	Sox10OE	Rank Change
Reported Sox10 or FGF signaling targets	Mitf	2.0	14.3	5272
	ErbB3	34.4	100.7	1978
	Mia	112.3	180.6	294
	Etv5	62.0	112.6	651
	Dusp6	459.2	200.5	-185
Genes associated with epithelial/mesenchymal or motile cell states	Postn	128.2	29.0	-6302
	Icam1	59.0	9.6	-6384
	Pvrl1	20.1	7.8	-3712
	Krt14	820.5	98.6	-1398
	Vim	293.3	1057.3	195
	Fyn	16.2	40.5	3187
	Enpp2	2.2	139.3	10873
	Limk1	3.9	15.1	4380
Genes associated with stem or differentiated cell states	FoxA1	63.9	24.9	-2843
	Esr1	4.2	0.3	-2961
	Pgr	5.7	2.0	-1952
	Myb	48.9	12.6	-4958
	Hmga1	74.5	20.0	-4071
	Runx1	62.5	8.7	-6829
	Id2	146.1	409.2	318
	Lbh	49.3	165.6	1359
	Dnmt3b	1.2	4.9	2630
	Tert	2.7	6.6	2067
	Kit	8.7	36.6	5429
	Tnfrsf19	0.0	161.6	13685

Figure 7. FGF Signaling Is Required for Sox10-Induced Cell Motility

(A) Sox10^{OE} organoids were grown in 3D culture in the presence of vehicle or 1.0 μM FGFRi. Scale bar, 100 μm.

(B) Fraction of Sox10^{OE} organoids with extruded satellite cells after 6 days (y axis) in the presence of vehicle or 1.0 μM FGFRi.

(C) Sox10^{OE} organoids were grown in 3D culture in SFM with EGF alone or EGF, FGF2, and FGF10. Scale bar, 40 μm.

(D) Gene Ontology terms associated with significantly down- or upregulated genes following Sox10^{OE} (top) and example notable genes with altered expression by Sox10^{OE} (bottom).

Error bars represent SD.

and the overexpression of Sox10 elicited an EMT-like delamination of cells (Figure 7A). However, this cell delamination was significantly attenuated in organoids that were exposed to the FGFRi, as indicated by the absence of satellite cells surrounding the primary organoid (Figures 7A and 7B). Sox10^{OE} organoids that were grown in media without FGF also failed to extrude satellite cells, confirming that it is inhibition of FGF signaling by the FGFRi that mediates this effect (Figure 7C). These data suggest that the potentiation of FGF signaling can be one effector of Sox10 that mediates cell delamination and that a pan-FGFRi blocks Sox10-induced motility in fMaSC-derived mammary organoids.

Transcriptome Analyses of Sox10^{OE} Cells Indicate Potential Mediators of Stem and EMT Functions

To more comprehensively profile the state changes elicited by Sox10 and to identify other potential direct or indirect targets of Sox10 that could mediate the stem/progenitor and EMT-like functions of Sox10, we performed transcriptome profiling of Sox10^{OE} cells through RNA sequencing (Table S2). In parallel, we also isolated and RNA-sequenced control organoid cells that did not overexpress Sox10 for comparison. To assess the quality of the sequencing data, we determined if previously described targets of Sox10 were upregulated in response to Sox10 overexpression. Published targets such as Mitf, Mia, and ErbB3 all showed elevated expression in Sox10^{OE} cells (Bondurand et al., 2000; Graf et al., 2014; Prasad et al., 2011) (Figure 7D). We also analyzed targets of FGF signaling, given our data linking Sox10 and FGF signaling. Among the targets induced by Sox10, we found that the FGF-positive signaling regulator Etv5 was upregulated, while the FGF negative regulator Dusp6 was downregulated (Figure 7D). This is consistent with the positive FGF-Sox10 loop indicated by our data, in which FGF acts to induce Sox10, while activated Sox10 then reinforces FGF signaling. These data validate that the differential expression of molecules between Sox10^{OE} and control cells can be used to identify targets of Sox10 or signaling network changes initiated by Sox10.

We next identified genes that were significantly differentially expressed in response to Sox10^{OE}. Gene ontology analysis with these gene lists indicated significant reprogramming of cellular function that is consistent with the observed phenotypic changes in Sox10^{OE} cells (Figure 7D; Table S2). For example, Sox10^{OE} cells delaminate from the primary organoid where

they tend to remain quiescent, and indeed this analysis finds genes associated with migration are upregulated with Sox10^{OE}, while genes associated with proliferation and adhesion are downregulated with Sox10^{OE}. Similarly, Sox10^{OE} cells in organoids lose differentiation marker expression and gain stem/progenitor function during this process, and indeed genes associated with differentiation are downregulated with Sox10^{OE}. These transcription data thus provide a hypothesis generating resource to determine how Sox10 elicits important state changes in normal or transformed mammary cells.

Notably, ErbB2 and the estrogen and progesterone hormone receptors all showed reduced expression levels following Sox10 overexpression. Sox10 is preferentially expressed in triple-negative breast cancers that lack these three receptors (Figure 1F). These data suggest that Sox10 may be one mechanism of functionally specifying this triple-negative state.

DISCUSSION

Our studies have used diverse strategies to reveal important roles for Sox10 in stem and progenitor functions within mammary cells. This is first indicated by the significant correlation between Sox10 expression and two aggressive subtypes of breast cancer that have previously been described as stem-like (basal-like) or EMT-like (claudin-low). We then present data that Sox10 consistently labels cells with stem/progenitor qualities in multiple contexts that include fetal, adult, and 3D cultured mammary tissues. Sox10 may be a cell state regulatory node in mammary cells, as deleting Sox10 decreased stem/progenitor functions, while its ectopic activation both expanded stem/progenitor activity and induced EMT. This suggests that relative expression levels of Sox10 can mediate either stem-like or EMT-like responses depending on context.

The link between Sox10 and both stem- and EMT-like cell functions is reminiscent of the published links between CSCs and EMT (Oskarsson et al., 2014). Importantly, it has been unclear to what extent CSCs are stem-like, given that their mesenchymal properties and transcriptome profiles often do not resemble those of bone fide stem cells. The enhanced motility of mesenchymalized cells may endow them with greater capacity to aggregate and form polyclonal ''tumorspheres'' in suspension cultures or to invade and form tumors more efficiently in xenograft assays. These properties are clearly independent of stemness measured by transcription profiling, and should not be used as surrogates for stem cell function. These concerns have led to the rebranding of CSCs as ''tumor-'' or ''xenograft-initiating cells,'' which suggests the distinction between the stem-like cells in tumors identified transcriptionally, and the more EMT-like CSCs.

The data described here present clear evidence that the stem cell and mesenchymal states are related and can be interconverted in stem-like cells. We find that a single factor, Sox10, is able to contribute to cells entering each of these two states, and critically, we show that it does so independently of the other state. Sox10+ cells that have not undergone EMT show increased levels of stemness in multiple contexts, while EMT occurs independent of stem cell activity. The separation of these states removes the aforementioned concerns about conflating stemness with properties of mesenchymal cells, and demonstrates that a single molecule such as Sox10 can link these two distinct states. Importantly, this affirms the link between stem-like and mesenchymal states and defines a molecular mechanism by which these state conversions can take place.

These data also yield predictions about how mammary cells acquire stem cell-like properties in normal and cancerous states and how these mechanisms may contribute to metastatic disease. The capacity of Sox10 to promote both stem-like and EMT-like behaviors suggests that Sox10 could be a factor that mediates these two functions that are hypothesized to be directly responsible for tumor initiation and progression. Most notably, we have modeled the sequential stages of metastatic behavior using only Sox10 in 3D mammary cell culture, as we find that (1) Sox10^{+} cells preferably form primary organoids, (2) Sox10^{OE} activates EMT to elicit delamination and migration of cells away from the primary organoid, and (3) reduction of Sox10 levels in these cells reverses the EMT and initiates the establishment of separate organoids at secondary sites. It is easy to visualize how this could similarly play out in Sox10+ tumors, in which microenvironmental or genomic changes could induce fluctuations in Sox10 expression levels that cycle cells through these stem-like and EMT states to mediate metastasis.

Our findings also have implications for how stem/progenitor cell states may be specified in mammary cells. As discussed in the introduction, the balanced activation of specific lineage determining factors is a mechanism capable of mediating stem-like functions in cells. This model fits with observations of Sox family transcription factors, where Sox molecules have antagonistic relationships with other factors at cell-fate decision points. By applying this model to Sox10 and mammary cells, our data indicate that Sox10 may specify the basal lineage in mammary cells. This is apparent in the expression data, where Sox10 preferentially labels the basal cell fraction in the adult mammary gland, and the functional data, as Sox10^{OE} can elicit EMT in mammary cells, and basal cells can be considered ''partial EMT'' based on their morphology. Furthermore, this model predicts that Sox10 should promote stem-like qualities when in balance with other factors. This is supported by our data linking Sox10 expression and function to stem-like properties and our data demonstrating that lower levels of Sox10 expression increase efficiency of bi-lineage sphere formation and self-renewal. These data thus support a model in which cell-fate decisions and stemness in mammary cells are regulated by a balance of lineage specifiers, of which Sox10 is one critical player that favors a basal lineage. However, there are pieces of our data that do not neatly fit this model, such as that Sox10^{neg} cells produce mostly basal-like organoids and Sox10^{OE} elicits cells that appear less differentiated. This suggests that a function of Sox10 may be to provide cell-state plasticity, instead of, or in addition to, a role in lineage specification.

As described in the Introduction, there is not a consensus on the localization and frequency for MaSCs. Our data and the balanced lineage specifier model suggest that a significant reservoir of Sox10-expressing poised basal cells exists and that these cells could adopt activated stem/progenitor cell properties by the acquisition of antagonistic factors that bring Sox10 levels into an equilibrium that favors a stem cell state.

This is consistent with work that indicates the majority of single basal cells have the potential to generate full mammary glands (Prater et al., 2014). Evaluating this model will require a better understanding of how Sox10 works in concert with other, presumably pro-luminal factors, such as Elf5, Gata3, and Notch signaling, among others. Similarly, it will be key to evaluate the relationship of Sox10 with basal lineage regulators such as p63 and Slug and the stem-cell marker Lgr5 (Oakes et al., 2014).

Finally, two of our most striking results are that the use of an FGFR inhibitor profoundly affects the expression of Sox10 and the delamination phenotype induced through $Sox10^{OE}$. Notably, the deletion of *FGFR1* and *FGFR2* results in the loss of the transplantation competent population of mammary stem cells and compromises ductal remodeling, which mirror the roles for Sox10 in stem cell competence and cell motility shown here (Pond et al., 2013). Extrinsic signaling mechanisms in the stem cell niche that regulate the frequency and output of stem cells are potential targets for cancer prevention or treatment. Thus, it will be key to determine if blocking FGF signaling also antagonizes the expression or downstream effects of Sox10 (or other Sox family transcription factors) in vivo in normal mammary tissue or tumors. Together, these data imply a central role for FGF signaling and Sox10 in normal mammary function and indicate that tight control is required to prevent it from eliciting malignant functions.

EXPERIMENTAL PROCEDURES

Mice

Mice were housed in accordance with NIH guidelines in Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-accredited facilities at the Salk Institute. All experimental protocols were approved by the Salk Institute Institutional Animal Care and Use Committee.

Mammary Cell Preparation

Single-cell preparations of fetal mammary cells were obtained by pooling freshly dissected fetal mammary rudiments from euthanized embryos into dissociation media (Epicult-B Basal medium [STEMCELL Technologies] supplemented with 5% fetal bovine serum [FBS], penicillin/streptomycin, fungizone, hydrocortisone, collagenase, and hyaluronidase). Rudiments were then dissociated to single cells by sequentially incubating them in dissociation medium for 1.5 hr at 37°C with gentle agitation, exposing them to ammonium chloride for 4 min on ice to remove erythrocytes and triturating them with dispase and DNase. Final suspensions were passed through a 40-μm filter to remove aggregated cells and stored in Hank's balanced salt solution with 2% FBS for flow cytometry. Single-cell preparations of adult mammary cells were made by dissecting out and mincing the #4 mammary glands from 6- to 12-week-old virgin female mice. Glands were then dissociated by agitating them for 3–6 hr at 37°C in the same dissociation media. Cells were further processed as with the fetal cells, except that trypsin and Accutase (Life Technologies) were also utilized prior to dispase treatment to facilitate disaggregation. Final suspensions were passed through a 40-μm filter to remove cell clusters and stored in Hank's balanced salt solution with 2% FBS for flow cytometry.

Immunostaining and Confocal Analyses

Mammary tissues were immunostained through direct or indirect immunofluorescence. Confocal microscopy was performed with equipment from the Waitt Advanced Biophotonics Center at the Salk Institute, including Zeiss 780 inverted laser scanning confocal microscopes. Details of tissue preparation and staining protocol are included in Supplemental Experimental Procedures.

3D Organoid Culture

To generate organoids, single mammary cells were plated at 50–650 cells per well in 96-well ultra low-adhesion plates (Costar) with Matrigel. Cells were plated in either restricted serum-free media (Epicult-B media with B-supplement [STEMCELL Technologies] containing heparin and penicillin/streptomycin and defined growth factors such as EGF, FGF2, and/or FGF10) or in serum-based MCF10A media (DMEM/F12 with 5% horse serum, hydrocortisone, cholera toxin, insulin, and ciproflaxin, supplemented with B27 supplement and EGF). Description of the plating protocol and analysis of these cells is in Supplemental Experimental Procedures.

4D Organoid Culture and Imaging

m2rtTA fMaSCs were infected with LV-TRE-hSox10-2A-NLSVenus and plated onto glass-bottom 35-mm dishes with a Matrigel bed in restricted serum-free media. After 72 hr, organoids were given fresh media and dox to induce Sox10/Venus expression. 8–24 hr later, cells were imaged at 10-min intervals with a Zeiss CSU Spinning Disk Confocal Microscope in a climate-controlled environment of 5% CO_2 and 37°C. Images were assembled into movies using Imaris imaging software.

RNA Sequencing and Bioinformatic Analyses

RNA isolation, sequencing, and analysis are described in detail in Supplemental Experimental Procedures.

Statistical Analyses

A two-tailed Student's t test was used to quantify significance. p values were represented as follows: $^{*}p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.0001$.

Additional experimental procedures are described in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the RNA-sequencing data reported in this paper is GEO: GSE71300.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, 11 figures, two tables, and two movies and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.08.040.

AUTHOR CONTRIBUTIONS

C.D. and G.M.W. designed the study. B.T.S., C.J., C.D., J.C.H., and C.M.P. performed bioinformatic analyses. B.T.S. and C.L.T. performed transplants. C.D. acquired all other data. C.D. wrote the manuscript; all authors facilitated revisions. G.M.W. supervised the study.

ACKNOWLEDGMENTS

Thanks to Rose Rodewald for lab management, Karissa Huang and Lisa Zeng for genotyping, Wenjun Guo for providing LV-TRE-Sox10, Michael Wegner for $Sox10^{flox}$ mice, Andrew Loftus for 4D imaging assistance, Sven Heinz, Manching Ku, Chris Benner, Max Shokhirev, and Yang Dai for RNA-sequencing assistance. Cancer Center Core Grant CA014195, NIH 5T32CA009370, Department of Defense W81XWH-12-1-0106/ W81XWH-12-1-0107, Susan G. Komen Foundation SAC110036, and the BCRF supported this work. Janssen Pharmaceuticals funded analyses of FGFRi effects on fMaSCs, with assistance from Tim Perera and Eli Jovcheva. C.D. is supported by NRSA fellowship F32CA174430.

Received: February 2, 2015
Revised: June 14, 2015
Accepted: August 12, 2015
Published: September 10, 2015

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.

Barker, N., Ridgway, R.A., van Es, J.H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J., and Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. Nature *457*, 608–611.

Bondurand, N., Pingault, V., Goerich, D.E., Lemort, N., Sock, E., Le Caignec, C., Wegner, M., and Goossens, M. (2000). Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. Hum. Mol. Genet. *9*, 1907–1917.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. *3*, 730–737.

Chen, J., Li, Y., Yu, T.S., McKay, R.M., Burns, D.K., Kernie, S.G., and Parada, L.F. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. Nature *488*, 522–526.

Chen, Z., Huang, J., Liu, Y., Dattilo, L.K., Huh, S.H., Ornitz, D., and Beebe, D.C. (2014). FGF signaling activates a Sox9-Sox10 pathway for the formation and branching morphogenesis of mouse ocular glands. Development *141*, 2691–2701.

Cimino-Mathews, A., Subhawong, A.P., Elwood, H., Warzecha, H.N., Sharma, R., Park, B.H., Taube, J.M., Illei, P.B., and Argani, P. (2013). Neural crest transcription factor Sox10 is preferentially expressed in triple-negative and metaplastic breast carcinomas. Hum. Pathol. *44*, 959–965.

Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J., et al. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nat. Med. *17*, 1086–1093.

Graf, S.A., Busch, C., Bosserhoff, A.K., Besch, R., and Berking, C. (2014). SOX10 promotes melanoma cell invasion by regulating melanoma inhibitory activity. J. Invest. Dermatol. *134*, 2212–2220.

Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zürrer-Härdi, U., Bell, G., et al. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. Cell *148*, 1015–1028.

Hornig, J., Fröb, F., Vogl, M.R., Hermans-Borgmeyer, I., Tamm, E.R., and Wegner, M. (2013). The transcription factors Sox10 and Myrf define an essential regulatory network module in differentiating oligodendrocytes. PLoS Genet. *9*, e1003907.

Ivanov, S.V., Panaccione, A., Nonaka, D., Prasad, M.L., Boyd, K.L., Brown, B., Guo, Y., Sewell, A., and Yarbrough, W.G. (2013). Diagnostic SOX10 gene signatures in salivary adenoid cystic and breast basal-like carcinomas. Br. J. Cancer *109*, 444–451.

Kim, Y.J., Lim, H., Li, Z., Oh, Y., Kovlyagina, I., Choi, I.Y., Dong, X., and Lee, G. (2014). Generation of multipotent induced neural crest by direct reprogramming of human postnatal fibroblasts with a single transcription factor. Cell Stem Cell *15*, 497–506.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells *26*, 903–911.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., et al.; kConFab (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913.

Loh, K.M., and Lim, B. (2011). A precarious balance: pluripotency factors as lineage specifiers. Cell Stem Cell *8*, 363–369.

Lu, P., Ewald, A.J., Martin, G.R., and Werb, Z. (2008). Genetic mosaic analysis reveals FGF receptor 2 function in terminal end buds during mammary gland branching morphogenesis. Dev. Biol. *321*, 77–87.

Mailleux, A.A., Spencer-Dene, B., Dillon, C., Ndiaye, D., Savona-Baron, C., Itoh, N., Kato, S., Dickson, C., Thiery, J.P., and Bellusci, S. (2002). Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. Development *129*, 53–60.

Masson, N., and Ratcliffe, P.J. (2014). Hypoxia signaling pathways in cancer metabolism: the importance of co-selecting interconnected physiological pathways. Cancer Metab. *2*, 3.

Merlos-Suárez, A., Barriga, F.M., Jung, P., Iglesias, M., Céspedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Muñoz, P., et al. (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell *8*, 511–524.

Najm, F.J., Lager, A.M., Zaremba, A., Wyatt, K., Caprariello, A.V., Factor, D.C., Karl, R.T., Maeda, T., Miller, R.H., and Tesar, P.J. (2013). Transcription factor-mediated reprogramming of fibroblasts to expandable, myelinogenic oligodendrocyte progenitor cells. Nat. Biotechnol. *31*, 426–433.

Nassour, M., Idoux-Gillet, Y., Selmi, A., Côme, C., Faraldo, M.L., Deugnier, M.A., and Savagner, P. (2012). Slug controls stem/progenitor cell growth dynamics during mammary gland morphogenesis. PLoS ONE *7*, e53498.

Oakes, S.R., Gallego-Ortega, D., and Ormandy, C.J. (2014). The mammary cellular hierarchy and breast cancer. Cell. Mol. Life Sci. *71*, 4301–4324.

Oskarsson, T., Batlle, E., and Massagué, J. (2014). Metastatic stem cells: sources, niches, and vital pathways. Cell Stem Cell *14*, 306–321.

Pond, A.C., Bin, X., Batts, T., Roarty, K., Hilsenbeck, S., and Rosen, J.M. (2013). Fibroblast growth factor receptor signaling is essential for normal mammary gland development and stem cell function. Stem Cells *31*, 178–189.

Prasad, M.K., Reed, X., Gorkin, D.U., Cronin, J.C., McAdow, A.R., Chain, K., Hodonsky, C.J., Jones, E.A., Svaren, J., Antonellis, A., et al. (2011). SOX10 directly modulates ERBB3 transcription via an intronic neural crest enhancer. BMC Dev. Biol. *11*, 40.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*, R68.

Prater, M.D., Petit, V., Alasdair Russell, I., Giraddi, R.R., Shehata, M., Menon, S., Schulte, R., Kalajzic, I., Rath, N., Olson, M.F., et al. (2014). Mammary stem cells have myoepithelial cell properties. Nat. Cell Biol. *16*, 942–950, 1–7.

Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2014). In situ identification of bipotent stem cells in the mammary gland. Nature *506*, 322–327.

Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell *12*, 15–30.

Schepers, A.G., Snippert, H.J., Stange, D.E., van den Born, M., van Es, J.H., van de Wetering, M., and Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science *337*, 730–735.

Schwitalla, S., Fingerle, A.A., Cammareri, P., Nebelsiek, T., Göktuna, S.I., Ziegler, P.K., Canli, O., Heijmans, J., Huels, D.J., Moreaux, G., et al. (2013). Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. Cell *152*, 25–38.

Seymour, P.A., Shih, H.P., Patel, N.A., Freude, K.K., Xie, R., Lim, C.J., and Sander, M. (2012). A Sox9/Fgf feed-forward loop maintains pancreatic organ identity. Development *139*, 3363–3372.

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.

Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I.A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C.J., and Stingl, J. (2012). Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. Breast Cancer Res. *14*, R134.

Spike, B.T., Engle, D.D., Lin, J.C., Cheung, S.K., La, J., and Wahl, G.M. (2012). A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. Cell Stem Cell *10*, 183–197.

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

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A., et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature *503*, 218–223.

Tomasetti, C., and Vogelstein, B. (2015). Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science *347*, 78–81.

Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. Nature *479*, 189–193.

Yang, N., Zuchero, J.B., Ahlenius, H., Marro, S., Ng, Y.H., Vierbuchen, T., Hawkins, J.S., Geissler, R., Barres, B.A., and Wernig, M. (2013). Generation of oligodendroglial cells by direct lineage conversion. Nat. Biotechnol. *31*, 434–439.