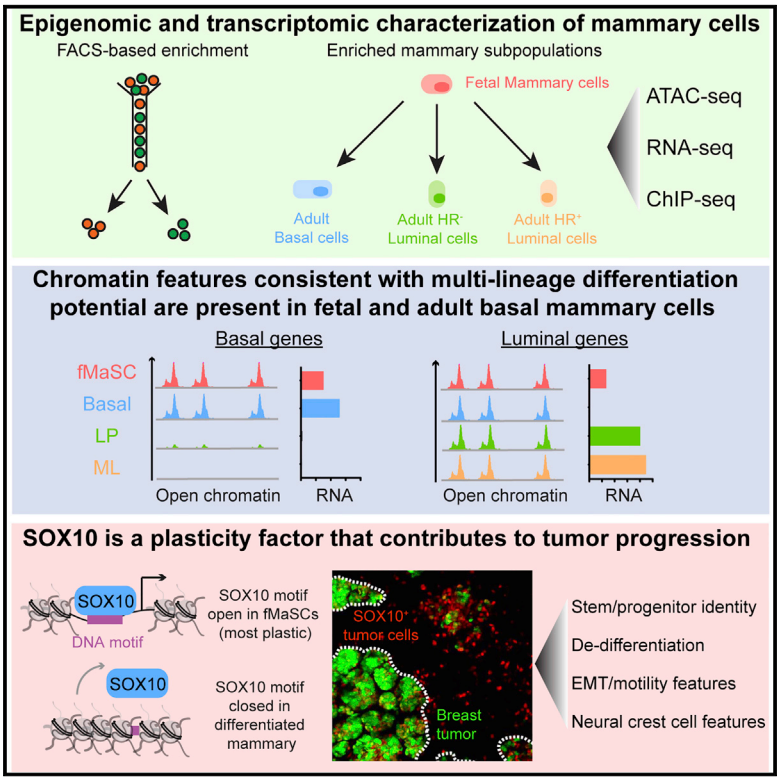
Article

Cancer Cell

Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity

Graphical Abstract



Highlights

- Developmental changes reveal mammary differentiation state regulators
- Multi-lineage differentiation potential is evident in adult basal mammary cells
- SOX10^{+} expression correlates with mammary tumor cell state plasticity
- SOX10 activity can elicit neural crest-like features in mammary tumors

Authors

Christopher Dravis, Chi-Yeh Chung, Nikki K. Lytle, ..., Christy L. Trejo, Tannishtha Reya, Geoffrey M. Wahl

Correspondence

cdravis@salk.edu (C.D.), wahl@salk.edu (G.M.W.)

In Brief

Dravis et al. use chromatin accessibility assays and transcriptional profiling during mammary development to identify factors that mediate breast cancer cell state interconversions and find SOX10 as a key factor. Mammary tumor cells expressing high SOX10 acquire a motile, neural crest-like state.

Dravis et al., 2018, Cancer Cell *34*, 466–482
September 10, 2018
https://doi.org/10.1016/j.ccell.2018.08.001

CellPress

CellPress

Cancer Cell
Article

Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity

Christopher Dravis,[1,4,*] Chi-Yeh Chung,[1,4] Nikki K. Lytle,[2,3] Jaslem Herrera-Valdez,[1] Gidsela Luna,[1] Christy L. Trejo,[1] Tannishtha Reya,[2,3] and Geoffrey M. Wahl[1,5,*]

[1]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[2]Sanford Consortium for Regenerative Medicine, La Jolla, CA 92037, USA
[3]Departments of Pharmacology and Medicine, Moores Cancer Center, University of California San Diego School of Medicine, La Jolla, CA 92037, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: cdravis@salk.edu (C.D.), wahl@salk.edu (G.M.W.)
https://doi.org/10.1016/j.ccell.2018.08.001

SUMMARY

Cell state reprogramming during tumor progression complicates accurate diagnosis, compromises therapeutic effectiveness, and fuels metastatic dissemination. We used chromatin accessibility assays and transcriptional profiling during mammary development as an agnostic approach to identify factors that mediate cancer cell state interconversions. We show that fetal and adult basal cells share epigenetic features consistent with multi-lineage differentiation potential. We find that DNA-binding motifs for SOX transcription factors are enriched in chromatin that is accessible in stem/progenitor cells and inaccessible in differentiated cells. In both mouse and human tumors, SOX10 expression correlates with stem/progenitor identity, dedifferentiation, and invasive characteristics. Strikingly, we demonstrate that SOX10 binds to genes that regulate neural crest cell identity, and that SOX10-positive tumor cells exhibit neural crest cell features.

INTRODUCTION

At diagnosis, most tumors present as heterogeneous collections of tumor cells and stroma. Many factors promote the cell state changes that contribute to tumor heterogeneity, drug resistance, tumor metastasis, and poor patient outcomes (Marusyk et al., 2012; Wainwright and Scaffidi, 2017). Recent studies reveal that some oncogene-induced cell state changes that occur during tumor progression can be traced to mechanisms enabling cancer cells to adopt behaviors that are not part of their homeostatic repertoire (Ge and Fuchs, 2018). This behavior, which we will refer to as cell state instability or plasticity, has some of the characteristics of the lineage infidelity acquired during wound healing (Ge et al., 2017), or inflammation- or oncogene-associated reprogramming to a multi-potential embryonic or stem-like state (Wahl and Spike, 2017). Better understanding of the mechanisms that underlie cell state instability in tumor progression could create opportunities for therapeutic intervention.

The cancer stem cell (CSC) hypothesis was initially attractive because it predicted the existence of a cellular subpopulation uniquely able to generate intra-tumoral heterogeneity, and that therapeutic targeting of these cells would prevent subsequent tumor evolution. However, it is now established that even differentiated cells can be reprogrammed into stem-like cells, suggesting that cell state reprogramming is more common and occurs in more diverse cell types than previously thought (Schwitalla et al., 2013; Tata et al., 2013). Indeed, this type of reprogramming can be used to re-establish stem-like hierarchies in

Significance

Tumor cells can reprogram into different cell states, contributing to the intra-tumoral heterogeneity that can result in drug resistance and metastasis. As cancer is a caricature of normal development, we analyzed developmentally plastic fetal mammary cells and their differentiated descendants to identify putative regulators of cell state changes during normal and tumor development. These analyses identify the transcription factor SOX10 in the control of normal mammary development, and in the genesis of mammary tumor cell plasticity in murine models and human breast cancer. We show that one manifestation of high SOX10 expression is acquisition of a motile, neural crest-like state in mammary tumors. Our results hold significance for mitigation of tumor cell plasticity and for interception of motile tumor cells.

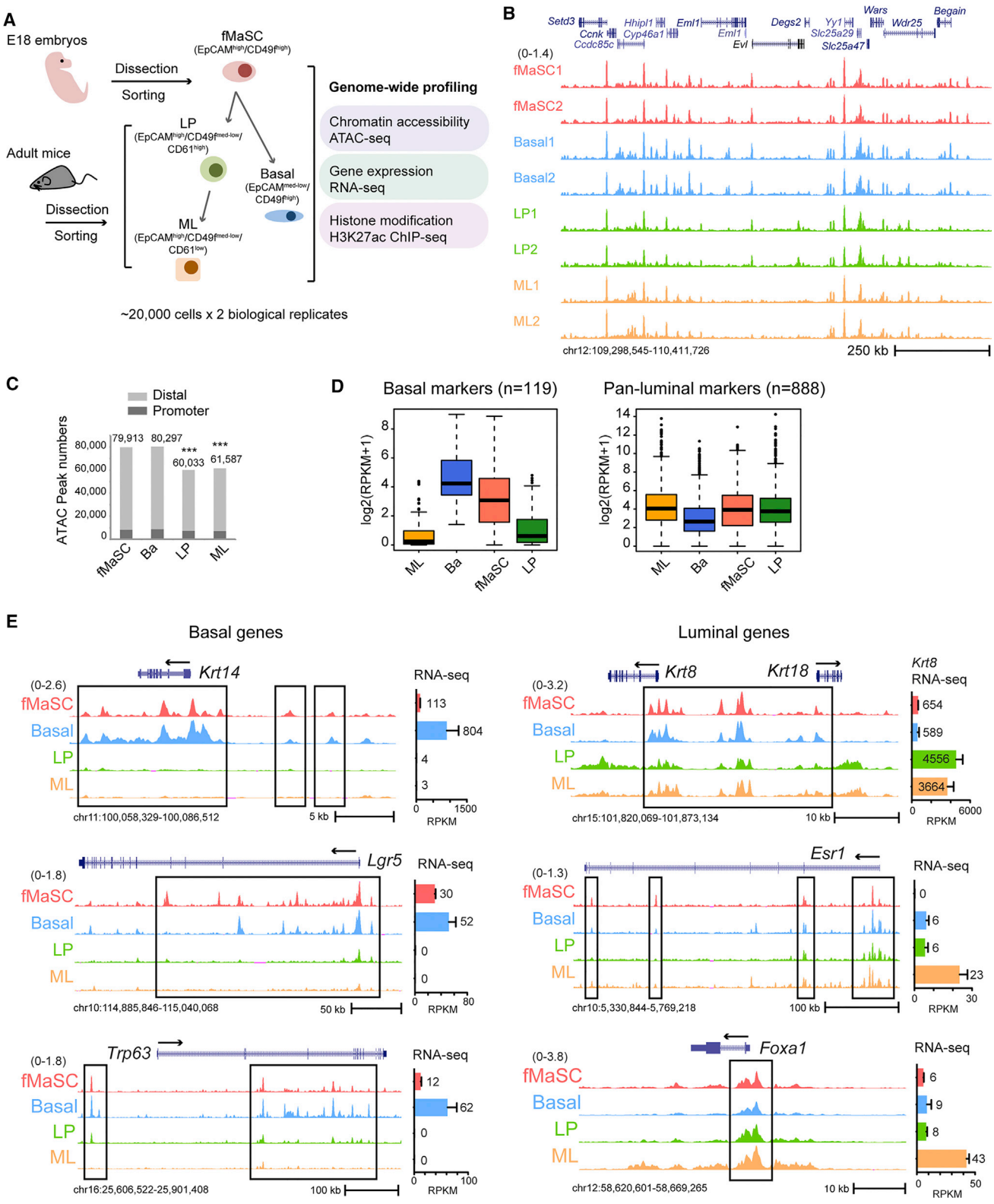


Figure 1. Multi-lineage Potential Is Present in Fetal and Adult Basal Mammary Cells

(A) Experimental strategy for epigenetic and transcriptional profiling of mammary cells.

(B) Representative ATAC-seq profiles of biological replicates from fMaSCs, adult basal cells, luminal progenitors (LP), and mature luminal cells (ML).

(legend continued on next page)

tumors even after elimination of putative CSCs (de Sousa e Melo et al., 2017; Shimokawa et al., 2017). These data suggest that eliminating phenotypically unstable cells will likely be fruitless, as other cells will take their place. Rather, abrogating the mechanisms by which tumor cells gain cell state plasticity may be more productive.

We focused on the relationship between mammary gland development and aggressive breast cancers to better understand the mechanisms by which differentiated cells revert to other cell states and by which intra-tumoral heterogeneity and malignancy arise. Despite its structural simplicity, the mammary gland undergoes impressive growth and invasive phases during development, cyclical expansive and apoptotic phases controlled by estrus cycles, and massive tissue expansion and involution associated with pregnancy and lactation (Inman et al., 2015). Clearly, mammary ducts must contain cells with significant growth, invasive, and multi-lineage potential. The coordinated cell state changes these cells undergo make the mammary tissue an excellent system in which to study mechanisms of cell state plasticity.

Whether adult mammary gland homeostasis requires a hierarchical relationship involving multi-potent mammary stem cells (MaSCs) has been controversial. The capacity for basal cells to form reconstituted glands in transplantation assays has supported the notion that multi-potent MaSCs reside in the basal fraction of the adult mammary gland (Shackleton et al., 2006; Stingl et al., 2006). However, lineage-tracing studies have produced conflicting results as to whether adult basal cells are multi-potent or unipotent (Davis et al., 2016; Giraddi et al., 2015; Rios et al., 2014; Van Keymeulen et al., 2011; Wang et al., 2015; Wuidart et al., 2016). On the other hand, studies have supported the conclusion that bipotent MaSCs are present in the fetus (i.e., fMaSCs) (Makarem et al., 2013; Spike et al., 2012; Van Keymeulen et al., 2011). Because lineage-tracing experiments measure cell fate only in the context of a native structure, and functional assays only measure the developmental potential of cells under non-native conditions, the development of agnostic molecular approaches may better predict the differentiation potential of mammary cells and enable identification of differentiation state regulators.

Cancer has been referred to as a caricature of normal development and of tissue renewal (see Wahl and Spike, 2017 for references). Therefore, to ascertain development correlates of breast cancer, we designed this study with three goals: (1) to generate epigenetic and transcriptomic maps of the developmentally plastic, bipotent fetal mammary stem cells (fMaSCs) and the adult lineages descended from them; (2) to identify genes, transcriptional regulators, and control regions associated with fMaSC bipotentiality that are altered upon adult lineage specification; and (3) to test whether such regulators are altered during cancer progression and contribute to the genesis of intra-tumoral heterogeneity.

RESULTS

Chromatin Features Indicative of Multi-lineage Potential Are Exclusively Present in Fetal and Adult Basal Mammary Cells

We first performed assay for transposase-accessible chromatin using sequencing (ATAC-seq) and RNA sequencing (RNA-seq) on biological replicates of mammary cell populations enriched for E18 fMaSCs, adult basal cells, luminal progenitor cells (LPs), and mature luminal cells (MLs) using fluorescence-activated cell sorting-based purification with previously established cell surface markers and corresponding phenotypic characterization (Figure 1A and Table S1) (Asselin-Labat et al., 2007; Makarem et al., 2013). ATAC-seq maps chromatin accessibility and indicates the potential of a flanking gene to be expressed. By contrast, RNA-seq maps transcript levels, and hence correlates more directly with cellular phenotype at the time of analysis.

ATAC-seq data were highly reproducible between biological replicates and showed clear enrichment at specific genomic regions (Figures 1B and S1A). The two replicates of each cell type were therefore combined for all downstream analyses to improve signal strength. The open chromatin regions indicated by the ATAC-seq signal also correlated with the active transcription marks H3K27ac and H3K4me3, but not with the repressive transcription mark H3K27me3 (Figures S1B and S1C). Similarly, there was significant correlation between chromatin accessibility and transcript level in the same cell types (Figure S1D). This is expected, as expressed genes generally require accessible chromatin for transcription.

Analysis of ATAC-seq profiles revealed that fMaSCs and adult basal cells possessed more open chromatin regions than either LPs or MLs (Figure 1C). Distal chromatin elements are associated with cell type specification during development (Shlyueva et al., 2014). Consistent with this, the majority of the increased chromatin accessibility in fMaSCs and adult basal cells was found in distal regions (Figure 1C). By measuring Shannon entropy, an indicator of cell specificity, we found that chromatin accessibility of distal regions was more cell type-specific than promoter accessibility (Figure S1E). To determine whether cell type-specific distal regions correlated with gene expression, we identified RNA-seq gene expression signatures for each cell type and evaluated activation levels of distal regions in these genes (Table S1). Notably, cell signature gene expression correlated well with ATAC-seq and H3K27ac signal (Figure S1F). Together, these observations indicate that distal region accessibility is linked to the expression of genes contributing to mammary cell identity.

We compared transcript levels of known lineage indicator genes in fMaSCs, basal cells, MLs, and LPs. As expected, adult basal cells showed elevated expression of genes previously associated with the basal lineage, LPs showed elevated expression of LP-associated genes, and MLs showed elevated

(C) Total numbers of ATAC-seq peaks in mammary cells, separated into promoter (<±3 kb transcription start site [TSS]) and distal regions (>±3 kb TSS). ***$p < 0.001$ (LP and ML versus both fMaSC and basal).

(D) RNA-seq expression of basal and luminal genes across mammary subpopulations, from two averaged biological replicates. The thick horizontal middle line is the median; height of the box is the interquartile range (IQR); dotted vertical line is 1.5× IQR; dots are the outliers.

(E) ATAC-seq and RNA-seq of basal and luminal genes across mammary cell subpopulations. Mean ± SEM (n = 2).

See also Figures S1 and S2; Table S1.

expression of ML-associated genes (Figure S1G). Pan-luminal markers such as *Krt8*, *Krt18*, and *Epcam* were expressed at significantly higher levels in LPs and MLs. Similarly, fMaSCs showed elevated expression of embryo-associated transcripts such as *Sox11* and *Hmga2*. Of note, fMaSCs also exhibited intermediate expression of both basal- and luminal-associated genes (Figure 1D). This is consistent with the notion that these primitive embryonic mammary cells exist in a differentiation-ambivalent state. Thus, the RNA-seq data affirm that adult mammary cells are phenotypically distinct and lineage restricted, whereas fMaSCs exhibit characteristics expected of cells in a developmentally plastic, multi-lineage state.

Consistent with the gene expression data, the ATAC-seq data show that fMaSCs exhibited open chromatin features at distal regions and promoters of both luminal and basal genes (Figures 1E and S2A). While basal cells exhibited open chromatin in regions associated with highly expressed basal genes, they unexpectedly manifested open chromatin features at putative regulatory elements associated with luminal genes that were expressed at low or undetectable levels (boxed regions, Figures 1E and S2A). Such bilineage open chromatin features of fMaSCs and basal cells were also observed systematically at the promoters of ML and basal associated genes (Figure S2B). Thus, the chromatin features in these cells indicate that fMaSCs and basal cells have the potential to express genes associated with specifying either basal or luminal cell identities.

For luminal cell types (LPs and MLs), open chromatin features were found only near luminal genes (boxed regions, Figures 1E and S2A). Basal genes were associated with closed chromatin features, and basal genes were expressed at levels close to background or not expressed (Figures 1E and S2A). Interestingly, LP-specific genes such as *Elf5* and *Kit* were accessible and expressed in LPs but not in MLs, whereas multiple ML-specific genes such as *FoxA1* and *Esr1* were accessible in both LPs and MLs. Thus, LPs have chromatin features indicating that they have the potential to express ML-associated genes.

Collectively, our transcriptome and chromatin analyses revealed (1) molecular correlates of the multi-lineage potential of fMaSCs and adult basal cells, (2) the predicted progenitor characteristics consistent with the ability of LPs to differentiate into ER-positive and ER-negative subtypes, and (3) a more restricted developmental potential of MLs.

Cell Type-Specific Chromatin Features Associate *Sox10* with the Mammary Stem Cell State

We used chromatin accessibility analyses to identify candidate transcriptional regulators of cell state changes that occur during mammary development. We identified chromatin regions that are uniquely open or closed in each of the four mammary cell populations (uniquely accessible region [UAR] and uniquely repressed region [URR], respectively) (Figures 2A, S3A, and S3B). The UARs and URRs represented regions with very low Shannon entropy, suggesting they are cell type-specific chromatin features; they also correlated strongly with histone H3K27 acetylation, an activation mark (Figures 2B, 2C, and S3C). The majority of UARs and URRs were located at distal regions of genes (Figure 2D), consistent with previous studies demonstrating the importance of distal elements in cell identity (Shlyueva et al., 2014). Interestingly, while the adult UARs and URRs correlate with cell type-specific chromatin activation and repression, respectively, as determined by previous chromatin immunoprecipitation coupled with next generation sequencing (ChIP-seq) data on the adult mammary populations (Pal et al., 2013), the fMaSC UARs and URRs do not exhibit such cell type specificity in the adult populations (Figure S3D). These comparisons both validate the quality of our data and demonstrate that the fMaSC-specific chromatin regions identified through our analyses are unique.

The Genomic Regions Enrichment of Annotations Tool (GREAT) enabled identification of genes likely controlled by these UARs/URRs in each cell type (Figure 2E and Table S2). The presence of basal-specific, LP-specific, and ML-specific genes in the UARs of the corresponding cell types suggests the relevance of these unique chromatin regions to regulating these genes in these cell types. In parallel, we performed GREAT analysis to identify genes controlled by active enhancer regions specific to human basal, LP, or ML mammary cells, using published subpopulation-specific ChIP-seq analyses (Pellacani et al., 2016), and found high levels of similarity to mouse epigenetic features (Figures S3E and S3F).

We next identified transcription factor (TF) motifs within the UARs and URRs. Homer revealed expected enrichment of the P63 and TEAD4 DNA-binding motifs in basal cells, the ELF5 DNA-binding motif in LPs, and the FOXA1 and Jun-AP1 DNA-binding motifs in MLs (Figure 2F). These TF DNA-binding motifs have also been mapped to uniquely active enhancers of the analogous populations of human mammary cells (Pellacani et al., 2016). Notably, binding motifs for SOX4, SOX9, SOX10, and NF1 were significantly enriched in fMaSC UARs compared with the other cell types (Figure 2F). Moreover, many regions containing these SOX motifs were specifically closed in MLs (enriched in ML URRs), which are the most differentiated (least developmentally plastic) of the four mammary cell types (Figure 2F). Of these SOX factors, SOX10 uniquely exerts potent cell reprogramming capacities *in vitro* (Kim et al., 2014) and is most differentially upregulated in the developmentally plastic fMaSC population (Figure 2G). Furthermore, in fMaSCs, UARs adjacent to highly expressed genes contained more SOX10 binding motifs than UARs adjacent to genes expressed at lower levels (a trend that was not observed with other TFs, such as NF1, P63, and FRA1). These data suggest an association between SOX10 binding and chromatin activation in fMaSCs. Collectively, these data associate SOX factors (SOX10 in particular) with the developmental plasticity and bipotentiality of fMaSCs.

SOX10 Is Expressed in Mammary Tumors from Mouse Models and Human Patients

The data reported above provide insight into potential factors that may be involved in transitions between the uncommitted state of fMaSCs and the differentiated states of adult cells. We determined whether levels of SOX10 correlate with changes in cell state during tumor progression using three mouse mammary tumor models. These models were chosen because of their transcriptomic relatedness to different intrinsic subtypes of human breast cancer (Pfefferle et al., 2013). We also used or developed *Sox10* reporters to enable the visualization and recovery of $SOX10^{high}$ and $SOX10^{low}$ tumor cells to evaluate correlations between SOX10 levels and gene expression changes. Finally, we

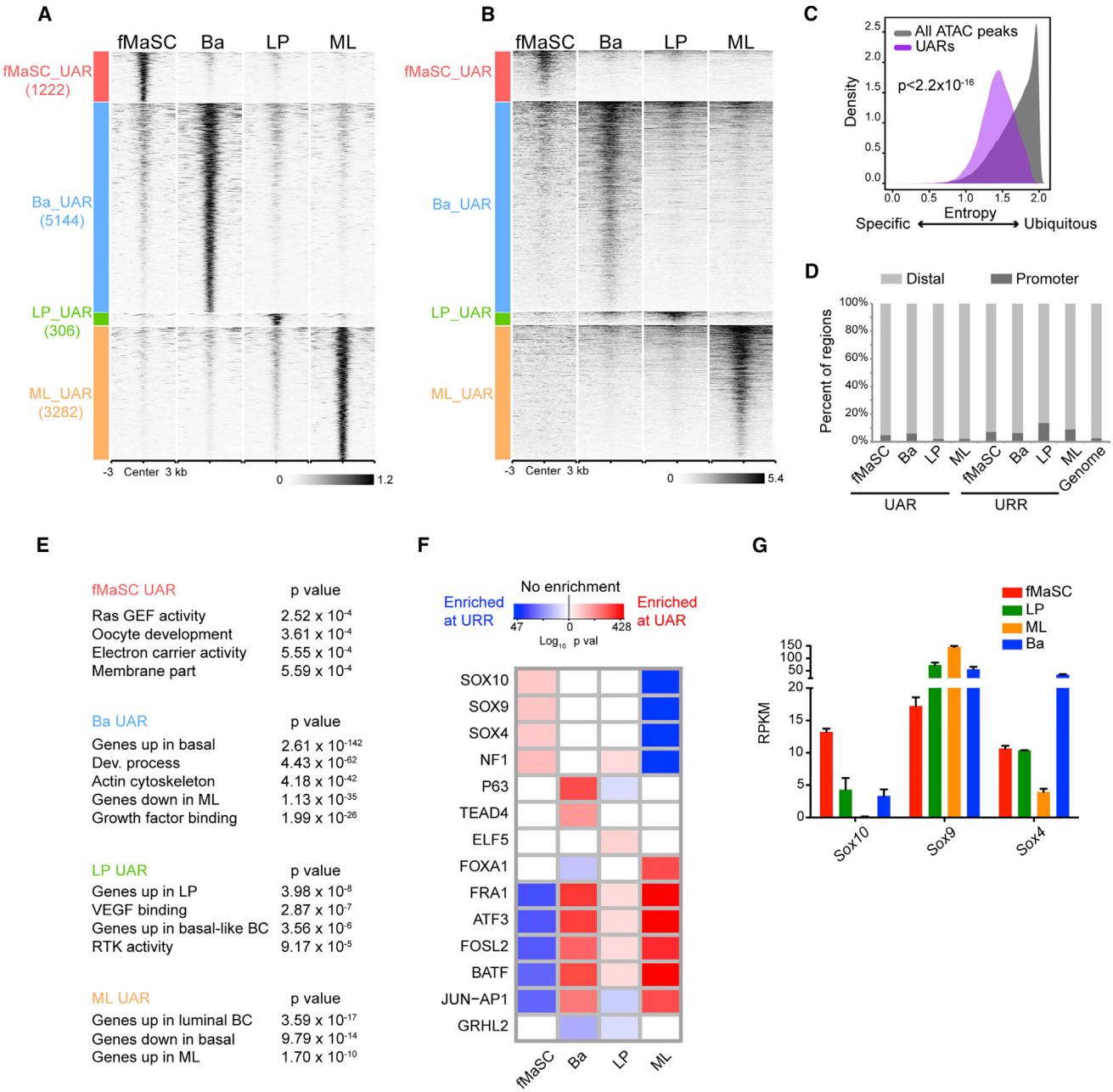


Figure 2. Chromatin Features Associate SOX10 with the Mammary Stem Cell State

(A and B) ATAC-seq signal at UARs (A) and corresponding H3K27ac signal (B) specific to the indicated mammary cell type; each row represents a specific genomic locus.

(C) Shannon entropy of UARs versus all ATAC-seq peaks.

(D) Percentage of UARs and URRs located at distal (>±3kb TSS) or promoter (<±3kb TSS) regions.

(E) GREAT analysis of genes associated with cell type-specific UARs.

(F) Enrichment of transcription factor motifs at UAR/URR across mammary cell subpopulations.

(G) Transcript level of Sox factors from (F). Mean ± SEM (n = 2).

See also Figure S3 and Table S2.

used CRISPR-based genome engineering to attach a biotin acceptor domain to the C terminus of SOX10 to perform highly specific ChIP studies to identify genes directly regulated by SOX10 (Figure 3A).

We used orthotopic transplantation of PY230 cells derived from an MMTV-PyMT tumor as a model for luminal-like breast cancers (Bao et al., 2015a; Pfefferle et al., 2013). We modified the *Sox10* locus in PY230 cells to express both a bright-red

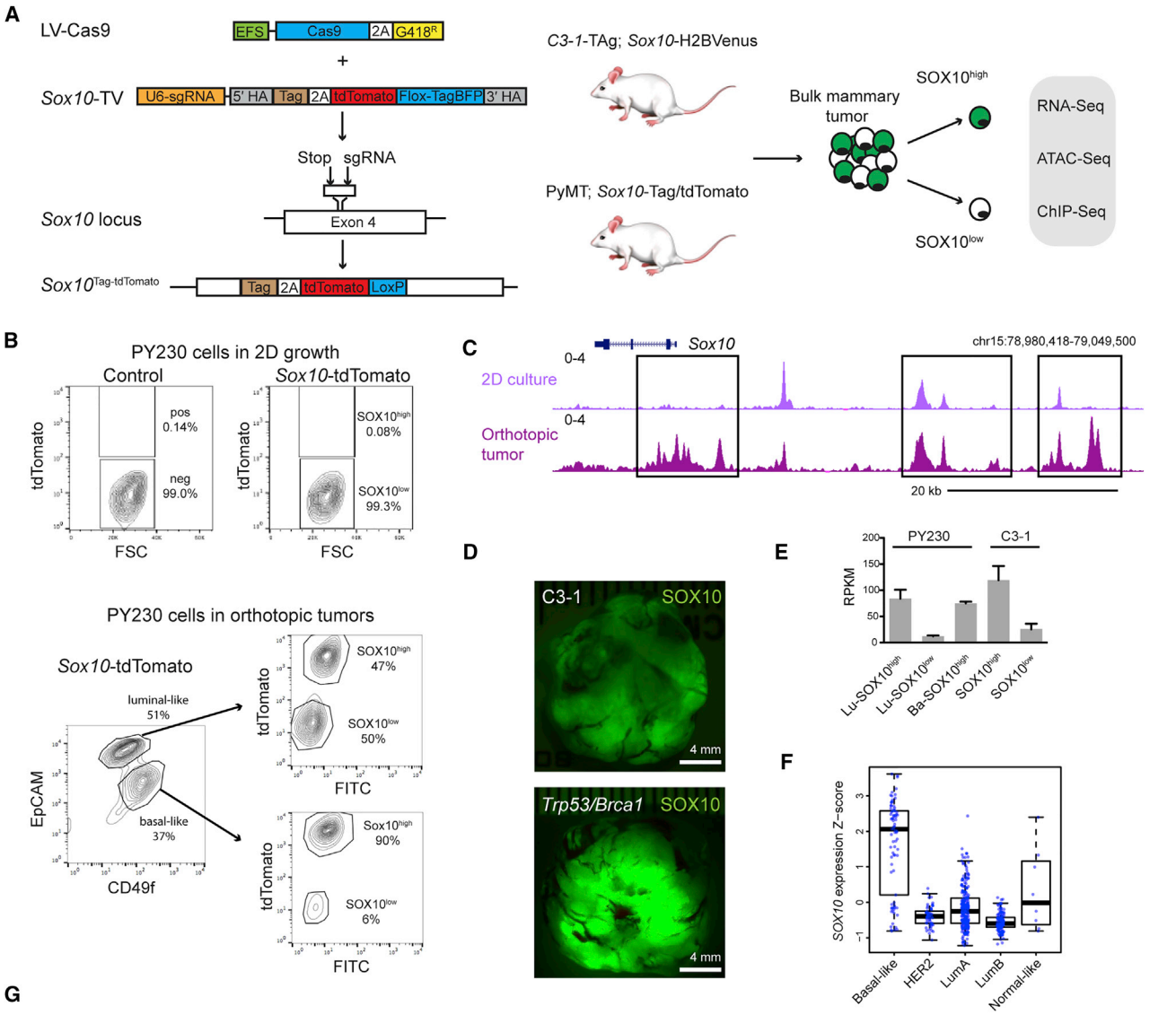


Figure 3. SOX10 Is Expressed in Mammary Tumors

(A) Strategy to modify the *Sox10* locus and characterize SOX10^{high} and SOX10^{low} tumor cells.

(B) tdTomato fluorescence (y axis) from control and *Sox10*^{tdTomato} tumor cells grown in 2D (top) and sorting strategy to isolate luminal-like and basal-like mammary tumor cells and evaluate tdTomato fluorescence (bottom).

(C) ATAC-seq of the *Sox10* locus in PY230 tumor cells grown in 2D or from orthotopic tumors.

(D) Whole-mount view of *Sox10* expression in C3-1 and *Trp53*;*Brca1* mammary tumors with a *Sox10*-H2BVenus reporter.

(E) *Sox10* transcript levels in tumor cells sorted by *Sox10* fluorescent reporter signal. Mean ± SEM (PY230, n = 2; C3-1 SOX10^{high}, n = 10; C3-1 SOX10^{low}, n = 4).

(legend continued on next page)

nuclear-localized fluorescent protein and a C-terminal biotinylatable epitope tag. Importantly, while PY230 cells did not express SOX10 when grown in two-dimensional (2D) culture on plastic, orthotopic transplantation generated tumors exhibiting robust but heterogeneous SOX10 expression (~>90% of basal-like tumor cells, and ~45%–70% of luminal-like tumor cells were SOX10$^+$) (Figure 3B). ATAC-seq profiling showed that the *Sox10* locus was inaccessible in cancer cells grown in 2D but was accessible in orthotopic tumors (Figure 3C). These data highlight the importance of the *in vivo* model for producing key contextual and molecular cues that can be lost in cell culture.

The C3-1 mouse is a transgenic mouse line that expresses the SV40 large T-antigen in mammary epithelia and develops tumors exhibiting features associated with human basal-like and claudin-low triple-negative breast cancers (Pfefferle et al., 2013). C3-1 animals were crossed with a mouse line containing a *Sox10*-H2BVenus BAC-transgene reporter to enable visualization of *Sox10* expression. Mammary tumors formed within 8 months of age, and fluorescence revealed robust (Figure 3D) and heterogeneous (Figure S4A) *Sox10* expression within these tumors. RNA-seq of SOX10^{high} and SOX10^{low} cells confirmed concordance between reporter signal and *Sox10* expression in the PY230 and C3-1 models (Figure 3E and Table S3).

The third model utilized *Trp53*^{flox}*Brca1*^{flox} mice in which mammary tumors were initiated by intra-ductal nipple injections of AAV-Cre, or by orthotopic transplantation of *Trp53*^{flox}*Brca1*^{flox} fMaSCs infected with lentivirus expressing Cre. P53 and BRCA1 inactivation are most frequently found in human basal-like triple-negative breast cancers (Turner et al., 2004). These mice were bred to the same BAC-transgenic *Sox10*-H2BVenus reporter mouse line described above to enable *in situ* detection of *Sox10* expression. Mammary tumors formed within 12 months, and fluorescence again revealed robust *Sox10* expression (Figure 3D).

Using the Cancer Genome Atlas (TCGA) database, we found that *SOX10* is expressed at the highest levels in the basal-like subtype of breast cancers (Figures 3F and S4B), which is consistent with our prior analyses using Metabric and UNC885 databases (Dravis et al., 2015). SOX10 protein expression in human basal-like breast cancers has been previously reported (Cimino-Mathews et al., 2013), and we were also able to visualize SOX10$^+$ cells in malignant tissue isolated from a human patient with ER$^-$PR$^-$HER2$^-$ breast cancer, but not in the adjacent benign tissue (Figures 3G and S4C).

Collectively, these data demonstrate that *SOX10* is expressed in two predominantly basal-like mouse breast cancer models, a mouse luminal-like mammary tumor model, and a subset of human breast cancers.

SOX10^{high} Tumor Cells Exhibit Mammary Stem/Progenitor Cell Features

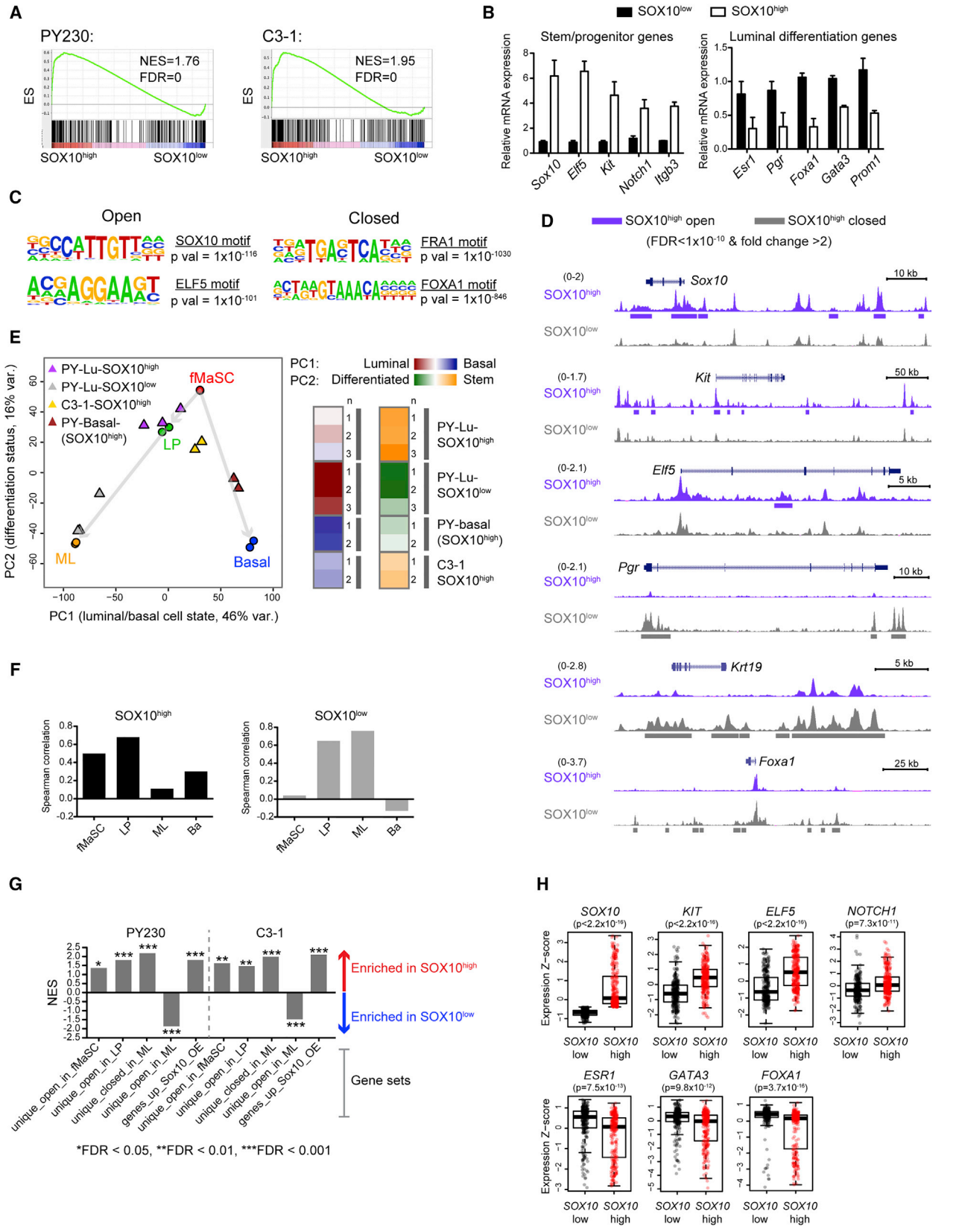
We then determined whether tumor cells expressing high SOX10 exhibit characteristics consistent with stem/progenitor cells. We used RNA-seq analysis to compare SOX10^{high} and SOX10^{low} luminal-like fractions from PY230 orthotopic and C3-1 autochthonous tumors. We could not perform this analysis for basal-like PY230 tumor cells, as the small subpopulation of tdTomato-negative tumor cells still expressed SOX10. We built transcriptome profiles of SOX10^{high} and SOX10^{low} luminal-like cell populations, and ascertained stem/progenitor relatedness using gene set enrichment analysis (GSEA) with signature gene lists representing stem/progenitor populations from the normal mammary gland (Table S4). These analyses revealed significant enrichment of fMaSC and LP signature genes for SOX10^{high} cells from both tumor models (Figures 4A and S5A). SOX10^{high} tumor cells also expressed higher levels of fMaSC- and LP-related genes compared with SOX10^{low} tumor cells (Figure 4B). Conversely, SOX10^{high} tumor cells exhibited reduced expression of genes associated with more differentiated MLs. These data indicate that SOX10^{high} tumor cells possess the stem/progenitor-related developmental plasticity associated with fMaSCs and LPs.

We determined whether the epigenetic profile of SOX10^{high} tumor cells also reflects a stem/progenitor identity. We first identified chromatin peaks that were differentially open or closed between SOX10^{high} and SOX10^{low} luminal PY230 cells (fold change >2; false discovery rate [FDR] < 1×10^{-10}). As expected, open peaks in SOX10^{high} tumor cells were significantly enriched for the SOX10 motif (Figure 4C). Interestingly, the ELF5 motif, which is associated with LPs, was also significantly open in SOX10^{high} cells. By contrast, FOXA1 and FRA1 motifs, which are important for adult differentiated cells (Figure 2F), were significantly closed in SOX10^{high} cells. GREAT analyses also showed that SOX10^{high} open peaks were associated with the LP gene signature, whereas SOX10^{high} closed peaks were associated with the ML signature (Figure S5B).

Analysis of chromatin accessibility at fMaSC- and LP-specific gene loci also revealed a strong correlation between *Sox10* expression and stem-like chromatin features. As expected, chromatin accessibility at the *Sox10* locus in SOX10^{high} tumor cells closely resembled that of fMaSCs and LPs, but not MLs (Figure S5C). Stem/progenitor-associated genes including *Kit* and *Elf5*, and fetal-specific genes such as *Sox11* and *Hmga2* also exhibited open chromatin in SOX10^{high} tumor cells. These differences were also apparent when comparing SOX10^{high} and SOX10^{low} tumor cells, as SOX10^{high} tumor cells had more open chromatin at fMaSC/LP-associated gene loci, and less open chromatin at the loci of ML-associated genes (Figure 4D).

Principal component analysis (PCA) enabled visualization of global changes in chromatin accessibility in each cell population. To best separate cell types, we used the UARs and URRs at which differences in chromatin accessibility best correlated with unique cell identity. This analysis revealed separation of normal cell types based on principal component 1 (PC1) and PC2, which we interpreted as scores for luminal-basal (PC1) and cell differentiation (PC2) chromatin states. These PCs display fMaSCs and LPs at the top of a differentiation trajectory,

(F) *SOX10* expression in human breast cancers from 2012 TCGA (n = 508). The thick horizontal middle line is the median; height of the box is the IQR; dotted vertical line is 1.5× IQR; dots are the outliers.
(G) Tissue section from ER$^-$PR$^-$HER2$^-$ breast tumor immunostained for SOX10, K8, and K14.
See also Figure S4 and Table S3.



(legend on next page)

and intermediate between luminal and basal states (Figure 4E). Notably, projecting the tumor cells onto PC1:luminal-basal and PC2:stem-differentiated dimensions indicates that the PY230 and C3-1 SOX10^{high} tumor cells localize to an intermediate differentiation state between luminal and basal. This is consistent with the interpretation that these tumor cell populations possess chromatin states resembling the mixed basal-luminal features of stem/progenitor-like fMaSCs and LPs.

We also used Spearman correlation to compare chromatin accessibility in SOX10^{high} and SOX10^{low} tumor cell populations with the UARs and URRs found in normal mammary cells. The chromatin accessibility of the SOX10^{high} PY230 tumor cells correlated significantly with unique chromatin features in LPs and fMaSCs (and, to a smaller extent, basal cells) (Figure 4F). On the other hand, SOX10l^{ow} PY230 tumor cells correlated strongly with LP and ML UARs. Thus, while the chromatin features of SOX10^{high} PY230 tumor cells correlate better with stem/progenitor populations than do SOX10^{low} tumor cells, both of these cell types possess blended chromatin features that are not apparent in the normal adult mammary cells from which they are derived.

We determined whether chromatin accessibility reflects gene expression using GSEA of the transcriptomes of SOX10^{high} and SOX10^{low} tumor cells. We used gene sets associated with UARs and URRs for fMaSCs, basal cells, LPs, and MLs. These analyses revealed that SOX10^{high} cells upregulated genes uniquely open in fMaSCs and LPs, and genes uniquely closed in MLs (Figure 4G). By contrast, SOX10^{low} tumor cells upregulated genes uniquely open in MLs. Thus, stem/progenitor identity, as indicated by chromatin accessibility in SOX10^{high} versus SOX10^{low} cells, correlated strongly with the transcriptome profiles of these cells. We infer that SOX10 contributes to the observed stem/progenitor identity, as there is significant enrichment in the SOX10^{high} tumor fraction for genes we previously showed are upregulated following SOX10 overexpression in an *in vitro* organoid culture model (Figure 4G) (Dravis et al., 2015).

Finally, we determined whether the associations with *Sox10* expression and stem/progenitor identity in mammary tumors could be extrapolated to human breast cancer. We used TCGA data to evaluate the expression of stem/progenitor-associated genes in SOX10^{high} versus SOX10^{low} breast cancers. These analyses revealed that SOX10^{high} tumors tend to express higher levels of stem/progenitor-associated genes and lower levels of ML-associated genes compared with SOX10^{low} tumors (Figures 4H and S5D).

Collectively, these data reveal that SOX10^{high} tumor cells exhibit chromatin and transcriptome features expected of stem/progenitor cells.

SOX10^{high} Cells within Mammary Tumors Exhibit Dedifferentiated and EMT-like Features

Ectopic overexpression of SOX10 reprograms mammary epithelial cells into a mesenchymal-like cell state (Dravis et al., 2015). Strikingly, analysis of sections from C3-1 mammary tumors revealed that SOX10^{high} cells expressed low levels of epithelial cytokeratins, whereas cells with lower SOX10 expression retained epithelial markers (Figures 5A and 5B). To better quantify the relationship between SOX10 and epithelial markers, we dissociated C3-1 mammary tumors to single cells, and found that >80%–90% of SOX10^{high} cells had undetectable levels of KRT8 and KRT14 (Figures 5C and 5D). SOX10^{high} tumor cells form tumorspheres in 3D culture conditions at low efficiency, and these tumorspheres exhibited high levels of SOX10 and low levels of cytokeratins (Figure 5E). Thus SOX10^{high} cells in these basal-like mammary tumors showed reduced levels of keratin markers associated with the epithelial state and mammary cell differentiation.

We analyzed the PY230 mammary tumor model to determine the generality of the relationship between *Sox10* expression and loss of epithelial features. Notably, SOX10^{high} PY230 mammary tumor cells also exhibited significant decreases in multiple epithelial and luminal mammary cell markers compared with SOX10^{low} mammary tumor cells (Figure 5F). SOX10^{high} cells also had increased expression of the mesenchymal/EMT markers *Vim*, *Snai2*, and *Twist1*. As ectopic expression of SOX10 in normal mammary cells also elicits dedifferentiation and mesenchymal-like features with similar corresponding gene expression changes (Dravis et al., 2015), we infer that SOX10 directly contributes to this cell state change.

Because SOX10 overexpression can also induce motility and mammary cell delamination in 3D culture (Dravis et al., 2015), we determined whether SOX10^{high} cells also locally invade in mouse tumor models *in vivo*. Strikingly, significant numbers of SOX10^{high} cells in PY230 tumors were found outside the primary tumor margin and in close proximity to tumor vasculature (Figure 5G and Video S1).

We examined the TCGA database to determine whether SOX10 is similarly linked to EMT and dedifferentiation in human breast cancer. We generated a rank-order list of human genes based on the correlation of their expression with *SOX10* expression across a panel of human breast tumors. Many EMT-related genes positively correlated with *SOX10* expression, whereas many epithelial/differentiation-related genes negatively correlated with *SOX10* expression (Figures 5H and S5E). Consistent with these data, SOX10 expression in malignant tissue correlated with undetectable or low expression of K14 and K8, compared with adjacent benign mammary tissue on the section,

Figure 4. SOX10^{+} Tumor Cells Exhibit Mammary Stem/Progenitor Features

(A) GSEA of fMaSC genes in SOX10^{high} versus SOX10^{low} tumor cells.

(B) Relative expression of LP and ML genes in PY230 tumor cells. Mean ± SEM (n = 2).

(C) TF motifs enriched in chromatin regions differentially accessible between SOX10^{high} versus SOX10^{low} PY230 tumor cells.

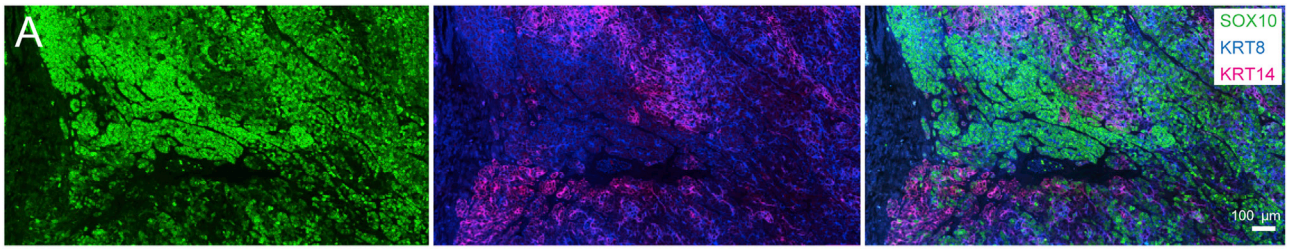
(D) ATAC-seq of PY230 tumor cells at stem/progenitor- and ML-associated loci.

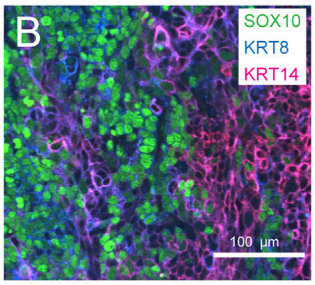
(E) PCA of normal and tumor mammary cell types using ATAC-seq signal (left) and the interpretation of projected tumor PC scores shown as heatmaps (right).

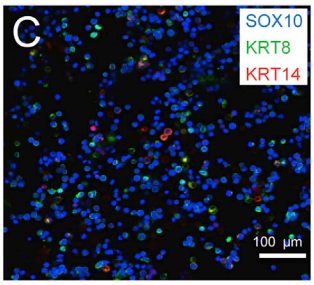
(F) Correlation of chromatin accessibility in PY230 tumor cells with UARs and URRs in normal mammary cells.

(G) GSEA of UAR- or URR-associated gene sets, and genes upregulated following *Sox10*^{OE} in SOX10^{high} versus SOX10^{low} tumor cells.

(H) Expression levels of stem/progenitor or ML genes in *SOX10*^{high} (upper 50%) and *SOX10*^{low} (lower 50%) human breast tumors, taken from RNA-seq of 2012 TCGA breast tumors (n = 528). The thick horizontal middle line is the median; height of the box is the IQR; dotted vertical line is 1.5× IQR; dots are the outliers. See also Figure S5 and Table S4.

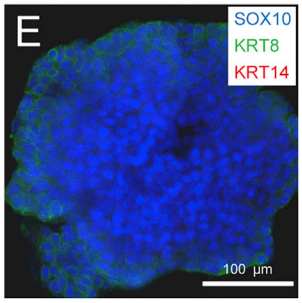


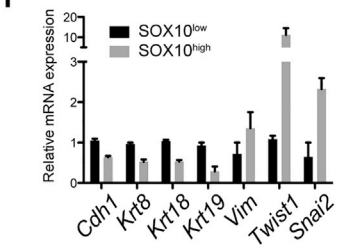


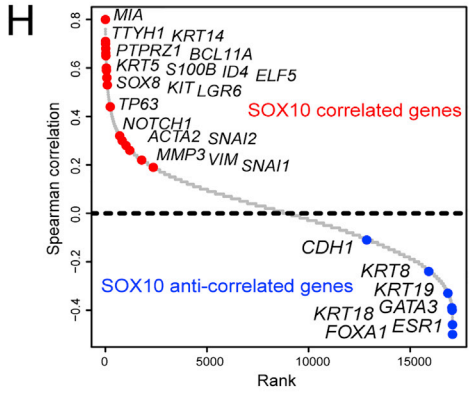


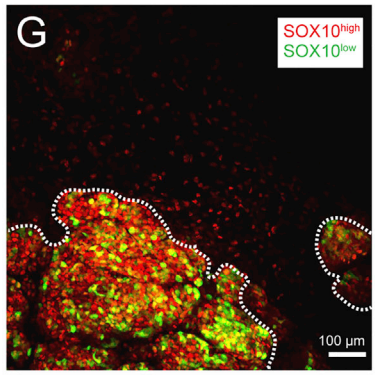
D

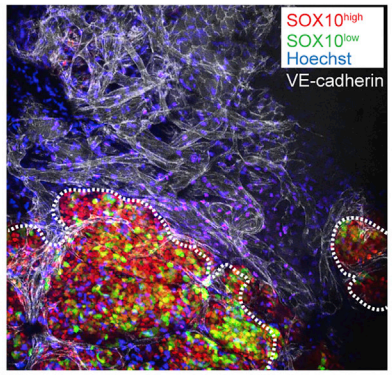
	Total cell count	SOX10^{high} K8^{+}	SOX10^{high} K14^{+}	SOX10^{high} K8^{-} K14^{-}
Tumor1	284	5.1% ± 2.1	6.0% ± 0.4	88.9% ± 2.5
Tumor2	886	9.2% ± 0.6	2.2% ± 2.0	88.6% ± 1.4
Tumor3	555	9.3% ± 4.3	7.1% ± 0.6	83.6% ± 3.7
Tumor4	229	5.7% ± 9.0	1.4% ± 2.7	92.9% ± 11.7

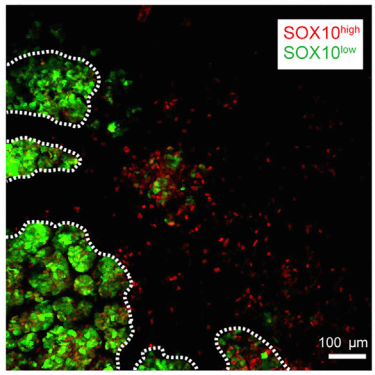


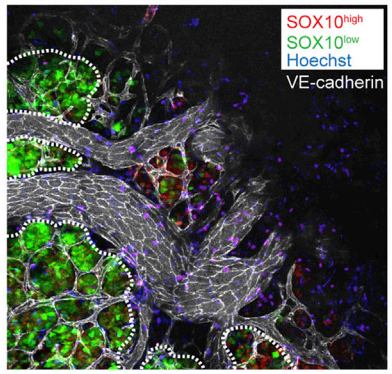


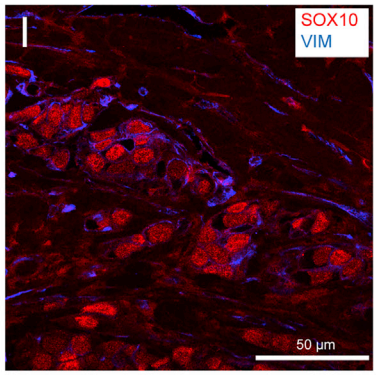


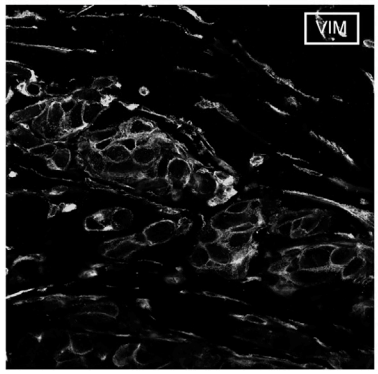












(legend on next page)

in the same human ER$^-$PR$^-$HER2$^-$ breast cancer sample (Figures 3G and S4C). Clear expression of the mesenchymal marker VIM could also be detected in many of the SOX10$^+$ tumor cells (Figure 5I).

Taken together, these data establish a link between SOX10 and partial EMT/dedifferentiation in breast cancer. The data further indicate links between SOX10 and local invasion and metastasis, critical features of cancer-associated mortality.

Elevated *Sox10* Expression Correlates with Neural Crest-like Features

SOX10 is a known specifier of neural crest cell (NCC) identity during embryonic development (Southard-Smith et al., 1998). Indeed, ectopic SOX10 expression, when combined with extrinsic factors, reprograms fibroblasts into an NCC-like state (Kim et al., 2014). Because *Sox10* appears to be highly expressed in mammary tumors, operates in a dysregulated microenvironment, and induces motility, partial EMT, and multi-lineage characteristics also present in NCC cells, we determined whether the phenotypes observed in SOX10^{high} mammary tumor cells reflect its ability to reprogram them into an NCC-like state.

We performed GSEA on SOX10^{high} versus SOX10^{low} tumor cells from PY230 tumors to ascertain enrichment for NCC gene sets (Table S5). These analyses showed enrichment for NCC-related genes in SOX10^{high} PY230 tumor cells (Figure 6A). Enrichment of the NCC gene list was also seen in SOX10^{high} C3-1 tumor cells (though below the significance threshold of FDR < 0.05). Transcript levels of critical NCC-specifying genes, including *Sox10*, *Sox8*, *Sox5*, *Twist1*, *Lmo4*, *Etv5*, and *Tfap2c*, were significantly higher in SOX10^{high} cells from both tumor models, whereas *Pax3*, *Dlx1*, *Id2*, *Prdm1*, and *Snai2* were enriched solely in PY230 tumors (Figures 6B and S6A).

We used ATAC-seq profiling of SOX10^{high} versus SOX10^{low} PY230 tumor cells to determine the chromatin accessibility at NCC-related genes. Consistent with the transcript data, we found that many key NCC-specifier genes showed more accessible chromatin in SOX10^{high} mammary tumor cells compared with normal mammary cells (Figures 6C and S6B). Moreover, NCC-related genes were significantly represented (41% of NCC genes) in the 3,563 loci exhibiting more accessible chromatin peaks in SOX10^{high} tumor cells (Figure 6D).

Finally, we determined whether human breast cancers also exhibit features of NCC reprogramming by performing GSEA with gene sets obtained from both differentiating and migrating NCCs. Both NCC-associated gene sets showed significant enrichment with *SOX10* correlated genes in human breast tumors (Figure 6E). Interestingly, co-expression network analysis using TCGA gene expression data (528 tumors) indicated that *SOX10* forms an interconnected network with genes such as *SOX8*, *FOXC1*, *SFRP1*, *WNT10A*, *ZEB2*, *SNAI2*, *TWIST1*, *NRP1*, and *EDN1*, suggesting that these genes might form a core regulatory network of NCC-like gene expression in breast tumors (Figure S6C).

These data indicate that mouse and potentially human SOX10^{high} mammary tumors exhibit molecular features of NCC. This suggests that SOX10 reprises a developmental role in reprogramming tumor cells to adopt NCC-like features.

Identification of Genes Bound by SOX10 Indicates a Direct Role in Cell State Interconversion that Promotes Tumor Development

The data presented above are consistent with the proposal that SOX10 directly contributes to cell state reprogramming in transformed mammary cells. We investigated this possibility more directly by identifying the direct transcriptional targets of SOX10 using ChIP-seq. As mentioned above, in *Sox10*^{tdTomato} PY230 cells a biotin acceptor domain is fused to the C terminus of the endogenous SOX10 protein. This tagged SOX10 was biotinylated by infecting these cells with a lentivirus expressing biotin ligase. We then used streptavidin to isolate SOX10 from these cells, using PY230 cells encoding unmodified SOX10 and infected with the same lentiviral biotin ligase as a negative control.

SOX10 ChIP-seq analysis of two independent PY230 tumors showed clear, sharp peaks compared with the control, and two biological replicates generated highly correlated data (Figures 7A and S7A). We combined the two biological replicates to increase the reliability of all downstream analyses. Using a stringent peak-calling cutoff (FDR < 1×10^{-100}), we identified 7,929 SOX10-binding peaks in PY230 cells, none of which were detected in the control (Figure 7B). The most of the SOX10-binding sites localized to sites distal to coding regions (Figure S7B) and were highly enriched for open chromatin (Figure S7C). As expected, motif analysis showed that SOX10 peaks were significantly enriched for the SOX10-binding motif (Figure S7D). We also observed enrichment of ATF3, ELF5, and NF1 motifs, suggesting that SOX10 may cooperate with these factors in regulating mammary cell states (Figure S7D).

To investigate whether SOX10 is associated with cell type-specific chromatin features, we examined SOX10 binding at UARs. Strikingly, SOX10 binding was significantly enriched at fMaSC and LP UARs, while showing very little binding at basal cell and ML UARs (Figures 7C and S7E). We next used the Binding and Expression Target Analysis (BETA) program to integrate

Figure 5. SOX10$^+$ Tumor Cells Exhibit Dedifferentiation and Mesenchymal Features

(A and B) Low-magnification (A) and high-magnification (B) images of C3-1;*Sox10*-H2BVenus mammary tumors immunostained for K8, K14, and GFP (SOX10).

(C) Single-cell dissociation of a C3-1;*Sox10*-H2BVenus mammary tumor immunostained for K8 and K14.

(D) Quantification of keratin status in four C3-1;*Sox10*-H2BVenus mammary tumors. Average percentage and 95% confidence interval from two images for each tumor are shown.

(E) Tumorsphere grown from C3-1;*Sox10*-H2BVenus mammary cells plated in 3D culture, immunostained for K8 and K14.

(F) Relative expression of differentiation and mesenchymal genes in PY230 tumor cells. Mean ± SEM (n = 2).

(G) PY230 *Sox10*^{tdTomato} tumor (outlined with dots) showing SOX10$^+$ cells (red) in the primary tumor margin and near vasculature (white). PY230 *Sox10*^{tdTomato} tumor cells were labeled with an LV-GFP to visualize tumor cells not expressing SOX10 (green).

(H) Rank-order list of *SOX10* co-expression genes in human breast tumors with epithelial (blue) and EMT-associated (red) genes highlighted.

(I) Tissue section from an ER$^-$PR$^-$HER2$^-$ human breast tumor immunostained for SOX10 and VIM.

See also Figure S5 and Video S1.

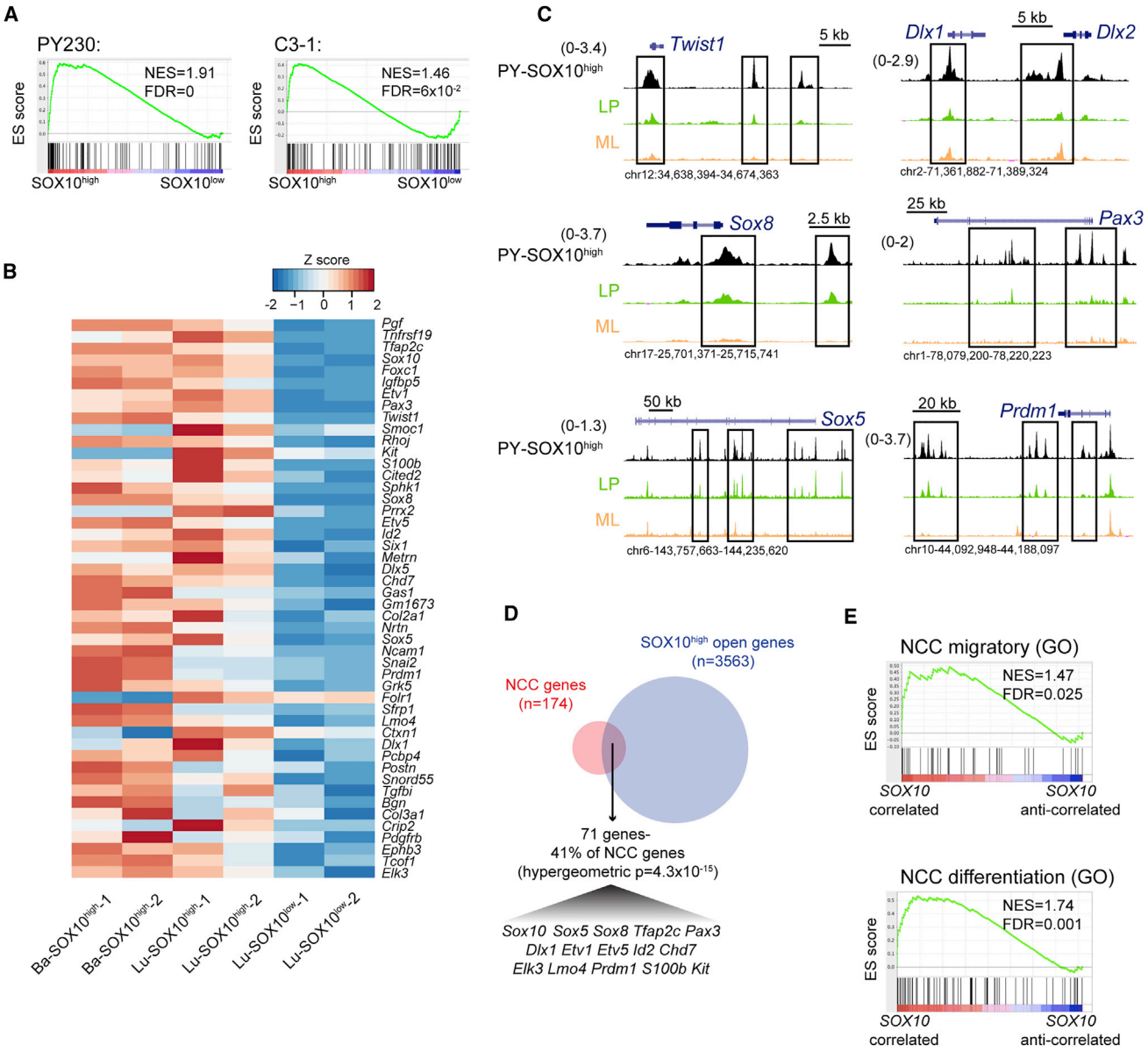


Figure 6. Neural Crest Cell Features Are Present in SOX10^{+} Tumor Cells

(A) GSEA of NCC-related genes in SOX10^{high} versus SOX10^{low} tumor cells.

(B) Heatmap of NCC-related genes that are >1.5-fold upregulated in SOX10^{high} cell fractions of PY230 tumors (n = 2).

(C) ATAC-seq of NCC-specifier genes in SOX10^{high} PY230 tumor cells compared with LP and ML.

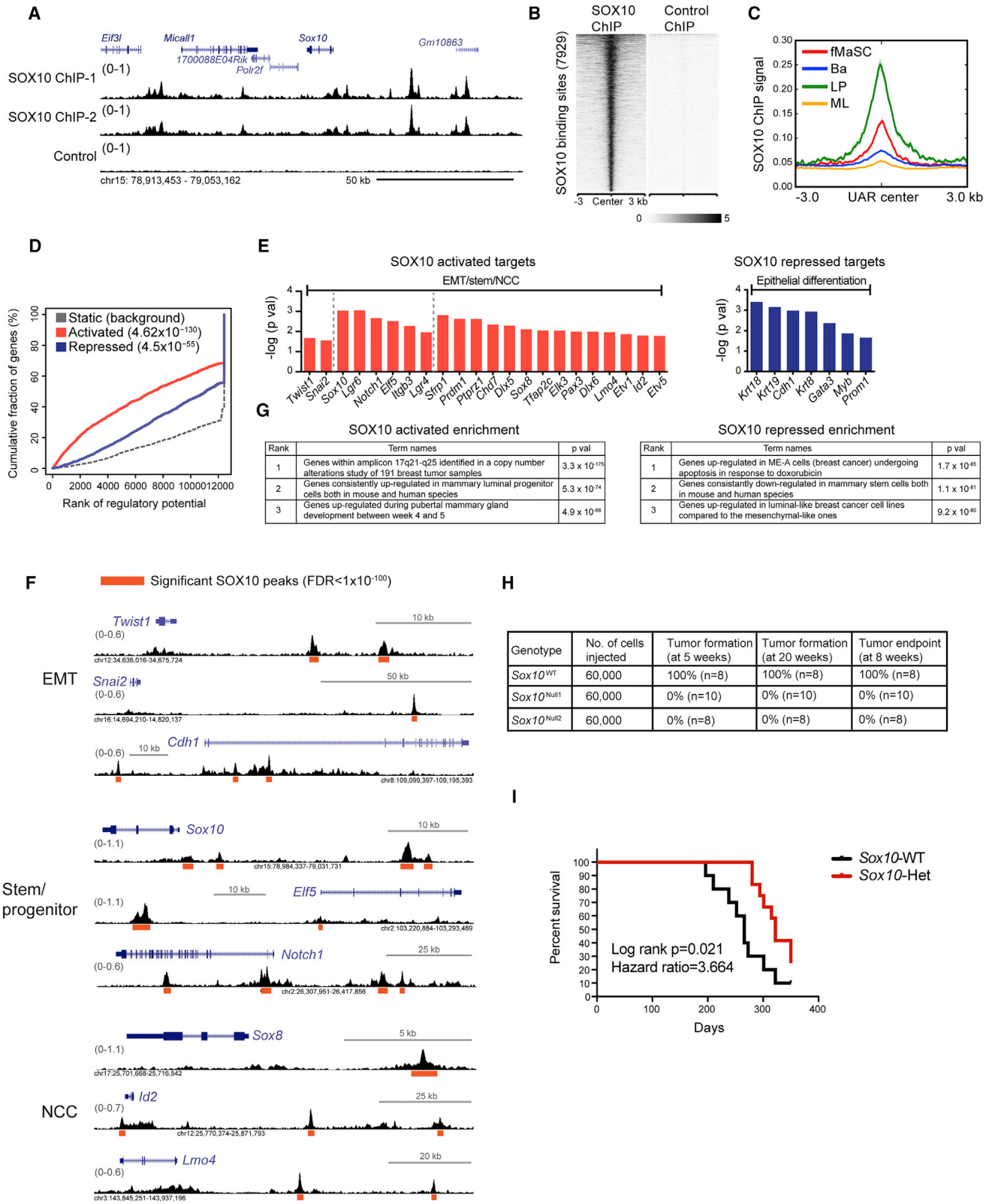
(D) Venn diagram showing overlap of NCC-related genes with genes showing more accessible chromatin in SOX10^{high} PY230 tumor cells.

(E) GSEA of *SOX10* co-expression genes from the TCGA with GO NCC-migratory (GO:0001755) and NCC-differentiation (GO:0014033) genes.

See also Figure S6 and Table S5.

SOX10 ChIP-seq and RNA-seq data. This analysis showed that SOX10 binding correlated with both transcriptional activation (1,526 genes) and repression (1,210 genes) (Figure 7D and Table S6). Many activated genes related to EMT, stem/progenitor identity, and NCC identity, and many repressed genes related to epithelial differentiation, exhibited SOX10 binding to both promoter and distal regions (Figures 7E, 7F, and S7F). GREAT analysis also revealed that SOX10-activated genes were associated with LPs, whereas SOX10-repressed genes were associated with apoptosis and epithelial differentiation in normal and cancerous mammary cells (Figure 7G). In addition, ClueGO network analysis showed that SOX10-activated genes were associated with neural development (related to NCC identity), cell migration, and developmental processes, and others such as metabolism and signal transduction (Figure S7G and Table S6).

These data implicate SOX10 in regulating cell state plasticity in normal and malignant mammary cells. To determine the



Rank	Term names	p val
1	Genes within amplicon 17q21-q25 identified in a copy number alterations study of 191 breast tumor samples	3.3×10^{-175}
2	Genes consistently up-regulated in mammary luminal progenitor cells both in mouse and human species	5.3×10^{-74}
3	Genes up-regulated during pubertal mammary gland development between week 4 and 5	4.9×10^{-66}

Rank	Term names	p val
1	Genes up-regulated in ME-A cells (breast cancer) undergoing apoptosis in response to doxorubicin	1.7×10^{-85}
2	Genes consistantly down-regulated in mammary stem cells both in mouse and human species	1.1×10^{-81}
3	Genes up-regulated in luminal-like breast cancer cell lines compared to the mesenchymal-like ones	9.2×10^{-80}

Genotype	No. of cells injected	Tumor formation (at 5 weeks)	Tumor formation (at 20 weeks)	Tumor endpoint (at 8 weeks)
Sox10^{WT}	60,000	100% (n=8)	100% (n=8)	100% (n=8)
Sox10^{Null1}	60,000	0% (n=10)	0% (n=10)	0% (n=10)
Sox10^{Null2}	60,000	0% (n=8)	0% (n=8)	0% (n=8)

Figure 7. SOX10 Correlates with Differentiation State and Functionally Contributes to Tumor Development

(A) Representative profiles of ChIP-seq from SOX10-biotinylated (two biological replicates) and control PY230 tumors. All ChIP-seq signals are shown as reads per million over input.

(legend continued on next page)

importance of SOX10 in tumor development, we used CRISPR to create *Sox10*^{Null} PY230 cells. These clones were viable and grew normally in 2D culture (Figure S7H), but failed to generate tumors after orthotopic transplantation (Figure 7H). These data indicate an essential function for *Sox10* in PY230 tumor growth. However, the orthotopic transplantation model does not distinguish between roles for SOX10 in tumor formation versus cancer cell engraftment. We turned to the autochthonous model by crossing *Sox10*^{WT} and *Sox10*$^{WT/Null}$ animals with the C3-1 mice to ascertain direct effects on tumor growth. Tumor progression to terminal endpoints was significantly delayed in *Sox10*$^{WT/Null}$ C3-1 mice compared with their wild-type littermates (Figure 7I). These data indicate that the loss of a single copy of *Sox10* significantly delays the events associated with tumor development.

Collectively, these functional data indicate direct roles for SOX10 in specifying cell state changes and promoting tumor development.

DISCUSSION

The epigenetic and transcriptomic databases presented here reveal relationships between bipotent fMaSCs, their luminal and basal cell descendants, and mouse models of mammary cancer. The data show that fMaSCs possess chromatin and transcriptional features that mirror the phenotypic plasticity they exhibit during development. Our analyses provide insights into the ability of basal, but not luminal, cells to act as facultative stem cells and for the observation that LPs, not basal cells, are the likely origins of human basal-like breast cancers. We identified candidate cell state regulators, and investigated one, SOX10, in detail in both mouse mammary cancers and human breast cancer. These studies reveal that SOX10 contributes to normal development and to cancer by regulating genes that control mammary stem/progenitor identity. SOX10 dysregulation in cancer engenders mesenchymal-like features that we show are associated with the acquisition of an embryonic neural crest cell-like state.

The lineage relationships between the component cells of the mammary gland remain controversial. In particular, lineage tracing and functional assays have yielded disparate interpretations concerning mammary cell hierarchy and the stem cell potential of adult mammary cells. These differences may reflect technical factors including promoter leakiness and sampling coverage with lineage-tracing systems, or *in vitro* growth conditions and the effects of transplantation. However, it must also be emphasized that lineage tracing reflects the lineage fate of cells constrained by normal tissue architecture. Thus, cells possessing multi-lineage potential may not be scored as such by lineage tracing if this capacity is context dependent.

We determined the chromatin architecture of loci associated with lineage decisions in mammary cells as alternative, agnostic indicators of lineage flexibility and show that the basal cell and fMaSC populations exhibit accessible chromatin at both basal and luminal regulatory genes. However, and in contrast to fMaSCs, basal cells express basal genes at high levels, but only weakly express some luminal genes. This suggests that basal cells have the potential to adopt either basal or luminal identities, but their behavior is normally constrained *in loco* to that of a basal identity. We infer that under transplantation conditions, the epigenetic state of the basal population enables cells within it to act as facultative bipotential stem cells. This may also be the case in human mammary tissues, as comparison of our mouse ATAC-seq data with published human mammary cell ChIP-seq data reveals that the analogous murine and human mammary populations share epigenetic features. The emerging capacities to perform similar transcriptomic and epigenetic studies at single-cell resolution should minimize caveats created by imprecision associated with analyzing enriched, though heterogeneous, bulk populations.

Mouse models have provided evidence that the LP population contains the cell of origin for stem-like triple-negative breast cancers (Lim et al., 2009), but the underlying molecular basis has remained unclear. Our combined analyses on mouse and human mammary tissues indicate that LPs are the adult cell type that most closely resembles the undifferentiated fMaSC state, and thus may be the cell type most prone to generate an unstable dedifferentiated state after oncogene activation and suppressor gene loss. By contrast, basal cells and fMaSCs did not closely associate by PCA despite both populations exhibiting poised chromatin indicative of multi-lineage potential. One possibility is that the more differentiated state of adult basal cells restricts their ability to acquire bipotentiality to a narrow set of conditions.

An important objective of our analysis was to identify cell state regulators that drive tumor cell plasticity and cell state reprogramming during tumor progression. We found that SOX10 motifs are important regulatory elements associated with state changes in mammary cells, and that human basal-like breast cancers exhibit significantly elevated *SOX10* expression. Importantly, tumor cells expressing high levels of SOX10 exhibit features of multi-lineage stem/progenitor cells, and lineage-associated chromatin features beyond those found in normal mammary cells. The functional and correlative data we present are consistent with the proposal that SOX10 can perturb gene expression to alter cell differentiation states in mouse and human mammary cancers. These data suggest that strategies to abrogate the induction of SOX10-mediated cell state changes may have significant utility in treating aggressive triple-negative and metaplastic breast cancers that currently lack targeted therapies.

(B) SOX10 and control ChIP-seq signal at all SOX10 binding sites (FDR < 1×10^{-100}).
(C) Average SOX10 ChIP-seq signal with 95% confidence interval (shaded regions) at UARs of each mammary cell type.
(D) BETA summary of SOX10 function as a transcriptional activator and repressor in PY230 cells.
(E) Specific activated and repressed targets of SOX10 binding from BETA.
(F) SOX10 ChIP-seq profiles at stem/progenitor, EMT, and NCC genes.
(G) GREAT analysis with genes positively or negatively regulated by SOX10 binding.
(H) Tumor formation following orthotopic transplantation of wild-type and *Sox10*^{Null} PY230 cells.
(I) Kaplan-Meier survival curve for C3-1 *Sox10*^{WT}or *Sox10*$^{WT/Null}$ animals.
See also Figure S7 and Table S6.

Our analyses reveal roles for SOX10 in cell state reprogramming. Elevated SOX10 levels in mammary cells correlated with reprogramming to a cell state with similarities to NCC, an extremely motile and multi-potent embryonic cell type. While some NCC features appear restricted to tumor cells, there is significant overlap between the NCC specification gene module and genes involved in stem/progenitor activity in normal mammary cells (Simoes-Costa and Bronner, 2015). This suggests that multi-lineage potential and migration in NCC and mammary cells may involve common molecular pathways. Clearly, strong parallels exist between the molecular and physiological mechanisms that drive NCC development and presumed stages of metastasis (Powell et al., 2013). NCC reprogramming in tumorigenesis has also been previously suggested to occur in a zebrafish model of melanoma (Kaufman et al., 2016). Melanoma also features prominent SOX10 expression, and is characterized by aggressive disease progression and a high percentage of single cells with the capacity to form new tumors (Quintana et al., 2008). These discoveries provide examples of how dysregulation of transcriptional regulators can reprogram tumor cells to acquire features that were not present in the tissue of origin. Our data suggest that SOX10 can contribute to intra-tumoral heterogeneity and the genesis of motile variants that contribute to metastasis.

Finally, we note that SOX10's role in the genesis of many parameters associated with tumor aggressiveness may be underappreciated because it is not expressed in 2D culture conditions. We also emphasize that SOX10 is likely to be among a much larger cohort of cell state regulators important for both mammary development and tumor progression. Importantly, we found that SOX10^{low} tumor cells also exhibit expanded lineage-associated chromatin features compared with normal mammary cells. These data suggest that there are other differentiation state regulators that contribute to breast tumor development and progression. The agnostic approaches described here should prove valuable for uncovering these regulators.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT RESOURCE AND SHARING
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Mice
  - Human Breast Tumor Samples
- METHOD DETAILS
  - Mammary Tumor Cell Isolation
  - Cell Labeling and Flow Cytometry
  - Immunostaining and Confocal Analyses
  - Mammary Tumor Cell Transplantation
  - Intravital Imaging
  - Mammary Tumor Survival
  - PY230 CRISPR-Based Genome Modification
  - 3-D Tumorsphere Culture
  - ATAC-seq and Data Analysis
  - RNA-seq and Data Analysis
  - TCGA Data Analysis
  - ChIP-seq and Data Analysis
- DATA AND SOFTWARE AVAILABILITY
- QUANTIFICATION AND STATISTICAL ANALYSES

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, six tables, and one video and can be found with this article online at https://doi.org/10.1016/j.ccell.2018.08.001.

ACKNOWLEDGMENTS

We thank Charlene Huang for lab assistance, Rose Rodewald, Cynthia Ramos, and Luke Wang for lab management, Karissa Huang, Alexis Roth, and Jasmine Padilla for genotyping, Alfredo Molinolo at the Moores Cancer Center Biorepository and Tissue Technology Shared Resource for human tumor samples, Conor Fitzpatrick and Caz O'Connor at the Salk Flow Cytometry Core, Max Shokirev at the Salk Bioinformatics Core, Manching Ku at the Salk Next Generation Sequencing Core, John Naughton at the Viral Vector Core, Bing Ren, David Gorkin, Sebastian Preissl, and Sven Heinz for assistance with ATAC- and ChIP-seq, and Jeff Rosen, Raj Giraddi, and David O'Keefe for critical evaluation of the manuscript. G.M.W. was supported by a Cancer Center Core grant (CA014195), NIH/NCI (R35 CA197687), the Susan G. Komen Foundation (SAC110036), and the BCRF. C.D. was supported by CA174430. N.K.L. was supported by GM007752 and CA206416, and T.R. by CA186043 and CA197699.

AUTHOR CONTRIBUTIONS

C.D., C.-Y.C., and G.M.W. designed the experiments; C.D., C.-Y.C., N.K.L., J.H.-V., G.L., and C.L.T. performed the studies; all authors contributed to results interpretation; C.D., C.-Y.C., and G.M.W. wrote the manuscript; T.R. and G.M.W. provided funding.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 15, 2017
Revised: April 16, 2018
Accepted: August 1, 2018
Published: August 30, 2018

REFERENCES

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

Bao, L., Cardiff, R.D., Steinbach, P., Messer, K.S., and Ellies, L.G. (2015a). Multipotent luminal mammary cancer stem cells model tumor heterogeneity. Breast Cancer Res. *17*, 137.

Bao, X., Rubin, A.J., Qu, K., Zhang, J., Giresi, P.G., Chang, H.Y., and Khavari, P.A. (2015b). A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. Genome Biol. *16*, 284.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

Bolstad, B. (2018). *preprocessCore: A Collection of Pre-processing Functions*. R package version 1.40.0, https://github.com/bmbolstad/preprocessCore.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, https://doi.org/10.1002/0471142727.mb2129s109.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Chung, C.Y., Sun, Z., Mullokandov, G., Bosch, A., Qadeer, Z.A., Cihan, E., Rapp, Z., Parsons, R., Aguirre-Ghiso, J.A., Farias, E.F., et al. (2016). Cbx8 acts non-canonically with Wdr5 to promote mammary tumorigenesis. Cell Rep. *16*, 472–486.

Cimino-Mathews, A., Ye, X., Meeker, A., Argani, P., and Emens, L.A. (2013). Metastatic triple-negative breast cancers at first relapse have fewer tumor-infiltrating lymphocytes than their matched primary breast tumors: a pilot study. Hum. Pathol. *44*, 2055–2063.

Corpening, J.C., Deal, K.K., Cantrell, V.A., Skelton, S.B., Buehler, D.P., and Southard-Smith, E.M. (2011). Isolation and live imaging of enteric progenitors based on Sox10-Histone2BVenus transgene expression. Genesis *49*, 599–618.

Davis, F.M., Lloyd-Lewis, B., Harris, O.B., Kozar, S., Winton, D.J., Muresan, L., and Watson, C.J. (2016). Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. Nat. Commun. *7*, 13053.

de Sousa e Melo, F., Kurtova, A.V., Harnoss, J.M., Kljavin, N., Hoeck, J.D., Hung, J., Anderson, J.E., Storm, E.E., Modrusan, Z., Koeppen, H., et al. (2017). A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. Nature *543*, 676–680.

de Vries, W.N., Binns, L.T., Fancher, K.S., Dean, J., Moore, R., Kemler, R., and Knowles, B.B. (2000). Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. Genesis *26*, 110–112.

Dravis, C., Spike, B.T., Harrell, J.C., Johns, C., Trejo, C.L., Southard-Smith, E.M., Perou, C.M., and Wahl, G.M. (2015). Sox10 regulates stem/progenitor and mesenchymal cell states in mammary epithelial cells. Cell Rep. *12*, 2035–2048.

Finzsch, M., Schreiner, S., Kichko, T., Reeh, P., Tamm, E.R., Bosl, M.R., Meijer, D., and Wegner, M. (2010). Sox10 is required for Schwann cell identity and progression beyond the immature Schwann cell stage. J. Cell Biol. *189*, 701–712.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.

Ge, Y., and Fuchs, E. (2018). Stretching the limits: from homeostasis to stem cell plasticity in wound healing and cancer. Nat. Rev. Genet. *19*, 311–325.

Ge, Y., Gomez, N.C., Adam, R.C., Nikolova, M., Yang, H., Verma, A., Lu, C.P., Polak, L., Yuan, S., Elemento, O., et al. (2017). Stem cell lineage infidelity drives wound repair and cancer. Cell *169*, 636–650.e14.

Giraddi, R.R., Shehata, M., Gallardo, M., Blasco, M.A., Simons, B.D., and Stingl, J. (2015). Stem and progenitor cell division kinetics during postnatal mouse mammary gland development. Nat. Commun. *6*, 8487.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Inman, J.L., Robertson, C., Mott, J.D., and Bissell, M.J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. Development *142*, 1028–1042.

Kaufman, C.K., Mosimann, C., Fan, Z.P., Yang, S., Thomas, A.J., Ablain, J., Tan, J.L., Fogley, R.D., van Rooijen, E., Hagedorn, E.J., et al. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. Science *351*, aad2197.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

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.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.

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. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913.

Makarem, M., Kannan, N., Nguyen, L.V., Knapp, D.J., Balani, S., Prater, M.D., Stingl, J., Raouf, A., Nemirovsky, O., Eirew, P., et al. (2013). Developmental changes in the in vitro activated regenerative activity of primitive mammary epithelial cells. PLoS Biol. *11*, e1001630.

Maroulakou, I.G., Anver, M., Garrett, L., and Green, J.E. (1994). Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. Proc. Natl. Acad. Sci. USA *91*, 11236–11240.

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? Nat. Rev. Cancer *12*, 323–334.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Pal, B., Bouras, T., Shi, W., Vaillant, F., Sheridan, J.M., Fu, N., Breslin, K., Jiang, K., Ritchie, M.E., Young, M., et al. (2013). Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. Cell Rep. *3*, 411–426.

Pellacani, D., Bilenky, M., Kannan, N., Heravi-Moussavi, A., Knapp, D., Gakkhar, S., Moksa, M., Carles, A., Moore, R., Mungall, A.J., et al. (2016). Analysis of normal human mammary epigenomes reveals cell-specific active enhancer states and associated transcription factor networks. Cell Rep. *17*, 2060–2074.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. *11*, 1650–1667.

Pfefferle, A.D., Herschkowitz, J.I., Usary, J., Harrell, J.C., Spike, B.T., Adams, J.R., Torres-Arzayus, M.I., Brown, M., Egan, S.E., Wahl, G.M., et al. (2013). Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol. *14*, R125.

Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Powell, D.R., Blasky, A.J., Britt, S.G., and Artinger, K.B. (2013). Riding the crest of the wave: parallels between the neural crest and cancer in epithelial-to-mesenchymal transition and migration. Wiley Interdiscip. Rev. Syst. Biol. Med. *5*, 511–522.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quintana, E., Shackleton, M., Sabel, M.S., Fullen, D.R., Johnson, T.M., and Morrison, S.J. (2008). Efficient tumour formation by single human melanoma cells. Nature *456*, 593–598.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

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.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. *6*, R33.

Schwitalla, S., Fingerle, A.A., Cammareri, P., Nebelsiek, T., Goktuna, 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.

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.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shimokawa, M., Ohta, Y., Nishikori, S., Matano, M., Takano, A., Fujii, M., Date, S., Sugimoto, S., Kanai, T., and Sato, T. (2017). Visualization and targeting of LGR5+ human colon cancer stem cells. Nature *545*, 187–192.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Simoes-Costa, M., and Bronner, M.E. (2015). Establishing neural crest identity: a gene regulatory recipe. Development *142*, 242–257.

Southard-Smith, E.M., Kos, L., and Pavan, W.J. (1998). Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. Nat. Genet. *18*, 60–64.

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.

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.

Turner, N., Tutt, A., and Ashworth, A. (2004). Hallmarks of 'BRCAness' in sporadic cancers. Nat. Rev. Cancer *4*, 814–819.

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.

Wahl, G.M., and Spike, B.T. (2017). Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. NPJ Breast Cancer *3*, 14.

Wainwright, E.N., and Scaffidi, P. (2017). Epigenetics and cancer stem cells: unleashing, hijacking, and restricting cellular plasticity. Trends Cancer *3*, 372–386.

Wang, D., Cai, C., Dong, X., Yu, Q.C., Zhang, X.O., Yang, L., and Zeng, Y.A. (2015). Identification of multipotent mammary stem cells by protein C receptor expression. Nature *517*, 81–84.

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y., and Liu, X.S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat. Protoc. *8*, 2502–2515.

Wuidart, A., Ousset, M., Rulands, S., Simons, B.D., Van Keymeulen, A., and Blanpain, C. (2016). Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. Genes Dev. *30*, 1261–1277.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952–1958.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Alexa Fluor® 647 anti-mouse CD326 (Ep-CAM)	Biolegend	RRID: AB_1134101
Brilliant Violet 421™ Streptavidin	Biolegend	Cat #: 405225
Biotin Rat Anti-Mouse TER-119/ Erythroid Cells	BD Biosciences	Cat #: 553672
Biotin Rat Anti-Mouse CD31	BD Biosciences	Cat #: 553371
Biotin Rat Anti-Mouse CD45	BD Biosciences	Cat #: 553078
Purified Rat Anti-Mouse CD16/CD32 (Mouse BD Fc Block™)	BD Biosciences	Cat #: 553142
Keratin, type II/ Cytokeratin 8	DSHB	Cat #: TROMA-1
Keratin 14 Polyclonal Antibody, Purified	Biolegend	RRID: AB_2565048
Anti-Vimentin Antibody	Millipore	Cat #: AB5733
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Invitrogen	RRID: AB_143157
Goat anti-Rat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 633	Invitrogen	RRID: AB_2535749
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 660	Invitrogen	RRID: AB_2535734
Goat anti-Rat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	RRID: AB_2534074
Goat anti-Chicken IgY (H+L) Secondary Antibody, Alexa Fluor 568	Invitrogen	RRID: AB_2534098
Purified anti-SOX-10 Antibody	Biolegend	RRID: AB_2629666
Alexa Fluor® 647 anti-mouse CD144 (VE-cadherin) Antibody	Biolegend	RRID: AB_10568319
H3K27ac polyclonal antibody	Diagenode	Cat #: C15410174
PE anti-mouse CD61 antibody	Biolegend	RRID: AB_313084
Bacterial and Virus Strains		
LentiCrisprV2	https://doi.org/10.1038/nmeth.3047	Addgene 52961
LV-CMV-eGFP	Salk Viral Vector Core	N/A
LV-SIN-Ubi-iCre-mCherry	Salk Viral Vector Core	N/A
Biological Samples		
Human Breast Cancer Sections	UCSD Biorepository and Tissue Technology	Alfredo Molinolo
Chemicals, Peptides, and Recombinant Proteins		
Scr7	SelleckChem	Cat#: S7742
EGF	Stem Cell Technologies	Cat#: 78006
bFGF	Stem Cell Technologies	Cat#: 78003
DAPI	Thermo Fisher Scientific	Cat#: 62248
Critical Commercial Assays		
EpiCult™-B Mouse Medium Kit	Stem Cell Technologies	Cat #: 05610
Growth Factor Reduced Matrigel	Corning	Cat#: 356231
Nextera DNA Library Preparation Kit	Illumina	CAT#: FC-121-1030
Nextera XT DNA Library Preparation Kit	Illumina	CAT#: FC-131-1096

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AMPure XP beads	Beckman	CAT#: A63881
Protein A magnetic beads	Invitrogen	CAT#: 10001D
Streptavidin magnetic beads	Pierce	CAT#: 88816
Deposited Data		
Raw sequencing and processed data (fastq, bigwig, bed and Excel files)	This paper	GEO: GSE116386
Mouse mammary cell histone mark ChIP-seq data (fastq files)	Pal et al., 2013	GEO: GSE43212
Human mammary cell enhancer genomic location (bed files)	Pellacani et al., 2016	https://www.cell.com/cms/attachment/2110641945/2083283388/mmc6.xlsx
Experimental Models: Cell Lines		
PY230 mammary tumor cell line	Bao et al., 2015a; Bao et al., 2015b	Lesley Ellies
Experimental Models: Organisms/Strains		
Sox10-H2BVenus	Corpening et al., 2011	
Sox10^{flox}	Finzsch et al., 2010	
C3-1-TAg	Maroulakou et al., 1994	
Trp53^{flox};*Brca1*^{flox}	Perou, CM	
Zp3-Cre	de Vries et al., 2000	
Oligonucleotides		
AGGCCAAGCCCTGACTGAGC	IDTDNA	sgRNA to target *Sox10* 3' locus for reporter and epitope tagging
CCAGCGACGGCGCGCTGCCT	IDTDNA	sgRNA-1 to target *Sox10* ORF
GGCGGCGGCCGGGAGCGACA	IDTDNA	sgRNA-2 to target *Sox10* ORF
Recombinant DNA		
pAD5-AAV Helper construct	Salk Viral Vector Core	
pdJ - psuedotype DJ for AAV	Salk Viral Vector Core	
Software and Algorithms		
Bowtie (v0.12.8)	http://bowtie-bio.sourceforge.net/	Langmead et al., 2009
MACS2 (v2.1.1)	http://liulab.dfci.harvard.edu/MACS/	Zhang et al., 2008
Bedtools (v2.20.1)	http://bedtools.readthedocs.io/en/latest/	Quinlan and Hall, 2010
Samtools (v1.3.1)	http://www.htslib.org/doc/samtools.html	Li et al., 2009
Deeptools (v2.4.1)	https://deeptools.readthedocs.io/en/latest/	Ramirez et al., 2014
GREAT (v3.0.0)	http://great.stanford.edu/public/html/index.php	McLean et al., 2010
SICER (v1.1)	https://home.gwu.edu/~wpeng/Software.htm	Zang et al., 2009
preprocessCore (R)	https://www.bioconductor.org/packages/release/bioc/html/preprocessCore.html	Bolstad, 2018
R (v3.4.4 for Mac OSX)	https://cran.r-project.org/bin/windows/base/	
Cytoscape (v3.3.0)	www.cytoscape.org	Shannon et al., 2003
Cytoscape ClueGO plugin	http://apps.cytoscape.org/apps/cluego	Bindea et al., 2009
Hisat2 (v2.0.5)	https://ccb.jhu.edu/software/hisat2/manual.shtml	Kim et al., 2015
Stringtie (v1.3.3)	http://ccb.jhu.edu/software/stringtie/index.shtml	Pertea et al., 2016
Ballgown (v1.0.1)	https://github.com/alyssafrazee/ballgown	Pertea et al., 2016
Homer (v4.8)	http://homer.ucsd.edu/homer/index.html	Heinz et al., 2010
GSEA	https://software.broadinstitute.org/gsea/index.jsp	Tamayo et al., 2005
BETA plus (v1.0.2)	http://cistrome.org/BETA/	Wang et al., 2013

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Hydrocortisone	Sigma Aldrich	Cat#: H4001
Collagenase/Hyaluronidase	Stem Cell Technologies	Cat#: 07912
Dispase	Stem Cell Technologies	Cat#: 07913
Trypsin	Thermo Fisher Scientific	Cat#: 25300-054
Matrigel (complete)	Corning	Cat#: 354234
Matrigel (growth factor reduced)	Corning	Cat#: 356231

CONTACT FOR REAGENT RESOURCE AND SHARING

Further information and requests for resources and software should be directed to and will be fulfilled by the Lead Contact, Geoffrey M. Wahl (wahl@salk.edu).

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Mice

The *Sox10*-H2BVenus (Corpening et al., 2011), *Sox10*^{flox} (Finzsch et al., 2010), C3-1-TAg (Maroulakou et al., 1994), and *Trp53*^{flox};*Brca1*^{flox} (Perou CM lab, unpublished data) mice have either been previously described or are nearing publication. For tumor studies, mice were maintained in an FVB background, except for the *Sox10*^{flox};*C3-1*-TAg study in which mice were in a mixed CD1/FVB/B6 background. The *Sox10*^{flox} LoxP cassette was deleted by a *Zp3*-Cre mouse (de Vries et al., 2000). Orthotopic transplantation of PY230 cells were performed with 6-10 week old adult wild-type C57bl/6 mice. All animals were handled in accordance with Salk Institute IACUC and AAALAC approved protocols and other ethics guidelines.

Human Breast Tumor Samples

Human breast tumor samples were provided by Moores Cancer Center at UC San Diego Health Comprehensive Biorepository, which is funded by the National Cancer Institute (NCI P30CA23100). Samples were isolated by Dr. Oluwole Fadare under protocol 161362, "Repeat ER, PR and HER2/neu testing in breast cancers", which was approved as a minimal risk study with waiver of informed consent by the Human Research and Protections Program at UC San Diego.

METHOD DETAILS

Mammary Tumor Cell Isolation

Mammary tumors were dissected out of freshly euthanized mice, minced into small pieces, placed in dissociation media (Epicult-B Basal medium (Stem Cell Technologies) supplemented with 5% FBS, pen/strep, fungizone, hydrocortisone, collagenase and hyaluronidase), and agitated with shaking for 3 hours at 37°C. Next, erythrocytes were removed with ammonium chloride exposure for 4 minutes on ice, followed by treatment and trituration with dispase and DNase. Final suspensions were passed through a 40 um filter to remove aggregated cells, and stored in Hank's Balanced Salt Solution with 2% FBS for flow cytometry.

Cell Labeling and Flow Cytometry

For labeling, the following antibodies were used: EpCAM-647, CD49f-APCCy7, CD61-PE, Streptavidin(SA)-BV421 (BioLegend), Biotin-Ter-119, Biotin-CD45, Biotin-CD31, and mouse BD Fc Block (BD Biosciences). For sorting tumor cells, DAPI and 421(Lineage)$^{+}$ cells were excluded, while EpCAM and CD49f were used to established luminal-like (EpCAM^{High};CD49f$^{Low\text{-}Med}$) and basal-like gates (EpCAM$^{Low\text{-}Med}$;CD49f^{High}), except for the C3-1-TAg tumors which were sorted off of EpCAM^{High} and Venus fluorescence.

Immunostaining and Confocal Analyses

For frozen sections, mammary tumors or lungs were fixed in formalin for 1 hour, cryoprotected in 30% sucrose with shaking overnight, and embedded in OCT. Tissue sections were washed 3 x 5 minutes in wash buffer (PBS-T containing 0.3% TritonX-100), blocked in 10% goat serum or 5% donkey serum for 1 hour, incubated with the primary antibodies for 8-12 hours at 4°C in blocking buffer, washed 4 x 20 minutes in wash buffer, incubated with the secondary antibodies for 1-2 hours at room temperature in blocking buffer, washed 4 x 20 minutes with wash buffer, and mounted (Vectashield, Vector Labs). For paraffin sections, mammary tumors were fixed in formalin overnight, stored in 70% ethanol, and processed for paraffin embedding and sectioning. Tissue sections were de-paraffinized, rehydrated, and treated with sodium citrate buffer for antigen retrieval, before being stained in the same process as with frozen sections. Tumorspheres and dissociated single cells were processed for immunofluorescence by fixing them in a Matrigel bed with formalin for 30 minutes, then directly starting the immunostain process as with frozen sections. Primary antibodies

used for immunofluorescence include: keratin-14 (Covance), keratin-8 (Troma-1, DSHB), Vimentin (Millipore), and SOX10 (BioLegend). Secondary antibodies used for immunofluorescence include: Alexa Fluors: 568 goat anti-rabbit, 633 goat anti-rat, 660 goat anti-rabbit, 488 goat anti-rat, and 568 goat anti-chicken, all from Molecular Probes (Invitrogen). Confocal microscopy was performed with equipment from the Waitt Advanced Biophotonics Center at the Salk Institute, including Zeiss LSM 880 with Airyscan and 780-inverted laser scanning confocal microscopes. For confocal images, Adobe Photoshop was used to increase field-wide brightness levels.

Mammary Tumor Cell Transplantation

PY230 cells grown in 2D culture were orthotopically transplanted into the #4 fat pads of 6-10 week old syngeneic mice. 10,000-200,000 cells were used per transplantation, unless otherwise indicated in text. Surgery and recovery of animals followed strict and IACUC-approved protocols.

Intravital Imaging

Sox10^{tdTomatoNLS} PY230 cells infected with LV-CMV-eGFP were orthotopically transplanted into 6-10 week-old female syngeneic mice. Tumors were allowed to develop until they reached approximately 0.25cm^3. For imaging, mice were anesthetized by IP injection of ketamine and xylazine (100/20 mg/kg) and maintained under anesthesia throughout the procedure using 1-2% (vol/vol) isoflurane gas mixed with oxygen. In order to visualize blood vessels and nuclei, mice were injected retro-orbitally with AlexaFluor 647 anti-mouse CD144 (VE-cadherin) antibody and Hoechst 33342 immediately following anesthesia induction. Tumors were exposed by carefully removing hair, skin, and connective tissue while keeping tumor vasculature intact. Mice were then placed inverted on an imaging apparatus, and each tumor was elevated and stabilized on a glass slide to reduce breathing artifacts. 80-150 micron images in 1024 × 1024 format were acquired with an HCX APO L20x objective on an upright Leica SP5 confocal system using Leica LAS AF 1.8.2 software. Videos were generated using Volocity 3D Image Analysis Software and compressed using Microsoft Video 1 compression.

Mammary Tumor Survival

For the *C3-1*-TAg;*Sox10* survival study, mice were examined for tumor development once every 7 days. Any mouse presenting with a mammary tumor of >10 mm size was considered as reaching the study endpoint, and the animal was recovered for euthanasia.

PY230 CRISPR-Based Genome Modification

PY230 cells were grown in media conditions that have previously been described (Bao et al., 2015a). For genome modification of PY230 cells, the PY230 cells were first infected at <3 MOI with CD0616, a lentivirus containing a floxed Cas9-2A-G418^R cassette modified from LentiCrisprV2 (Sanjana et al., 2014). G418-resistant cells were expanded as PY230-Cas9 cells. To target the *Sox10* locus for tdTomato, a targeting vector was designed in an AAV backbone sequentially containing: an sgRNA cassette versus genomic sequence proximally downstream of the stop codon (AGGCCAAGCCCTGACTGAGC), 268 bp 5' homology arm, in-frame AVI-V5-V5-2A-tdTomatoNLS cassette, LoxP-EFS-TagBFP-CW3SL-LoxP, 394 bp 3' homology arm. AAV was generated by transfection with PEI (Polysciences) in 293A cells with the AAV SOX10 TV, transfer plasmid, and DJ cap plasmid. PY230-Cas9 cells were infected with the AAV, TagBFP$^+$ cells were isolated by FACS and plated at clonal density, and PY230 clones were picked, expanded, screened by PCR, and sequenced to validated candidates. AAV-Cre was then used to remove the Cas9 and the TagBFP cassette. Protein lysates from PY230-*Sox10*^{tdTomato} mammary tumors that were immunoprecipitated with streptavidin and immunoblotted with a V5 antibody confirmed the specific presence of a single protein at SOX10's expected size of 60-70 kDa. To produce null alleles of *Sox10* in PY230 cells, a similar strategy was used to identify *Sox10*^{Null} clones, except the PY230-Cas9 cells were infected with an AAV sequentially containing: two sgRNAs cassettes targeting necessary coding regions near the start codon (CCAGCGACGGCG CGCTGCCT) and (GGCGGCGGCCGGGAGCGACA), and an EFS-tdTomatoNLS-WPRE cassette. Viability of resulting clones was confirmed by plating 100,000 cells *Sox10*^{Null} cells in 2D culture and quantifying their cell proliferation after 4 days, in comparison to the parental PY230 cell line.

3-D Tumorsphere Culture

For growth of C3-1-TAg cells in 3-D culture conditions, previously described protocols for mammary cell organoid formation 2% GFR Matrigel (Corning) were followed (Dravis et al., 2015).

ATAC-seq and Data Analysis

The transposition assay was performed as previously described (Buenrostro et al., 2015). Around 2x10^4 nuclei from each normal and tumor cell type were used in each reaction with 20 μl of transposition mix (10 μl 2x TD, 2 μl TDE1, 8 μl H2O; Illumina Nextera FC-121-1030) and incubated at 37°C for 30 minutes. qPCR was performed to determine the cycle number for 25% library saturation. Typically, 10-14 total cycles were performed. The library was purified with AMPure XP beads (Beckman A63881), and then analyzed by Agilent TapeStation, and 37, 75 or 125 bp paired-end sequencing, or 50 bp single-end sequencing was performed with Illumina HiSeq 2500 or NextSeq 500. ATAC-seq analysis was performed as previously described (Bao et al., 2015b). In brief, after quality check with FastQC, sequencing reads were mapped to the mouse genome (mm9) with Bowtie (Langmead et al., 2009), with these parameters: -m 1 -S -n 2 -l 30. Since we were only interested in the cutting sites of Tn5 that represents open chromatin regions,

paired-end samples were mapped as single-end, and bam files from the same paired-end sample were merged into one bam file. Duplicated reads were removed and peak calling was done with MACS2 (Zhang et al., 2008), with these parameters: –nolambda –nomodel –shift -100 –extsize 200. Bedgraph files generated were then converted into BigWig format and visualized on UCSC genome browser (https://genome.ucsc.edu/). Genome-wide average signal profile at genes was checked for each sample to ensure similar signal-to-noise level. Signal profiling, correlation analysis and k-means clustering were performed using deepTools (Ramirez et al., 2014). Functional annotation of peaks and peak-gene association were done with GREAT using the default ''basal plus extension'' parameter (McLean et al., 2010). Differential peaks were called with SICER-df-rb (Zang et al., 2009), with these parameters: window size: 100, gap size: 100, E-value: 0.01, FDR: 0.05. HOMER was used for motif enrichment analysis (Heinz et al., 2010). Only the top ten enriched motifs (ranked by p value) whose corresponding TFs are expressed (RPKM > 1) in the specific cell types are shown in Figure 2F. Comparison of genes associated with human mammary regulatory elements and mouse UARs was conducted with Jaccard index normalized to have mean 0 and standard deviation 1 for each UAR.

Shannon entropy was calculated as previously described (Schug et al., 2005). In brief, Shannon entropy H of each peak k among n different cell types is calculated as:

$$H_k = -\sum_{i=1}^{n} P(X_i) log_2 P(X_i)$$

Where:

$$P(X_i) = \frac{S_i}{\sum_{i=1}^{n} S_i}$$

Here, S_i = ATAC-seq signal at peak k in cell type i.

To isolate UARs and URRs (see Figure S2A), pairwise differential peaks (FC > 2 and FDR < $1x10^{-30}$) between each cell type were first determined using SICER-df, and enrichment score (ES) for each peak calculated as ES = FC x –log(FDR). Cell type specific regions were then isolated by cross comparison of peaks using bedTools intersect (Quinlan and Hall, 2010). Afterward, total enrichment score (TES) was calculated by adding up cell type specific ES. For example, the TES of fMaSC = $ES_{fMaSC\text{-}Ba}$ + $ES_{MaSC\text{-}LP}$ + $ES_{fMaSC\text{-}ML}$. Thus, cell-type specificity of UARs and URRs can be ranked by their TES.

PCA of normal and tumor mammary cells were analyzed using all UARs and URRs to best separate the normal cell types. To control for sample and peak variances, ATAC-seq signals of each sample were first normalized to mean=0 and SD=1, and then each peak was normalized with the mean signal of all cell types. Singular value decomposition was then performed on the normalized signal using the R function svd. PC scores were calculated by multiplying v with d and plotted. Presumed relationship between normal cells (differentiation trajectory) was plotted with arrows linking the PC score centroids of cell replicates. PC1 and PC2 projection for the tumor cells were plotted as heatmaps to interpret their chromatin state.

RNA-seq and Data Analysis

Low input bulk RNA-seq was performed using the Smart-seq2 protocol as previously described (Picelli et al., 2014). In brief, around 2000 cells from each cell type were processed. Full-length cDNA were generated, and their size distribution was checked with TapeStation to ensure good RNA quality. The cDNA were then amplified with 18 PCR cycles, tagmentated and amplified again with 10 PCR cycles using Nextera XT kit (Illumina FC-131-1096). The sequencing library was purified with AMPure XP beads. 50 bp or 75 bp single-end sequencing was performed with Illumina HiSeq 2500. Sequencing reads were quality checked with FastQC and mapped to the mouse genome (mm9) using Hisat2 (Kim et al., 2015). Transcripts were assembled by Stringtie. Transcript quantification and differential expression analysis were performed by Ballgown (Pertea et al., 2016). To exclude non-expressed genes, genes that have RPKM variance across samples < 1 are removed. Gene expression distributions between samples were checked to ensure similar transcriptome quality. Mammary cell signatures were generated using an entropy method similar to above and as previously described (Schug et al., 2005), and with this condition: fMaSC and basal signature: H < 0.8 & up in fMaSC/basal & RPKM in fMaSC/basal > 5; LP and ML signature: H < 1 & up in LP/ML & RPKM in LP/ML > 5. GSEA (http://software.broadinstitute.org/gsea/index.jsp) was performed using fold change pre-rank and the indicated gene sets.

TCGA Data Analysis

2012 and 2015 TCGA breast tumor gene expression data, and the SOX10 correlation data, were downloaded from cBioPortal (http://www.cbioportal.org/) (Cancer Genome Atlas Network, 2012; Gao et al., 2013). Samples with incomplete information were removed before analysis. Gene co-expression network was constructed using pair-wise Spearman correlation between samples with r > 0.35 as cutoff, and the Fruchterman-Reingold algorithm was used to connect the gene nodes using the R package igraph.

ChIP-seq and Data Analysis

Low input ChIP was performed as previously described with some modifications (Schmidl et al., 2015). For H3K27ac ChIP, around $2x10^4$ cells were crosslinked with 1% PFA at room temperature for 10 minutes, sonicated with Covaris M220 into 200-700 bp fragments, and incubated with anti-H3K27ac antibody (Diagenode C15410174) overnight at 4°C. Protein A beads pull down (Invitrogen 10001D), washing, on-beads tagmentation (Illumina Nextera FC-121-1030), reverse crosslinking, library amplification

(13 PCR cycles) and DNA purification were performed as described. The input library was prepared by tagmentation of 1 ng of reverse crosslinked input DNA and then amplified and processed together with the ChIP DNA. Multiple input preparations were pooled together to ensure sufficient complexity of the input library. For the SOX10 ChIP, $5x10^5$ control and SOX10-AVI/BirA cells were processed as above, except that streptavidin beads (Pierce #88816) were used for pull down, and that two additional washing steps with 2% SDS in PBS were performed to significantly remove ChIP backgrounds. The input library was prepared as above. All controls and ChIP samples were analyzed by Agilent TapeStation, and 50 bp single-end sequencing was performed with Illumina HiSeq 2500.

ChIP-seq data analyses were performed as previously described (Chung et al., 2016). In brief, sequencing reads were checked by FastQC and aligned to mm9 with Bowtie (Langmead et al., 2009). MACS2 was used to call peaks and generate Bedgraph files that show fold change enrichment over input (Zhang et al., 2008). Bedgraph files were then converted into BigWig files and uploaded to UCSC Genome Browser for visualization. As we have observed variable between samples signal-to-noise levels for the H3K27ac ChIP-seq, we normalized the ChIP signal using genome-wide quantile normalization at 100 bp window using the preprocessCore package in R, and then converted the Bedgraph files to BigWig files for further analyses. ChIP-seq profiling, motif analysis and GREAT annotation were conducted as described above. SOX10 target prediction was performed with BETA plus (Wang et al., 2013) using the following parameter: –da 0.25 -d 150000. Gene ontology network analysis of SOX10 targets was performed with ClueGO plug-in of Cytoscape (Bindea et al., 2009; Shannon et al., 2003). Fastq files of H3K4me3 and H3K27me3 ChIP-seq of mouse mammary cell populations were downloaded from GEO database (Pal et al., 2013). Reads were mapped and processed as described above. Signal profiling was performed using deepTools (Ramirez et al., 2014).

DATA AND SOFTWARE AVAILABILITY

The accession number for the ATAC-, RNA-, and ChIP-seq data in this paper is GEO: GSE116386.

QUANTIFICATION AND STATISTICAL ANALYSES

All statistical analyses, data processing and heatmap plotting were performed in R (https://www.r-project.org/), unless otherwise noted. Statistical significance of the difference in entropy distribution was calculated with Kolmogorov-Smirnov test. Statistical significance of the difference in the proportion of promoter versus distal ATAC-seq peaks was performed with two-sample test of equal proportions with two-sided alternative hypothesis. Statistical significance was calculated with unpaired two-tailed student's t test assuming equal variance, unless otherwise noted.