Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity

Citation: Chaffer, Christine L., Nemanja D. Marjanovic, Tony Lee, George Bell, Celina G. Kleer, Ferenc Reinhardt, Ana C. D'Alessio, Richard A. Young, and Robert A. Weinberg. "Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity." Cell 154:1 (3 July 2013), pp. 61-74.

As Published: http://dx.doi.org/10.1016/j.cell.2013.06.005

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/103883

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Published in final edited form as:
Cell. 2013 July 3; 154(1): 61–74. doi:10.1016/j.cell.2013.06.005.

Poised chromatin at the ZEB1 promoter enables cell plasticity and enhances tumorigenicity

Christine L Chaffer[1,*], **Nemanja D Marjanovic**[1,*], **Tony Lee**[1], **George Bell**[1], **Celina G Kleer**[2], **Ferenc Reinhardt**[1], **Ana C D'Alessio**[1], **Richard A Young**[1,3], and **Robert A Weinberg**[1,4]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

[2]University of Michigan Medical School, Department of Pathology, Ann Arbor MI, 48109, USA

[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4]Ludwig MIT Center for Molecular Oncology, Cambridge, MA 02139, USA

Summary

The recent discovery that normal and neoplastic epithelial cells re-enter the stem-cell state raised an intriguing possibility in the context of cancer pathogenesis: the aggressiveness of carcinomas derives not from their existing content of cancer stem cells (CSCs), but from their proclivity to generate new CSCs from non-CSC populations. Here we demonstrate that non-CSCs of human basal breast cancers are plastic cell populations that readily switch from a non-CSC to CSC-state. The observed cell plasticity is dependent on ZEB1, a key regulator of the epithelial-mesenchymal transition. We find plastic non-CSCs maintain the ZEB1 promoter in a bivalent chromatin configuration enabling them to respond readily to microenvironmental signals, such as TGFbeta. In response, the ZEB1 promoter converts from a bivalent to active chromatin configuration, ZEB1 transcription increases and non-CSCs subsequently enter the CSC state. Our findings support a dynamic model where interconversions between low and high tumorigenic states occur frequently, thereby increasing tumorigenic and malignant potential.

Introduction

Metastatic dissemination and disease relapse are critical determinants of cancer prognosis. The mechanisms underlying both processes remain poorly understood. Recent advances in understanding cellular hierarchies present within a variety of tumors have changed our perspective of neoplastic cell population organization. In particular, cell-surface antigen markers have revealed distinct subpopulations of neoplastic cells within tumors showing pronounced differences in tumor-initiating and metastatic powers (Visvader and Lindeman,

Corresponding authors: Christine Chaffer: chaffer@wi.mit.edu, Robert Weinberg: weinberg@wi.mit.edu.
*These authors contributed equally to this work

2012). Such evidence indicates that within individual tumors, genetically identical cancer cells may nonetheless reside in distinct phenotypic states.

Importantly, tumors derived from implanting highly tumorigenic subpopulations of cells exhibit the phenotypic heterogeneity of their predecessor tumors, in that they contain both highly and weakly tumorigenic cells (Visvader and Lindeman, 2012). Implicit is the notion that highly tumorigenic cells can self-renew and also divide asymmetrically into daughter cells with low tumorigenic potential. Parallels identified with cell hierarchies operating in normal adult tissues have led to coining of the term “cancer stem cell” (CSC) to describe the subset of neoplastic cells that reside in a highly tumorigenic state.

The simplest depiction would portray CSCs as residing at the apex of a cellular hierarchy and spawning, in a unidirectional manner, more differentiated non-CSC progeny. Cells in a number of cancer types conform to that model (Bonnet and Dick, 1997; Visvader and Lindeman, 2012). These studies imply that once a CSC has exited the CSC state it cannot re-enter it. This principle of unidirectionality holds great importance given the significance of CSCs for cancer development and, quite possibly, progression to metastatic disease.

A small number of studies now suggest that not all cancers strictly conform to the unidirectional hierarchical CSC model. We and others have recently demonstrated that non-CSCs can acquire CSC-like activity under certain conditions (Chaffer et al., 2011; Gupta et al., 2011; Roesch et al., 2010). These studies open the door to the possibility that there is likely to be greater plasticity in cancer cell populations – yielding bi-directional interconversions between CSC and non-CSC states – than is depicted in the simplest version of the CSC model.

It has remained unclear whether these interconversions are confined to specific types of cancer, how frequently they occur *in vivo*, and how they are achieved mechanistically. These interconversions are potentially important for cancer diagnosis, prognosis and therapy, given the now-extensive evidence that CSCs are intrinsically more prone to disseminate and, at the same time, exhibit resistance to many existing anti-tumor therapies (Dean et al., 2005; Malanchi et al., 2012). In the present study, we aimed to address the role of non-CSC-to-CSC conversions by determining their frequency in a cohort of breast cancer (BrCa) cell lines.

In fact, we find that non-CSC-to-CSC conversions occur frequently in certain subtypes of BrCa but not in others, and have uncovered a mechanism governing this transition. From a therapeutic standpoint, the plasticity that we describe suggests efforts to improve therapeutic outcome for cancer patients by specifically targeting CSCs must be further enhanced by coupling them with strategies designed to eliminate non-CSC-to-CSC interconversions or, at the very least, to eliminate the subpopulations of non-CSCs that are poised to become CSCs.

Results

CD44 status and tumorigenic potential

Cell surface antigens, such as CD44, CD24 and ESA, have been successfully used to isolate CSC-like populations from BrCa cell lines and primary tissues. Among these antigens, it is widely accepted that breast CSCs are contained exclusively in the $CD44^{hi}$ cell compartment (Al-Hajj et al., 2003; Fillmore and Kuperwasser, 2008; Mani et al., 2008; Visvader and Lindeman, 2012).

To test the notion that $CD44^{hi}$ status, on its own, would allow enrichment of CSCs from BrCa cell lines, we analyzed 5 basal BrCa lines (SUM149, SUM159, HCC38, HMLER and BPLER) and 3 luminal BrCa lines (MCF7, MCF7R and T47D). (In the clinic, luminal BrCa generally are less aggressive and hold a better prognosis, while basal BrCa behave in the opposite fashion). $CD44^{hi}$ and $CD44^{lo}$ populations were evident to various extents in all basal BrCa lines examined, while all luminal BrCa lines consisted only of $CD44^{lo}$ populations (Figure S1A, and data not shown). To compare the relative tumorigenic potentials of these subpopulations, we purified by FACS $CD44^{lo}$ fractions from the luminal lines and $CD44^{lo}$ and $CD44^{hi}$ populations from the basal lines. Purified populations were injected immediately into the mammary fat pads of NOD/SCID mice. In all these experiments, cell populations were only implanted in hosts if they were greater than 99.7% pure as judged by FACS (Figure S1A-C).

We found that $CD44^{lo}$ luminal lines required higher cell numbers and longer incubation times *in vivo* to generate tumors of equivalent size to those seeded by purified $CD44^{lo}$ basal lines (luminal: 1×10^6 cells and 12-16 weeks *in vivo*; basal: 5×10^5 cells and 6-10 weeks *in vivo*, Figure 1A and Table S1). In addition, basal $CD44^{hi}$-derived tumors were 3-40 fold larger than their $CD44^{lo}$ counterparts when equal cell numbers were injected (Figure 1A and Table S2). Limiting dilution analysis showed that basal $CD44^{hi}$ cell fractions were significantly enriched for CSC frequency compared to their $CD44^{lo}$ counterparts (~10-fold, Figure S1D and Table S3). These data demonstrate that $CD44^{hi}$ expression enriches for cells that naturally reside in basal BrCa cell lines and possess higher intrinsic tumor-initiating and growth potential. Moreover, they raised the question of how certain BrCa populations that apparently lacked all traces of tumor-initiating CSCs were able to generate tumors when injected into host mice.

$CD44^{hi}$ CSCs arising from basal $CD44^{lo}$ cell populations in vivo

In previous work, we demonstrated that non-CSCs derived from experimentally transformed human mammary basal epithelial cells (HMECs) could spontaneously generate *de novo* CSCs both *in vitro* and *in vivo* (Chaffer et al., 2011). In the present work, we first undertook to test the idea that non-CSC-to-CSC conversions occur frequently in a broad array of BrCa cell lines.

Accordingly, we used FACS to analyze the tumors described above that arose from basal or luminal $CD44^{lo}$ cells. We found that luminal $CD44^{lo}$–derived tumors comprised almost entirely $CD44^{lo}$ cells with a small but detectable subpopulation (avg. <0.32%) of $CD44^{hi}$ cells. This suggested that luminal BrCa cells apparently lacking $CD44^{hi}$ tumor-initiating

cells were nonetheless able to seed tumors by generating new CD44^{hi} cells, albeit at a low frequency. In marked contrast to the behavior of luminal cells, basal CD44^{lo}-derived tumors contained CD44^{hi} subpopulations ranging in size from 2-22% of tumor cells (Figure 1B). These findings indicated that basal CD44^{lo} populations efficiently generate CD44^{hi} populations *in vivo*, while luminal CD44^{lo} populations do so with dramatically lower efficiency.

Functional analysis of CD44^{hi} cells created in vivo

We next sought to demonstrate that CD44^{hi} cells arising *in vivo* from basal CD44^{lo} cells were functionally equivalent to CSCs that are naturally present in basal BrCa cell lines. To begin, we derived several cell lines from tumors arising from implanted CD44^{lo} basal cells (SUM149-, SUM159- and BPLER-CD44^{lo} tumors) depicted in Figure 1A, terming them ExV (reflecting their *ex vivo* derivation). Each of these tumor-derived ExV-cell lines contained both CD44^{lo} and CD44^{hi} cells (Figure 1B), which we termed ExV-CD44^{lo} and ExV-CD44^{hi} cells. We then used FACS to isolate pure (>99%) populations of ExV-CD44^{lo} and ExV-CD44^{hi} cells and immediately injected them orthotopically into NOD/SCID mice (Figure 1C and Figure S1E-G). In most cases, ExV-CD44^{hi} cells generated 13-23 fold larger tumors compared to their ExV-CD44^{lo} counterparts (SUM159 and BPLER ExV lines), and displayed increased tumor-initiating ability (SUM159 and SUM149 ExV lines) (Figure 1C and Table S4). These results illustrate that CD44^{hi} cells arising *in vivo* from basal CD44^{lo} cells behave much like the CSCs that are naturally present in basal BrCa cell populations, in that they exhibit elevated tumor-initiating and tumor growth potential relative to their CD44^{lo} counterparts.

ZEB1 drives CD44^{lo}-to-CD44^{hi} cellular plasticity

We then undertook to shed light on the mechanism(s) enabling the observed *in vivo* CD44^{lo}-to-CD44^{hi} conversions. Given the inherent difficulties of uncovering these mechanisms in an *in vivo* setting, we first sought mechanistic insights from an *in vitro* model system that we had previously developed, in which non-transformed basal mammary epithelial cells (HME-flopc cells) spontaneously undergo CD44^{lo}-to-CD44^{hi} conversions with high frequency (Chaffer et al., 2011). As demonstrated at the time, conversions of non-transformed immortalized human mammary epithelial cells closely paralleled the behavior of their corresponding transformed derivatives.

In this instance, we also drew from earlier work demonstrating that CD44^{hi} stem-like cells are more mesenchymal than their CD44^{lo} counterparts (Chaffer et al., 2011; Mani et al., 2008). This and subsequent work (Guo et al., 2012) demonstrated that passage through the cell-biological program termed the epithelial-mesenchymal transition (EMT) placed cells close to the epithelial stem cell (SC) state. (This EMT program is largely studied for its ability to confer mesenchymal traits on epithelial cells.) Accordingly, we purified CD44^{lo} and CD44^{hi} subpopulations from HME-flopc cells and confirmed that CD44^{hi} cells, which contained the SC-like cells, indeed resided in a more mesenchymal state than their more epithelial CD44^{lo} counterparts (Figure 2A). Subsequently, to identify a key mediator of non-CSC-to-CSC plasticity, we analyzed the expression of various transcription factors (EMT-

TFs) known to govern the EMT program. Here we found that *ZEB1* expression was significantly higher (~10-fold) in $CD44^{hi}$ compared to $CD44^{lo}$ cells (Figure 2B).

To determine whether ZEB1 contributed in a critical way to mediating the transition from the $CD44^{lo}$ to $CD44^{hi}$ state, we analyzed $CD44^{lo}$-to-$CD44^{hi}$ conversions in HME-flopc-$CD44^{lo}$ cells expressing either doxycycline-inducible control or *ZEB1*-targeted shRNAs. We first confirmed that, following doxycycline-mediated induction, each of these shRNA vectors was capable, on its own, of achieving 80 to 90% ZEB1 knockdown (ZEB1-kd) in $CD44^{hi}$ cells. We also noted no differences in cell proliferation between cell populations expressing the control shRNA and those expressing shRNAs targeting *ZEB1* mRNA (Figure S2A-B).

We proceeded to purify $CD44^{lo}$ cells expressing the various shRNA vectors by FACS, introduced them into 2D cultures, and used FACS to monitor resulting cultures propagated in the presence or absence of doxycycline over the next 16 days. In the continued presence of doxycycline, the ability of $CD44^{lo}$ ZEB1-kd cells to convert into the $CD44^{hi}$ state was reduced by 75% (sh1) and 67% (sh2) relative to cultures expressing the control shRNA (Figure 2C). However, when doxycycline was withdrawn at day 8, permitting cells to continue growing in the absence of doxycycline for an additional 8 days, FACS analysis showed that $CD44^{lo}$ ZEB1-kd cells soon regained their ability to convert to the $CD44^{hi}$ state (Figure 2C). As such, the ability of $CD44^{lo}$ cells to activate ZEB1 expression appeared to be an important determinant of their ability to enter into the $CD44^{hi}$ state and thus an important determinant of cell plasticity in this model system.

Modulation of $CD44^{lo}$-to-$CD44^{hi}$ conversions by the MIR200 family

The expression of the *ZEB1* gene is tightly regulated by an interactive network involving ZEB1 itself, its relative ZEB2, and members of the MIR200 family of microRNAs (Gregory et al., 2008; Wellner et al., 2009). Thus, ZEB1 can serve to repress expression of the MIR200 miRNAs, while the latter can inhibit both the function and/or reduce the stability of the mRNAs specifying ZEB1 and ZEB2; hence, these mutually antagonistic elements constitute a circuit that operates as a bi-stable switch, governing the residence of cells in either the mesenchymal or epithelial state. Accordingly, we confirmed that ZEB1-kd in our system resulted in a decrease in *ZEB2* mRNA and concomitant increases in *MIR200B* and *MIR200C* levels (Figure S2A and S4).

In light of the mutually antagonistic actions of ZEB1 and the MIR200 miRNAs, we determined whether addition of synthetic inhibitors or mimetics (chemically synthesized, single-stranded, modified RNAs) influenced spontaneous $CD44^{lo}$-to-$CD44^{hi}$ conversions in HME-flopc-$CD44^{lo}$ cells. Indeed, as we found, miR200b or miR200c inhibitors significantly increased the rate of $CD44^{lo}$ to $CD44^{hi}$ conversions (Figure 2D).

We next undertook to determine whether the ability of the MIR200 family to affect $CD44^{lo}$-to-$CD44^{hi}$ conversions derived largely from modulation of *ZEB1* transcript levels or, alternatively, from the involvement of other MIR200 targets. Here we found that in the presence of ZEB1-kd, the miR200c inhibitor was unable to provoke a $CD44^{lo}$-to-$CD44^{hi}$ conversion (Figure 2E).

Together, these various results highlighted two important points: 1) ZEB1 is a key mediator of spontaneous $CD44^{lo}$-to-$CD44^{hi}$ conversions in non-transformed HMECs, acting through repression of the *MIR200* family and, quite possibly, other still-unidentified targets; and 2) *MIR200* family modulation of $CD44^{lo}$-to-$CD44^{hi}$ conversion derives from effects on *ZEB1* mRNA levels.

We extended these observations by examining the consequences of *ZEB1* knockdown in HME-flopc-$CD44^{lo}$ cells that had previously been transformed with *SV40-Early Region* and oncogenic *RAS in vitro* (Figure 2F). We found similar dynamics observed previously in the untransformed HME-flopc cells operated in their transformed derivatives.

Together, these results demonstrated ZEB1 is a key mediator of $CD44^{lo}$-to-$CD44^{hi}$ conversions in both non-transformed and transformed HMECs *in vitro*, and supported the previously reported notion that the dynamics of epithelial versus mesenchymal plasticity are quite similar in hTERT-immortalized cells and their transformed derivatives (Chaffer et al., 2011).

ZEB1/MIR200c are differentially expressed in BrCa cell populations

To determine whether ZEB1 often functions as a key player in driving BrCa cell plasticity, we assessed whether ZEB1 and MIR200 expression was indicative of the $CD44^{lo}$ versus $CD44^{hi}$ state in a broader array of BrCa lines. To do so, we first analyzed ZEB1 protein expression in luminal, basal $CD44^{lo}$, and basal $CD44^{hi}$ BrCa cell lines. ZEB1 was not detected in all luminal lines; however, in basal BrCa cell lines, ZEB1 was detectable in both populations, being 4-fold higher in matched pairs of $CD44^{hi}$ compared to $CD44^{lo}$ cells (Figure 3A).

Given the tight regulatory loop between ZEB1 and the *MIR200* family, we next examined how ZEB1 protein levels are regulated in these BrCa cell lines. Accordingly, we analyzed *ZEB1*, *MIR200B* and *MIR200C* mRNA expression levels. We found luminal lines expressed very low levels of *ZEB1* mRNA and very high levels of *MIR200B/C*. Conversely, basal $CD44^{lo}$ cells generally expressed modest but nonetheless detectable levels of both *ZEB1* mRNA and *MIR200B/C* whereas basal $CD44^{hi}$ cells expressed high levels of *ZEB1* mRNA and low-to-absent *MIR200B/C* levels (Figure 3B-C). We further confirmed that these patterns of *ZEB1* and *MIR200B/C* differential expression were maintained in the basal ExV-$CD44^{lo}$ and ExV-$CD44^{hi}$ cell populations (Figure S3A).

These observations demonstrated that, as predicted from their known interactions, the expression of ZEB1 and *MIR200B/C* varies inversely in these various BrCa cell lines, and that the ZEB1 EMT-TF is expressed at far higher levels in basal BrCa cells, which represent a class of tumors that generally carry worse clinical prognosis. Moreover, they indicated that the expression pattern initially observed in the HME-flopc cells lines was closely echoed by the human basal BrCa lines examined.

Role of ZEB1-mediated $CD44^{lo}$-to-$CD44^{hi}$ conversions in tumor initiation and growth

The data implicating ZEB1 in $CD44^{lo}$-to-$CD44^{hi}$ cell plasticity *in vitro* did not shed light on whether it plays a similar role *in vivo*, specifically in basal BrCa cells. Consequently, we

introduced the same doxycycline-inducible control or ZEB1-targeted shRNAs used earlier into cell populations of the HCC38 and SUM159 human breast cancer cell lines and the experimentally transformed HMLER cells (Figure S3B). We confirmed that ZEB1-kd achieved in the presence of doxycycline did not affect cell proliferation rates in monolayer culture (Figure S3C). FACS-purified $CD44^{lo}$ populations of control or doxycycline-induced ZEB1-kd cells were then injected orthotopically into NOD/SCID mice immediately following FACS-purification in order to analyze the effects of ZEB1-kd on $CD44^{lo}$-to-$CD44^{hi}$ conversions *in vivo* and on tumorigenicity. Animals were administered doxycycline (2g/1L) for the duration of the experiment.

Strikingly, HMLER-$CD44^{lo}$ cells gave rise to tumors of 0.1g on average after 8-10 weeks of growth *in vivo*, while their ZEB1-kd counterparts failed to form tumors (Figure 3D), indicating that ZEB1-mediated $CD44^{lo}$-to-$CD44^{hi}$ conversions were essential for tumor-initiating potential. That is, the ability of HMLER-$CD44^{lo}$ carcinoma cells to initiate tumors appeared to depend critically on the ability of these cells to spontaneously generate CSCs *in vivo*, which depended in turn on their ability to activate expression of their own endogenous *ZEB1* gene. Similar results were obtained with HCC38-$CD44^{lo}$ cells. In SUM159-$CD44^{lo}$ ZEB1-kd cells, tumor size was significantly decreased (0.39g average for controls; 0.03g and 0.18g on average for sh1 and sh2 respectively), and tumor-initiating potential was decreased from 100% in control cells to 67% in ZEB1-kd cells. Together, these data demonstrate that the ability of $CD44^{lo}$ basal BrCa cell populations to up-regulate ZEB1 expression is generally a critical determinant of their tumor-initiating potential and overall tumor growth.

Role of ZEB1 in the $CD44^{hi}$ stem-cell state

Given the importance of ZEB1 in enabling spontaneous $CD44^{lo}$-to-$CD44^{hi}$ conversions, we wondered whether the continued expression of ZEB1 was required thereafter for maintenance of the resulting $CD44^{hi}$ cell state; alternatively, other regulatory loops might become activated that then obviate the need for high ZEB1 expression for initial entrance into the $CD44^{hi}$ state. Accordingly, we used FACS to monitor the ability of purified nontransformed HME-flopc-$CD44^{hi}$ cells to maintain their $CD44^{hi}$ marker profile in culture over a 16 day period in the presence or absence of ZEB1-kd. Interestingly, we found that $CD44^{hi}$ ZEB1-kd cells maintained their $CD44^{hi}$ marker profile (Figure 4A and S4).

Next we functionally tested the SC activity of $CD44^{hi}$ ZEB1-kd cells by mammosphere-forming ability in 3D culture (Dontu et al., 2003). We found mammosphere formation was reduced by 80-99% compared to control. Furthermore, addition of miR200 inhibitors to $CD44^{hi}$ ZEB1-kd cells could not restore mammosphere-forming ability (Figure 4B). Together, these data demonstrate that ZEB1 is required for initial acquisition of both $CD44^{hi}$ expression and stem-like activity of $CD44^{hi}$ cells and subsequent maintenance of SC activity, but is not required for long-term maintenance of high cell-surface CD44 expression.

We further examined the effect of ZEB1-kd in HME-flopc-$CD44^{hi}$ cells that had been transformed by the introduction of *SV40-Early Region* and oncogenic *RAS* genes, i.e., the HMLER-flopc cells. Similarly, FACS analysis confirmed that transformed $CD44^{hi}$ ZEB1-kd cells maintained their $CD44^{hi}$ phenotype (data not shown). We also observed that ZEB1-kd

decreased tumorsphere formation, an *in vitro* surrogate measure of CSC-like activity (60% inhibition, $p<0.003$), and significantly reduced tumor burden *in vivo* ($p<0.05$) (Figure 4C-D).

To summarize, together with our earlier results, these showed that ZEB1 is required for conversion from the $CD44^{lo}$ to $CD44^{hi}$ state and also for maintenance of $CD44^{hi}$ stem-like/CSC-like activity. Once cells are residing in the $CD44^{hi}$ state, however, ZEB1 and CD44 expression can be uncoupled, in that cells with ZEB1-kd functionally lose their stem-like/CSC features while still maintaining high CD44 expression. Stated differently, these data suggest that $CD44^{hi}$ cells can constitute heterogeneous cell populations, in which $CD44^{hi}ZEB1^{+}$ signifies great enrichment of cells residing in a CSC-like state, whereas $CD44^{hi}ZEB1^{-}$ signifies cells that are non-CSCs.

Phenotypic plasticity and chromatin configuration of the *ZEB1* promoter

The above-described experiments provided clear indication that ZEB1 function is necessary for the generation of $CD44^{hi}$ stem-like cells from $CD44^{lo}$ cells. Still, these observations did not provide insight into why ZEB1 was readily induced in basal $CD44^{lo}$ cells but not in luminal $CD44^{lo}$ cells. We reasoned that current models of epigenetic regulation might illuminate these differences in behavior, as global epigenetic differences have been observed between luminal and basal-type BrCa (Maruyama et al., 2011). More specifically, we chose to use chromatin immunoprecipitation followed by quantitative real-time PCR (ChIP-qPCR) to examine the chromatin state at the *ZEB1* promoter.

The functional state of chromatin has been defined largely by patterns of covalent modifications to the N-terminal domains of histones and is indicative of transcriptional activity. Thus, trimethylation of lysine 4 of the histone H3 subunit (H3K4me3) is associated with transcriptional initiation (Guenther et al., 2007), dimethylation of lysine 79 of the same subunit (H3K79me2) is associated with transcriptional elongation (Mueller et al., 2007), and the combination of H3K4me3 and H3K79me2 indicates an actively transcribed gene. In contrast, trimethylation of lysine 27 (H3K27me3) is often associated with transcriptional repression mediated by the Polycomb group of proteins (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002).

In embryonic stem (ES) cells, the promoters of many genes encoding key developmental regulators are associated with both the permissive H3K4me3 and the restrictive H3K27me3 modifications (yielding so-called ‘bivalent’ domains) (Bernstein et al., 2006). This bivalent chromatin is thought to keep these genes repressed, but nonetheless poised for rapid transcriptional activation in response to subsequent signaling decisions favoring differentiation. These findings indicate that this combination of histone modifications may be a signature of regulators that are required to rapidly switch cell state. Indeed, through publically available databases we determined that ZEB1 maintains a bivalent chromatin configuration in ES cells (Figure S5A).

We speculated that ZEB1 might exhibit a bivalent chromatin state in those cells in which it was possible to switch between low and high ZEB1 expression. To pursue this notion, we first analyzed purified $CD44^{lo}$ or $CD44^{hi}$ HME-flopc populations. $CD44^{hi}$ cells displayed

chromatin methylation patterns indicating active transcription at the *ZEB1* promoter as determined by the presence of both H3K4me3 and H3K79me2 marks (Figure 5A-B). In contrast, we found that CD44^{lo} cells exhibited bivalent chromatin methylation patterns at the *ZEB1* promoter, as determined by the presence of both H3K4me3, and H3K27me3. These data indicated that immortalized untransformed basal mammary epithelial cells with the ability to spontaneously switch to a CD44^{hi} stem-like state do indeed maintain their *ZEB1*-promoter in a poised, bivalent configuration.

We wished to extend these observations to human BrCa cell lines, including luminal, basal CD44^{lo} and basal CD44^{hi} cell populations. Our previous data had shown that luminal cell lines express very low levels of *ZEB1* mRNA (Figure 3A-B). Perhaps unsurprisingly, we found that all luminal cell lines exhibited only repressive chromatin methylation patterns at *ZEB*1, as determined by the presence of H3K27me3 and the relative absence of both H3K4me3 and H3K79me2 (Figure 5C). Further, we found that the chromatin at the *ZEB1* promoter in basal CD44^{hi} cells is characterized by the presence of H3K4me3 and H3K79me2 and the relative absence of H3K27me3, indicating active transcription. These chromatin modifications conform with our earlier expression data indicating that all basal CD44^{hi} cells express high levels of ZEB1 (Figure 3A-C).

Provocatively, we found that the *ZEB1* promoter in basal CD44^{lo} BrCa cells resided in a bivalent chromatin state, where both H3K4me3 and H3K27me3 modifications were detected, much as we had found in the immortalized CD44^{lo} HME cells. To demonstrate that both histone modifications reside simultaneously in specific regions in the basal CD44^{lo} BrCa cells, we performed sequential-ChIP analysis for H3K4me3 followed by H3K27me3 ChIP, as well as the reverse ChIP experiment (Figure S5B). Together, these results demonstrate that both immortalized and neoplastic basal CD44^{lo} populations maintain ZEB1 in a unique bivalent chromatin state, poised for activation, and provided a mechanistic insight as to why *ZEB1* is readily induced in basal CD44^{lo} cells but not in luminal CD44^{lo} cells, which maintain the *ZEB1* promoter in a repressed state.

Microenvironmental stimuli and CD44^{lo}-to-CD44^{hi} conversions

Extensive evidence indicates that activation of the EMT program and entrance into a stem cell state is generally triggered by contextual signals received by normal and neoplastic epithelial cells (Mani et al., 2008; Thiery et al., 2009). Among these signals, TGFbeta has been shown to potently upregulate ZEB1 expression (Gregory et al., 2011). Accordingly, we examined whether TGFbeta could enhance the spontaneous CD44^{lo}-to-CD44^{hi} transitions of immortalized HMECs. Indeed, TGFbeta induced a dose-dependent increase, while the SB431542 TGFbeta receptor inhibitor inhibited such conversions. In the context of ZEB1-kd, however, the ability of TGFbeta to induce transitions was abolished (Figure 6A and S6A-B). Similarly, we also found that transformed HME-flopc-CD44^{lo} cells could not transition from the CD44^{lo} to CD44^{hi} state in response to TGFbeta if ZEB1 expression was inhibited (Figure 6B). Hence, TGFbeta can enhance CD44^{lo}-to-CD44^{hi} transitions in both normal and transformed cells in a fashion that is dependent upon induction of ZEB1.

To further confirm that TGFbeta-driven CD44^{lo}-to-CD44^{hi} conversions were achieved through ZEB1, we showed that *ZEB1* mRNA was induced and inhibited by TGFbeta and

SB431542 treatment respectively (Figure S6C). As controls, we found that TGFbeta target genes (PAI-1 and GADD45B) were induced and repressed in these cells in response to modulation of TGFbeta signaling (Figure S6C), confirming that overall TFGbeta signaling operated in these cells as anticipated.

We extended these findings to a series of human BrCa cell lines and speculated that basal $CD44^{lo}$ cells, but not luminal $CD44^{lo}$ cells, would readily transit to a $CD44^{hi}$ state in response to TGFbeta. To test this notion, luminal (MCF7Ras and ZR-75-1) and basal (HMLER and HCC38) $CD44^{lo}$ cells were treated for 4 days with TGFbeta (2 or 20 ng/ml). FACS analysis at Day 4 showed that the luminal cells did not generate any $CD44^{hi}$ cells in response to TGFbeta treatment, while the basal $CD44^{lo}$ cells responded in a dose-dependent manner to TGFbeta by generating $CD44^{hi}$ cells (Figure 6C). We further demonstrate that TGFbeta stimulation led to an increase in ZEB1 protein levels (Figure S6E). Together, these data indicate that one important contextual signal - TGFbeta - enhances the rate of $CD44^{lo}$-to-$CD44^{hi}$ transitions in basal breast cancer $CD44^{lo}$ cell populations, and that this response is dependent upon induction of ZEB1 expression.

Modulation of the chromatin status at the ZEB1 promoter by TGFbeta

Having shown that TGFbeta induces $CD44^{lo}$-to-$CD44^{hi}$ conversions in basal $CD44^{lo}$ cells, we examined whether this effect coincided directly with changes in the histone modification patterns at the *ZEB1* promoter. To do so, we performed ChiP-qPCR on basal HME-flopc-$CD44^{lo}$ cells expressing doxycycline-inducible *ZEB1*-shRNA treated with Control (HCl), TGFbeta (2ng/ml) or the SB431542 TGFbeta receptor inhibitor (10 μM).

Performing this experiment in a ZEB1-kd intracellular environment was essential in order to maintain a homogeneous $CD44^{lo}$ population; otherwise, TGFbeta treatment would cause the $CD44^{lo}$ cells to transition to a $CD44^{hi}$ state, in which case, the $CD44^{hi}$ cells harboring active histone modifications at the *ZEB1* promoter would mask any changes occurring specifically in $CD44^{lo}$ cells.

ZEB1 knockdown in basal $CD44^{lo}$ cells was induced by exposure to doxycycline for 5 days and cells were subsequently treated with TGFbeta, SB431542, or control (PBS or DMSO). After a further 5 days in the continued presence of doxycycline, FACS analysis confirmed that the $CD44^{lo}$-ZEB1-kd cells remained as pure $CD44^{lo}$ populations (Figure 6D)

We then performed ChIP-qPCR at the *ZEB1* promoter to compare changes in the levels of histone modifications in control versus TGFbeta or SB531542 treatment. While no significant differences in the H3K4me3 or H3K79me2 methylation patterns were observed across all treatment groups, we found that TGFbeta treatment did indeed lead to a marked decrease in the repressive H3K27me3 mark associated with the *ZEB1* promoter in basal $CD44^{lo}$ cells (Figure 6E). Together, these data demonstrate that TGFbeta enables cells to transition from the bivalent chromatin status to active chromatin state at the *ZEB1* promoter, doing so, at least in part, through the removal of the H3K27me3 repressive mark.

Wishing to put into context these changes in the histone modifications at the Z*EB1* promoter, we also followed changes in histone marks following control, TGFbeta or

SB531542 treatment at the promoters of two TGFbeta-responsive genes (GADD45B, PAI-1) and a TGFbeta-non-responsive gene (HPRT1). We found that the magnitude of changes in histone modifications causing active transcription at known TGFbeta target genes closely correlated with the changes that we had previously observed at the *ZEB1* promoter in response to TGFbeta treatment (Figure S6D). These provide further evidence that *ZEB1* is a bona fide TGFbeta target gene in these cells.

Intrinsic responsiveness of luminal BrCa cells to exogenous ZEB1

Knowing that luminal cells are indeed responsive to TGFbeta treatment (as determined by TGFbeta-mediated up-regulation of pSMAD2 (Figure S6E)), we reasoned that the inability of luminal cells to undergo a $CD44^{lo}$-to-$CD44^{hi}$ switch in response to TGFbeta treatment might be due to their inability to activate *ZEB1* transcription or, quite possibly, to an intrinsic lack of responsiveness of these cells to ZEB1 signaling. To explore these alternatives, we forced expression of ZEB1 in MCF7Ras cells (Figure S6F). We observed a progressively increasing population of $CD44^{hi}$ cells in MCF7R-ZEB1 cells over a 2-week time course (Figure 6F). Furthermore, we found that ZEB1 overexpression increased tumorsphere formation *in vitro* and tumorigenicity *in vivo*, in a dose-dependent manner (Figure 6G-H). These data indicate that while the endogenous *ZEB1* promoter in these luminal cells is repressed, they are nevertheless intrinsically responsive to this EMT-TF if its expression is forced.

Assessment of ZEB1 and MIR200B/C in clinical cases of breast cancer

As described above, we found that ZEB1 pays an important role in promoting $CD44^{lo}$-to-$CD44^{hi}$ conversions and in maintaining the CSC-like state in cells that already reside in the $CD44^{hi}$ state. Indeed, both of these processes contribute to enhanced tumor initiation and growth. We were interested in relating these observations to the properties of clinical cases of breast cancer. To pursue this question, we accessed data from the Cancer Genome Atlas Network (CGAN, 2012).

We first assessed the relative abundance of MIR200 family members across all breast tumors represented in the CGAN database and found that *MIR200C* accounts for 93% (sd=5%) of mature miRNA in the MIR200BC family, with minority representation of the related *MIR200B* (6%) and *MIR429* (1%) family members (Figure 7A); given the dominant presence of *MIR200C* over its other family members, this allowed us to focus subsequent measurements on levels of *MIR200C*. We also found that the levels of *ZEB1* and *MIR200C* expression are inversely correlated in basal ($p=9.4\times10^{-4}$; $r^2=0.13$), luminal A ($p=2.8\times10^{-4}$; $r^2=0.07$) and luminal B ($p=6.8\times10^{-4}$; $r^2=0.12$) subtypes but not in *HER2*-overexpressing BrCa cells (Figure 7B). We then compared the absolute abundance of ZEB1 and *MIR200C* in BrCa subtypes and found that surprisingly, in contrast to the observations described above, *ZEB1* mRNA appeared to be significantly more abundant in normal and luminal A BrCa subtypes compared to the basal subtype (Figure 7C).

Given our earlier observations in cultured cells that luminal BrCa lines do not express *ZEB1* and that the chromatin at the ZEB1 promoter resides in a repressed state in those same lines, we reasoned that this apparent conflict with the relatively high levels of *ZEB1* and *MIR200C*

mRNA in clinical cases of Luminal A-type BrCa might be explained in either of two ways: our studies of cultured cancer cell lines failed to properly reflect the behavior of corresponding cells in living tissues, or the data in the CGAN database was confounded by strong contamination of carcinoma cells with adjacent stromal cells expressing high levels of *ZEB1*.

To resolve between these alternatives, we analyzed ZEB1 protein expression in a tissue microarray of breast cancer biopsies (Figure 7D and Figure S7). We found that ZEB1 protein is present at high levels in the stromata of all breast cancer subtypes. Interestingly, however, comparison of ZEB1 protein specifically in cancer cells showed that triple-negative (TN) BrCa cells have significantly higher levels of ZEB1 protein compared to Luminal A cancer cells (Figure 7D, $p = 0.017$). In fact, ZEB1 was not present in the carcinoma cells of any Luminal A BrCa sample examined (0/91 samples). From these data, we concluded that the strong *ZEB1*-positive signature produced by the TCGA analysis data in Luminal A BrCa compared to TN BrCa (Figure 7C) was entirely attributable to stromal cells present as significant contaminants in the Luminal A samples. Our data further highlight the difficulty of interpreting such global genomic analyses performed on whole tumor digests, where the relative representations of both carcinoma cells and stromal cells cannot be accounted for.

Together with our demonstration in BrCa cell lines that ZEB1 is the driver of the *de novo* generation of CSCs from non-CSC cell populations, the high expression of ZEB1 in TN BrCa cells raises the possibility that the more aggressive nature of clinical TN-type BrCa compared to Luminal-type BrCa may be, in part, attributable to the ability of TN BrCas to readily activate *ZEB1* expression in response to microenvironmental stimuli and subsequently, to create a continuous source of highly tumorigenic CSCs.

Discussion

The present work reveals that the dynamics of interconversion between epithelial non-CSC and mesenchymal/CSC states are important determinants of normal and neoplastic epithelial tissue behavior. In contrast to the widely accepted CSC model, where CSCs give rise to non-CSC progeny in a unidirectional manner, we have demonstrated that in certain carcinoma subtypes, notably basal carcinomas of the breast, neoplastic cells can readily convert from a $CD44^{lo}$ to a $CD44^{hi}$ state (Figure 7E). Given the strong enrichment of CSCs in the $CD44^{hi}$ state and their virtual absence in $CD44^{lo}$ cells (Al-Hajj et al., 2003), this suggested an interconversion between non-CSCs and CSCs, as indeed we demonstrated directly.

Our analyses indicate that this plasticity is not a universal property of all breast carcinomas. Plastic behavior was associated with 4/5 basal-type BrCa cell lines, while luminal $CD44^{lo}$ populations (4/4) were extremely inefficient at switching from the $CD44^{lo}$-to-$CD44^{hi}$ phenotype. These findings demonstrate fundamental differences in the biology driving basal-versus luminal- type tumors and suggest that the well-documented aggressive behavior of basal-type BrCa's may be traced, in no small part, to this plasticity and the associated ability to generate carcinoma cells with enhanced tumor-initiating powers. The discovery of ZEB1, a well-characterized EMT-TF, as a key mediator of $CD44^{lo}$-to-$CD44^{hi}$ plasticity is

consistent with the idea that the epithelial-mesenchymal transition generates cells with CSC-like activity (Mani et al., 2008; Morel et al., 2008). While we do not rule out the possibility that other EMT transcription factors functioning upstream of ZEB1 may also drive non-CSC-to-CSC conversions, we have demonstrated that ZEB1 plays a critical, rate-limiting role in governing basal BrCa cell plasticity.

We found that the chromatin configuration associated with the *ZEB1* promoter in luminal $CD44^{lo}$ cells was repressed, while in basal $CD44^{lo}$ cells it was maintained in a bivalent/poised configuration, corresponding with the respective inability and ability of these two BrCa cell types to generate *de novo* CSC-like cells. As we argue here, differences in chromatin configuration appear to be responsible for the profound differences in cell plasticity. This yields, in turn, the interesting notion that the aggressiveness of certain breast carcinomas may not be determined by their steady-state concentrations of $CD44^{hi}$ stem-like cells; instead, their content of non-CSC cells with a proclivity to readily spawn $CD44^{hi}$ stem-like derivatives may strongly influence the overall malignant behavior of these tumors. Stated differently, the bivalency of the *ZEB1* promoter in carcinoma cells may represent a useful prognostic parameter of tumor aggressiveness, a notion that will require extensive clinical testing and validation. Since such bivalent chromatin is already present in certain immortalized, non-transformed human mammary epithelial cells, this might suggest that the establishment of such bivalency occurs during the normal ontogeny of this lineage differentiation.

The nature of *ZEB1*-associated poised chromatin in basal $CD44^{lo}$ cells suggests that those cells may readily and efficiently re-enter a stem-like state given the appropriate stimulus. Indeed, we demonstrated that TGFbeta, a well-known EMT-inducing stimulus (Gregory et al., 2011), can efficiently promote non-CSC-to-CSC conversion. The same TGFbeta stimulus failed to induce luminal $CD44^{lo}$ cells to convert to the CSC state (Figure 7E). Here, bivalency associated with the *ZEB1* promoter permits basal-type non-CSCs to respond to the same stimulus in a qualitatively different manner than luminal type non-CSCs. In that regard, it is plausible that basal non-CSCs located in an inflammatory microenvironment that is rich in EMT-inducing heterotypic signals may respond to local stimuli by switching to a $CD44^{hi}$ CSC state; the resulting cells may then significantly enhance the aggressiveness of the tumors in which they reside.

At present, it seems plausible that disseminating CSCs are the principal agents of metastasis, as they are endowed with multiple traits that are essential for completion of most of the steps of the invasion-metastasis cascade (Thiery et al., 2009). However, in light of the plasticity that we can now ascribe to basal $CD44^{lo}$ non-CSCs, it is conceivable that they too may leave a primary tumor and, following arrival at secondary tissue sites, create pools of newly formed CSCs that are critical to subsequent spawning of metastatic deposits. If validated, this would suggest that certain tumors are clinically aggressive because they can dispatch non-CSCs (which are usually far more numerous than CSCs) to distant sites as founders of new metastatic colonies following non-CSC-to-CSC conversions. In the same manner, plastic basal-$CD44^{lo}$ cells may also contribute to disease recurrence. These and other considerations suggest that therapies directed at preventing non-CSC-to-CSC conversions

should be considered as essential components of adjuvant therapies for breast cancer patients and, quite possibly, patients suffering other types of neoplastic disease.

Experimental Procedures

Animals

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

Cell culture

Cells were cultured as described in Table S5.

Vectors and viral infections

pBabe SV40-ER (Zeocin), pBabe H-Ras (Puromycin), PRRL-GFP, pLV-Tomato vectors, production of virus and infection of target cells have been previously described (Elenbaas et al., 2001; Shaner et al., 2004). Infected cells were selected with Zeocin (100 μg/ml) and Puromycin (2 μg/ml). shRNA were purchased from Open Biosystems.

RNA preparation and qRT-PCR analysis

Total RNA was isolated using the RNeasy Micro kit (Qiagen). Reverse transcription was performed with miScript II RT Kit, miScript and Qantitect Primer Assays were used to detect miRNAs and mRNA (Qiagen).

ChiP-qPCR

Chromatin immunoprecipitation was performed as previously described (Lee et al., 2006).

Mammosphere culture

Mammosphere culture was performed as previously described (Dontu et al., 2003).

Statistical analysis

Data are presented as mean ± SEM. Student's t test was used to compare two groups ($p < 0.05$ was considered significant) unless otherwise indicated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Advanced Medical Research Foundation (AMRF), the Breast Cancer Research Foundation (RAW) and NIH grants HG002668 (RAY) and CA146445 (RAY and TIL). RAW is an American Cancer Society Research Professor and a Daniel K. Ludwig Foundation Cancer Research Professor. The authors

thank the core Flow Cytometry Facilities at Whitehead Institute for Biomedical Research and The Koch Institute of MIT and Elinor Ng-Eaton for technical support, and Paul Bernath for critical review of the manuscript.

References

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003; 100:3983–3988. [PubMed: 12629218]

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125:315–326. [PubMed: 16630819]

Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997; 3:730–737. [PubMed: 9212098]

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science. 2002; 298:1039–1043. [PubMed: 12351676]

CGAN. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490:61–70. [PubMed: 23000897]

Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, Brooks M, Reinhardt F, Su Y, Polyak K, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A. 2011; 108:7950–7955. [PubMed: 21498687]

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell. 2002; 111:185–196. [PubMed: 12408863]

Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nat Rev Cancer. 2005; 5:275–284. [PubMed: 15803154]

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, Wicha MS. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. 2003; 17:1253–1270. [PubMed: 12756227]

Elenbaas B, Spirio L, Koerner F, Fleming MD, Zimonjic DB, Donaher JL, Popescu NC, Hahn WC, Weinberg RA. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. 2001; 15:50–65. [PubMed: 11156605]

Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast cancer research : BCR. 2008; 10:R25. [PubMed: 18366788]

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008; 10:593–601. [PubMed: 18376396]

Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S, Morris M, Wyatt L, Farshid G, Lim YY, et al. An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. Mol Biol Cell. 2011; 22:1686–1698. [PubMed: 21411626]

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007; 130:77–88. [PubMed: 17632057]

Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, Itzkovitz S, Noske A, Zurrer-Hardi U, Bell G, et al. Slug and Sox9 cooperatively determine the mammary stem cell state. Cell. 2012; 148:1015–1028. [PubMed: 22385965]

Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011; 146:633–644. [PubMed: 21854987]

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 2002; 16:2893–2905. [PubMed: 12435631]

Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc. 2006; 1:729–748. [PubMed: 17406303]

Malanchi I, Santamaria-Martinez A, Susanto E, Peng H, Lehr HA, Delaloye JF, Huelsken J. Interactions between cancer stem cells and their niche govern metastatic colonization. Nature. 2012; 481:85–89. [PubMed: 22158103]

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008; 133:704–715. [PubMed: 18485877]

Maruyama R, Choudhury S, Kowalczyk A, Bessarabova M, Beresford-Smith B, Conway T, Kaspi A, Wu Z, Nikolskaya T, Merino VF, et al. Epigenetic regulation of cell type-specific expression patterns in the human mammary epithelium. PLoS Genet. 2011; 7:e1001369. [PubMed: 21533021]

Morel AP, Lievre M, Thomas C, Hinkal G, Ansieau S, Puisieux A. Generation of breast cancer stem cells through epithelial-mesenchymal transition. PLoS One. 2008; 3:e2888. [PubMed: 18682804]

Mueller D, Bach C, Zeisig D, Garcia-Cuellar MP, Monroe S, Sreekumar A, Zhou R, Nesvizhskii A, Chinnaiyan A, Hess JL, et al. A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood. 2007; 110:4445–4454. [PubMed: 17855633]

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell. 2002; 111:197–208. [PubMed: 12408864]

Roesch A, Fukunaga-Kalabis M, Schmidt EC, Zabierowski SE, Brafford PA, Vultur A, Basu D, Gimotty P, Vogt T, Herlyn M. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell. 2010; 141:583–594. [PubMed: 20478252]

Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol. 2004; 22:1567–1572. [PubMed: 15558047]

Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. Cell. 2009; 139:871–890. [PubMed: 19945376]

Visvader JE, Lindeman GJ. Cancer stem cells: current status and evolving complexities. Cell Stem Cell. 2012; 10:717–728. [PubMed: 22704512]

Wellner U, Schubert J, Burk UC, Schmalhofer O, Zhu F, Sonntag A, Waldvogel B, Vannier C, Darling D, zur Hausen A, et al. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nat Cell Biol. 2009; 11:1487–1495. [PubMed: 19935649]

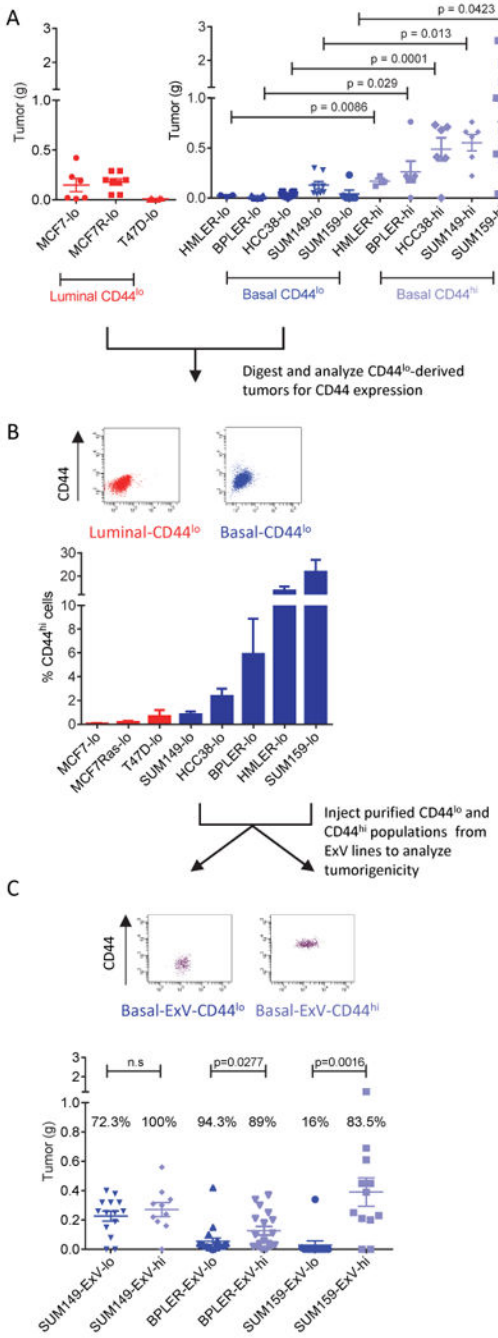


Figure 1. Basal breast cancer CD44^{lo} non-CSC cell populations spontaneously switch to a CD44^{hi} CSC state in vivo

(A) Tumorigenicity of FACS-purified luminal BrCa CD44^{lo} cells or basal BrCa CD44^{lo} and CD44^{hi} cell populations following orthotopic injection into NOD/SCID mice (n ≥ 6/group).

(B) Representative FACS plots for CD44 expression and quantification of CD44^{hi} cells generated from luminal or basal CD44^{lo}-derived tumors generated in (A).

(C) Basal CD44^{lo} digested tumors from (B) were cultured *in vitro* to generate *ex vivo* cell lines (ExV). ExV lines were purified by FACS into CD44^{lo} and CD44^{hi} components and

injected orthotopically into NOD/SCID mice (n ≥ 8/group). Tumor incidence displayed as percentages on the graph.

Data represented as mean ± SEM. See also Figure S1.

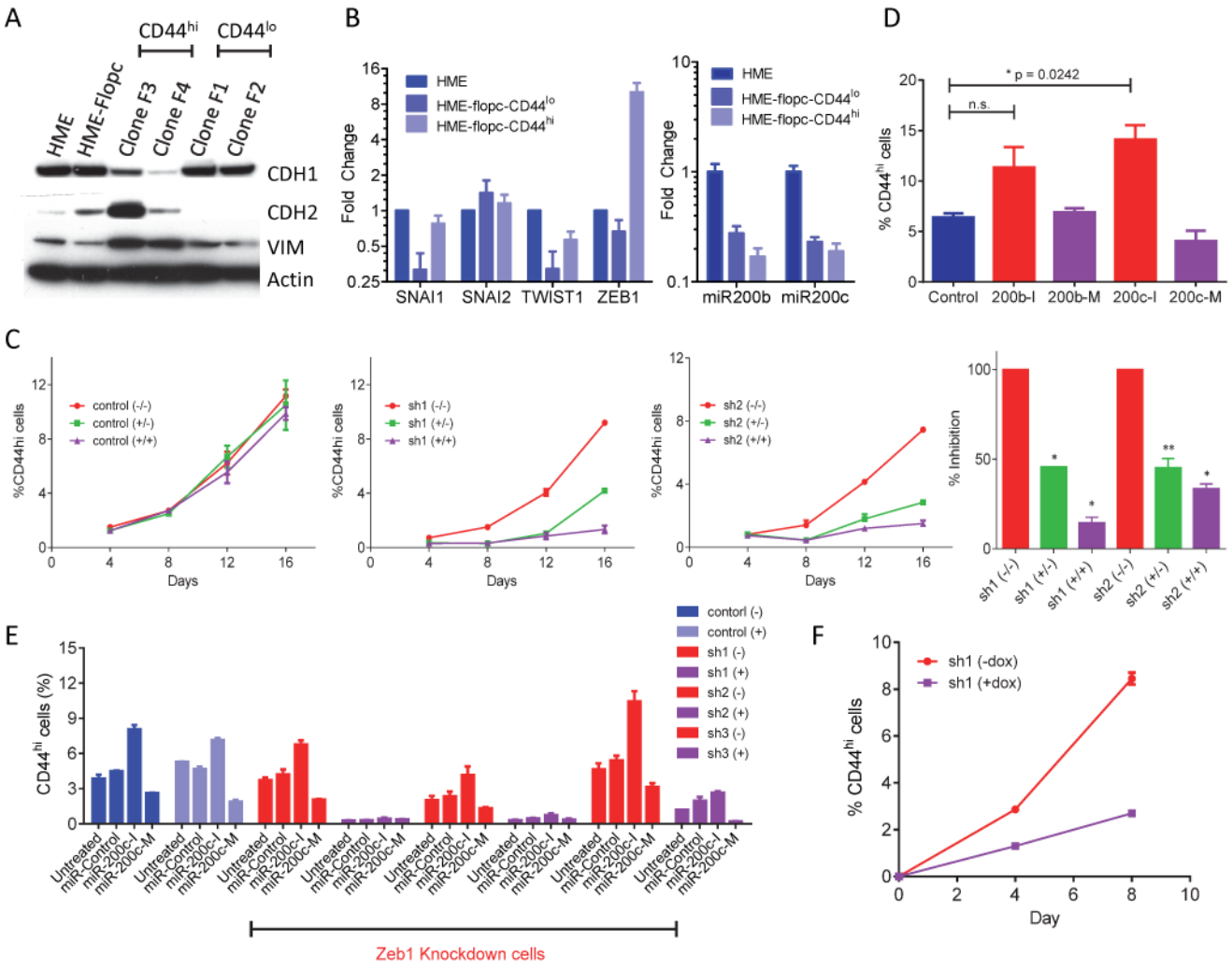


Figure 2. ZEB1 is an essential mediator of CD44^{lo}-to-CD44^{hi} cell transitions

(A) Western blot for markers of the epithelial (CDH1) or mesenchymal (CDH2, VIM) phenotype in immortalized human mammary epithelial cells (HME), HME-flopc and single cell clones derived from HME-flopc population enriched for the CD44^{lo} phenotype (clones F1 and F2) or CD44^{hi} phenotype (clones F3 and F4).

(B) qPCR for EMT transcription factors and *MIR200B/C* in non-transformed CD44^{lo} (HME and HME-flopc-CD44^{lo}) or HME-flopc-CD44^{hi} cells.

(C) FACS analysis for the ability of HME-flopc-CD44^{lo} cells to switch to the CD44^{hi} cell state. Cells express a doxycycline (dox) inducible control shRNA (control) or shRNA targeting *ZEB1* (sh1 and sh2). -/- no dox, +/- dox on for 8 days then removed for the remaining 8 days, +/+ dox on for the duration of the experiment. Inhibition (%) at day 16 is also shown (*$p < 0.0001$, **$p < 0.0008$, compared to (-/-).

(D) Purified HME-flopc-CD44^{lo} cells treated with *MIR200B/C* inhibitors (I) or mimetics (M) to determine effects on switching from CD44^{lo} to CD44^{hi} cell state. Data are mean ± SEM.

(E) Purified HME-flopc-CD44^{lo} cells expressing a dox-inducible control shRNA (control) or shRNA targeting *ZEB1* (sh1, sh2 or sh3) were analyzed for their ability to switch to the CD44^{hi} state in the presence (+) or absence (-) of dox, and in response to a *MIR200C* inhibitor (I) or mimetic (M).

(F) Transformed HME-flopc-CD44^{lo} cells (with *SV40-Early Region* and *RAS* oncoprotein) expressing a dox-inducible shRNA targeting ZEB1 (sh1) were analyzed for conversion to the CD44^{hi} state in the presence (+) or absence (-) of ZEB1-knockdown. Cells were monitored by FACS for 8 days. Data represented as mean ± SEM. See also Figure S2.

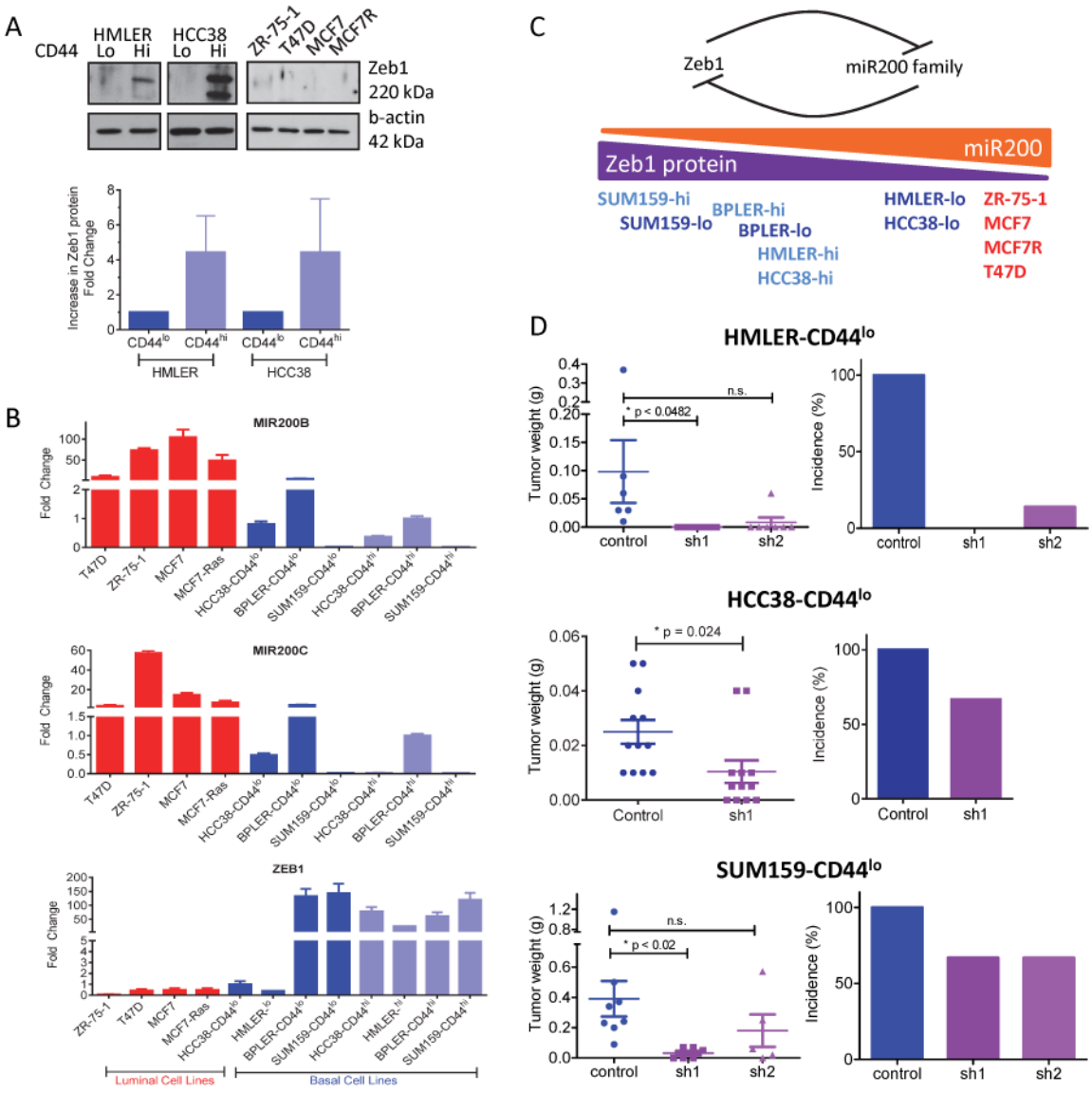


Figure 3. Inhibition of CD44^{lo}-to-CD44^{hi} conversions by blocking ZEB1 decreases tumorigenicity

(A) Western blot comparing the expression of ZEB1 in basal BrCa cell lines (HMLER and HCC38) purified for CD44^{lo} or CD44^{hi} subpopulations, and luminal BrCa cell lines (ZR-75-1, T47D, MCF7 and MCF7R). Quantification of differential ZEB1 expression in basal cell lines (n = 4).

(B) qPCR assessing *ZEB1*, *MIR200B* and *MIR200C* mRNA expression in BrCa cell lines.

(C) Schematic illustrating expression of *MIR200* family members and ZEB1 protein expression in basal (CD44^{lo} and CD44^{hi} subpopulations) and luminal BrCa cell lines.

(D) Purified CD44^{lo} cells from HMLER or HCC38 and SUM159 basal BrCa cell lines expressing dox-inducible control shRNA (control) or shRNA targeting *ZEB1* (sh1 and sh2) were analyzed for tumorigenic potential. Final tumor mass and incidence are represented (n ≥ 5/group). Data represented as mean ± SEM. See also Figure S3

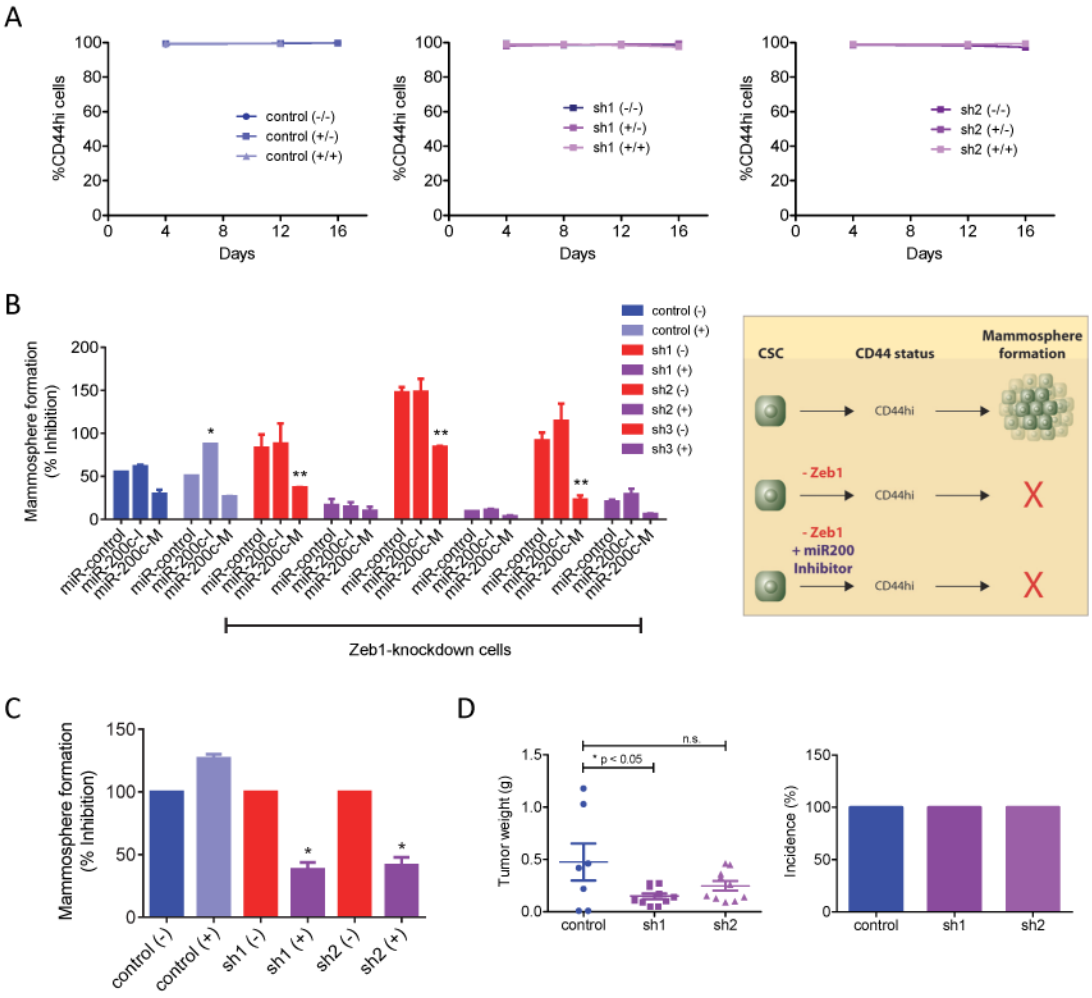


Figure 4. ZEB1 is essential for the stem cell/CSC activity of CD44^{hi} cells

(A) FACS analysis for CD44 expression in purified HME-flopc-CD44^{hi} cells. Cells express a doxycycline (dox) inducible control shRNA (control) or shRNA targeting *ZEB1* (sh1 and sh2). -/- no dox, +/- dox on for 8 days then removed for the remaining 8 days, +/+ dox on for the duration of the experiment. The percentage of spontaneously arising CD44^{lo} cells was determined by FACS over a 16 day time period.

(B) Purified HME-flopc-CD44^{hi} cells expressing control or shRNA targeting *ZEB1* (sh1, sh2 and sh3) were assessed for mammosphere-forming ability with or without dox-induction. Cells were treated with a *MIR200C* inhibitor (I) or mimetic (M). $p<0.001$, two-way ANOVA followed by Tukey's multiple comparisons test, *- different to miR-control and miR-200c-M, **- different to miR-control and miR-200c-I

(C) Transformed HME-flopc-CD44^{hi} cells expressing control, sh1 or sh2 were assessed for mammosphere formation with or without dox-induction ($p<0.0001$, one-way ANOVA followed by Tukey's multiple comparisons test, *-different to sh(-)).

(D) Transformed HME-flopc-CD44^{hi} cells (control, sh1 and sh2) were purified by FACS and implanted into the fat pad of NOD/SCID mice (n = 8/group). Tumor weight and incidence are shown.

Data represented as mean ± SEM. See also Figure S4

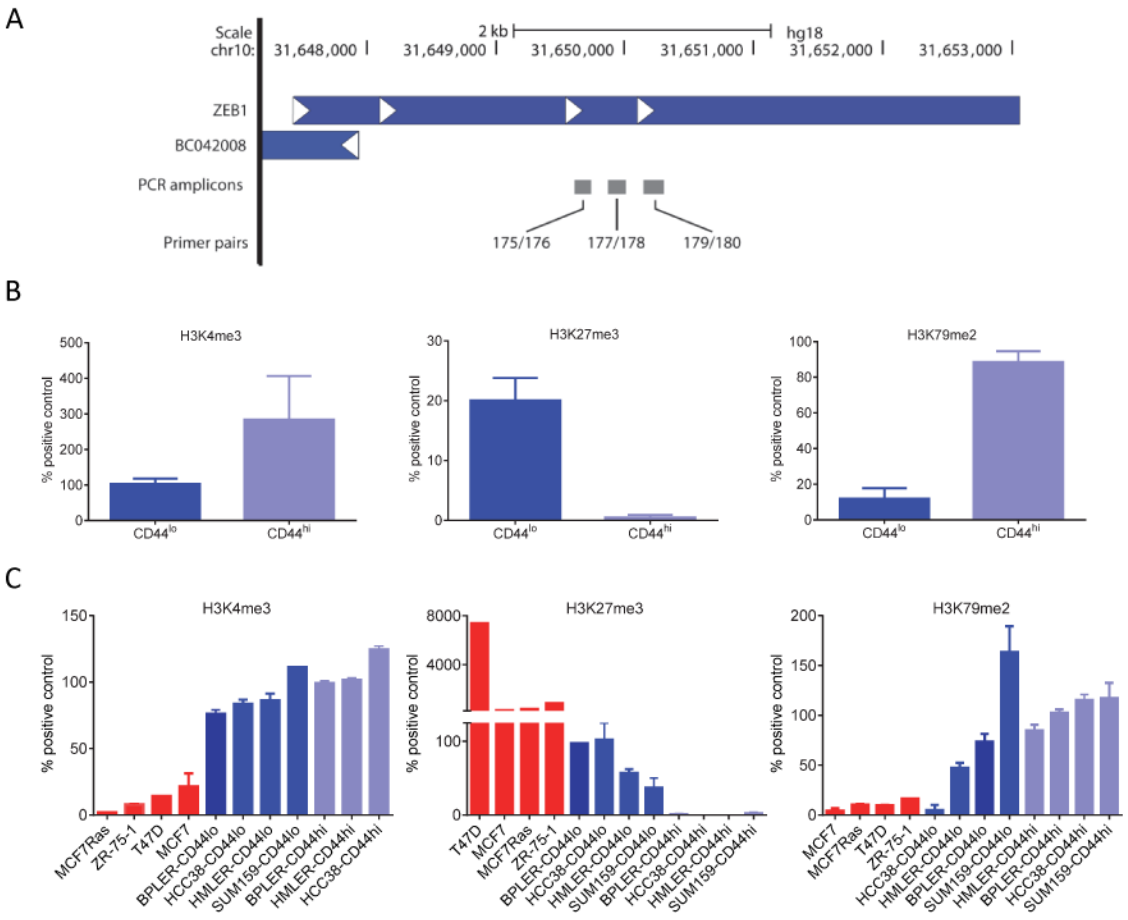


Figure 5. The ZEB1 promoter is maintained in a bivalent chromatin state in basal CD44^{lo} non-CSCs

(A) Schematic showing the location of primer sets used for ChIP-qPCR.

(B) and (C) ChIP-qPCR for the H3K4me3, H3K27me3 and H3K79me2 histone modifications at the *ZEB1* promoter in (B) non-transformed CD44^{lo} or CD44^{hi} cells and (C) luminal CD44^{lo} cells and basal CD44^{lo} and CD44^{hi} sorted populations. Data are mean ± SEM of biological duplicates performed as technical replicates.

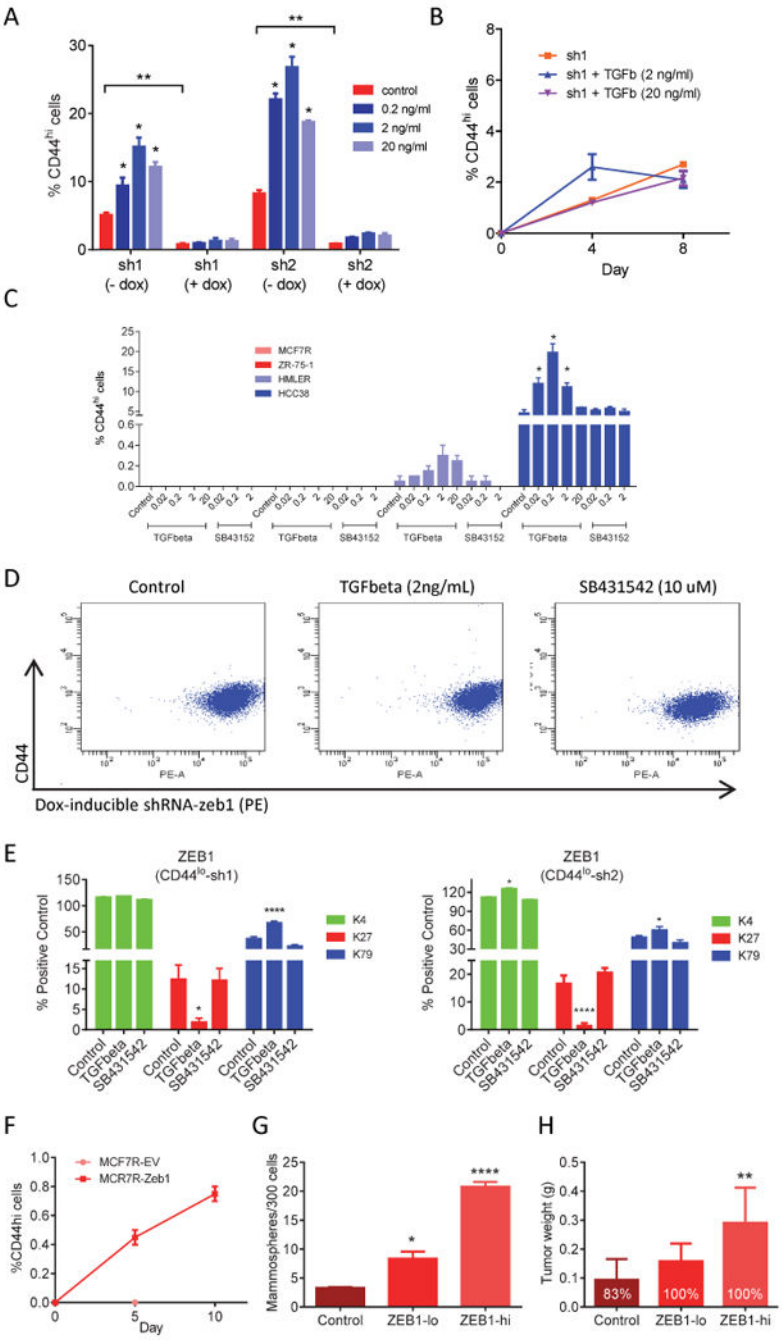


Figure 6. TGFbeta can induce CD44^{lo}-to-CD44^{hi} switching and modulates the chromatin at the ZEB1 promoter

(A) Purified HME-flopc-CD44^{lo} cells expressing dox-inducible shRNA targeting *ZEB1* (sh1 and sh2) were monitored by FACS for their ability to switch to the CD44^{hi} state following TGFbeta treatment *in vitro*. *p<0.0001 - different to control, **p<0.001- different to control (-dox).

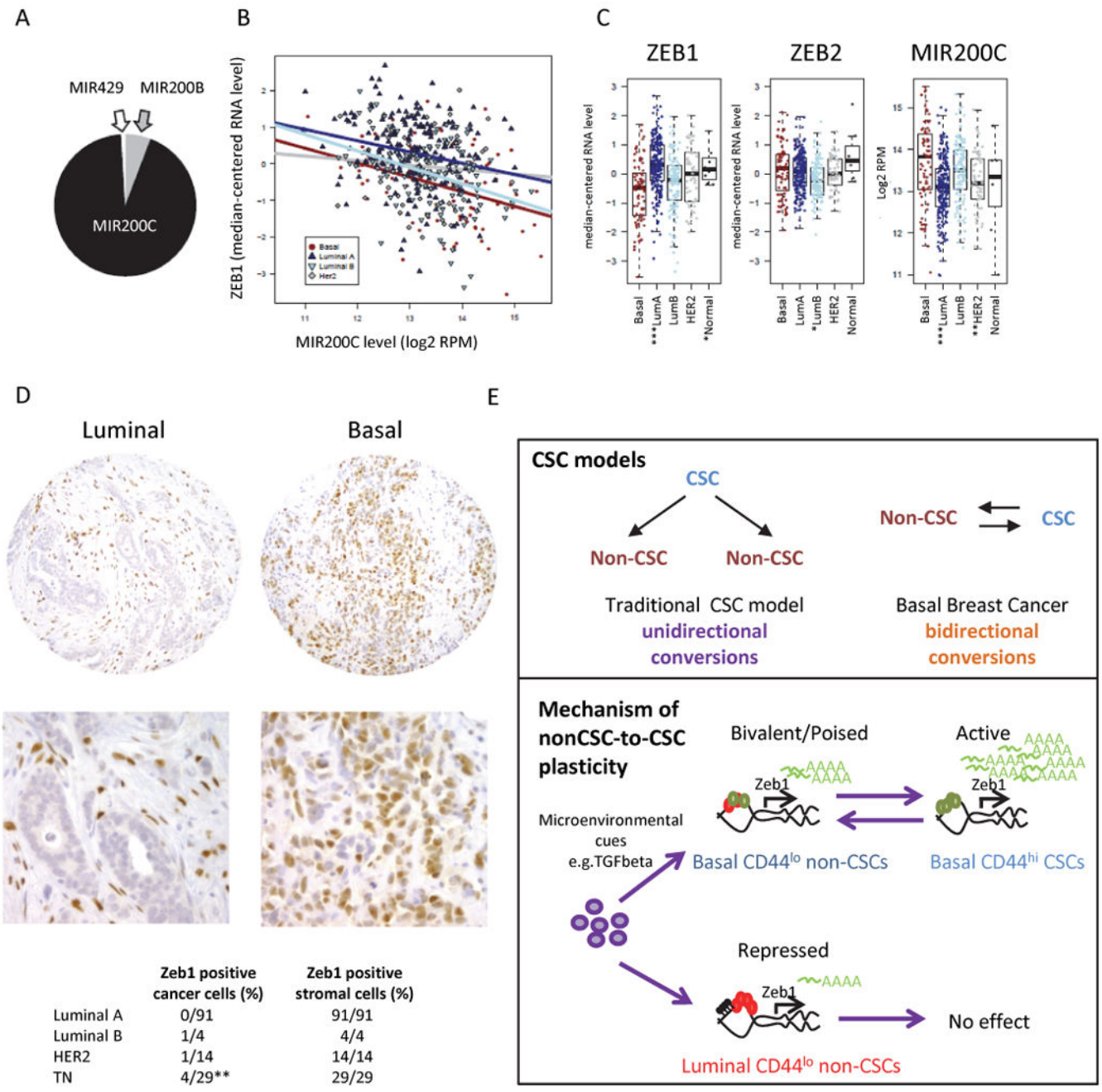
(B) Transformed HME-flopc-CD44^{lo} cells expressing sh1 targeting *ZEB1* were treated with TGFbeta and monitored by FACS for switching to the CD44^{hi} state.

(C) Purified $CD44^{lo}$ cells from luminal (MCF7R and ZR-75-1) and basal (HMLER and HCC38) BrCa cell lines monitored by FACS for switching to the $CD44^{hi}$ state following TGFbeta or SB431542 treatment *in vitro* *p<0.0001 – two-way ANOVA followed by Tukey's multiple comparisons test.

(D) Representative FACS plots of HME-flopc-$CD44^{lo}$ cells expressing sh1 or sh2 targeting *ZEB1* treated with control (PBS), TGFbeta (2ng/ml) or SB431542 (10μM).

(E) ChIP-qPCR for the H3K4me3, H3K27me3 and H3K79me2 histone modifications at the *ZEB1* promoter in cells from (D) (*p<0.0001, n = 4, two-way ANOVA followed by Tukey's multiple comparison test, different to control and SB431542).

(F-H) MCF7R cells expressing a dox-inducible empty vector (control) or *ZEB1* overexpression construct were treated with dox and monitored by FACS for their ability to switch to the $CD44^{hi}$ state (F), for the ability to form tumorspheres in vitro, *p<0.0001, one-way ANOVA and Tukey's multiple comparisons test, different to control (G) and for tumorigenicity in vivo (tumor initiating ability marked as percentages on each bar), *p = 0.03, one-way ANOVA and Tukey's multiple comparisons test, different to Control and ZEB1-lo (H). Data represented as mean ± SEM. See also Figure S6.



	Zeb1 positive cancer cells (%)	Zeb1 positive stromal cells (%)
Luminal A	0/91	91/91
Luminal B	1/4	4/4
HER2	1/14	14/14
TN	4/29**	29/29

Figure 7. ZEB1 in clinical cases of breast cancer

(A) *MIR200C* accounts for 93% (sd=5%) of mature miRNA in the *MIR200BC* family, which also includes *MIR200B* (6%) and *MIR429* (1%), so subsequent analysis uses *MIR200C* to represent the *MIR200BC* family.

(B) Levels of *ZEB1* and *MIR200C* are inversely correlated in basal (p=9.4e-4; r2=0.13), luminal A (p=2.8e-4; r2=0.07) and luminal B (p=6.8e-4; r2=0.12) subtypes (but not Her2). *ZEB1* is shown as median-centered values and *MIR200C* by log2-transformed reads per million mapped reads (RPM).

(C) mRNA abundance of *ZEB1*, *ZEB2*, and *MIR200C* by subtype. Asterisks indicate a difference compared to the basal subtype ($p<0.05$; ANOVA with Dunnett post-hoc): *=$p<0.05$; **=$p<0.01$; *=$p<0.05$.

(D) Human BrCa tissue array stained with an antibody targeting ZEB1 (100x and 400x images provided). ** $p=0.017$, TN compared to Luminal A, Fisher's Exact Test followed by Bonferroni correction for multiple-hypothesis testing. See also Figure S7.

(E) Schematic depicting: 1) an alternative CSC model for basal-type BrCa that includes bidirectional conversions between CSCs and non-CSCs, and 2) a model of non-CSC-to-CSC conversion that includes a microenvironmental stimulus acting on non-CSCs harboring bivalent chromatin marks at the *ZEB1* promoter enabling a rapid activation of *ZEB1* and switch to a CSC state.