

# Inducible formation of leading cells driven by CD44 switching gives rise to collective invasion

Cuixia Yang<sup>1,2#</sup>, Manlin Cao<sup>3#</sup>, Yiwen Liu<sup>1</sup>, Yiqing He<sup>1</sup>, Yan Du<sup>1</sup>, Guoliang Zhang<sup>1</sup>, Feng Gao<sup>1,2\*</sup>

## Abstract

Collective invasion into adjacent tissue is a hallmark of luminal breast cancer, with about 20% of cases that eventually undergo metastasis. It remained unclear how less aggressive luminal-like breast cancer transit to invasive cancer. Our study revealed that CD44<sup>hi</sup> cancer cells are the leading subpopulation in collective invading cancer cells, which could efficiently lead the collective invasion of CD44<sup>lo</sup>/follower cells. CD44<sup>hi</sup>/leading subpopulation showed specific gene signature of a cohort of hybrid epithelial/mesenchymal state genes and key functional co-regulators of collective invasion, which was distinct from CD44<sup>lo</sup>/follower cells. However, the CD44<sup>hi</sup>/leading cells, in partial-EMT state, were readily switching to CD44<sup>lo</sup> phenotype along with collective movements and vice versa, which is spontaneous and sensitive to tumor microenvironment. The CD44<sup>lo</sup>-to-CD44<sup>hi</sup> conversion is accompanied with a shift of CD44s-to-CD44v, but not corresponding to the conversion of non-CSC-to-CSC. Therefore, the CD44<sup>hi</sup> leader cells are not a stable subpopulation in breast tumors. This plasticity and ability to generate CD44<sup>hi</sup> carcinoma cells with enhanced invasion-initiating powers might be responsible for the transition from in situ to invasive behavior of luminal-type breast cancer.

**Key words:** leading cells/ luminal breast cancer/ collective invasion/ phenotype plasticity

**Running title:** CD44 switching in collective invasion

<sup>1</sup>(1)Department of Molecular Biology Laboratory, Shanghai Sixth People's Hospital, Shanghai Jiaotong University, Shanghai 200233, China;

(2) Department of Clinical Laboratory, Shanghai Sixth People's Hospital, Shanghai Jiaotong University, Shanghai 200233, China;

(3) Department of Rehabilitation Medicine, Shanghai Sixth People's Hospital, Shanghai Jiao Tong University, Shanghai 200233, China

#Both are co-first authors.

Corresponding author: Tel: +86-21-64369181-58702; Fax: +86-21-63701361; E-mail: gao3507@126.com

Disclosure of Potential Conflicts of Interest: No potential conflicts of interest were disclosed.

# Significance

Now, the mechanisms involved in local invasion and distant metastasis are still unclear. We identified a switch of CD44 that drives leader cell formation during collective invasion in luminal breast cancer. We provided evidence that interconversions between low and high CD44 states occur frequently during collective invasion. Furthermore, these findings demonstrated that the CD44<sup>hi</sup>/leader cells featuring partial EMT are inducible and attainable in response to tumor microenvironment. The CD44<sup>lo</sup> cancer cells are plastic that readily shift to CD44<sup>hi</sup> state, accompanied with shifts of CD44s-to-CD44v, thereby increasing tumorigenic and malignant potential. There are many “non-invasiveness” epithelial/follower cells with reversible invasive potential within an individual tumor, that casting some challenges on molecular targeting therapy.

## Introduction

The challenge to understand the nature of tumor cells that initiate metastases is of great clinical significance. Invasion into adjacent tissue often followed by metastasis is a hallmark of breast cancer (BrCa). Growing evidence showed that the metastatic dissemination of epithelial breast tumors can occur in a manner as collective cells (Friedl et al, 2012) without activation of the epithelial-mesenchymal transition (EMT) program (Fischer et al, 2015; Zheng et al, 2015), in which cancer cells locally invade the peritumoral stroma as cohesive clusters while maintaining cell-cell junctions, even traveling as clustering circulating-tumor cells (Aceto et al, 2014; Yu et al, 2013) and more efficiently seeding to form metastases than single circulating tumor cells (Cheung et al, 2016; Zajac et al, 2018). EMT models could not adequately describe the invasive behaviors of breast cancer cells in collective movement (Shamir et al, 2014). The nature of the cancer cells that give rise to cohesive invading is still poorly understood.

At the onset of collective cell invasion, a subset of cancer cells appear at the leading edge acquires a distinct phenotype with characteristic morphology and motility. It was revealed that basal genes, such as cytokeratin 14 (K14) or p63 (Cheung et al, 2013), and some signaling moleculars, such as Erk1/2 and RhoA located at the leading front cells (Chapnick & Liu, 2014; Reffay et al, 2014). Those leading cells can facilitate invasion by forming invasive protrusions, which provide traction and trigger modifications of the ECM in leader-follower cooperative manner (Nguyen-Ngoc et al, 2012). It is more of challenge for epithelial cancers to drive leader cell formation and initiate cell spread in a collective manner. The ability to acquire a distinct “leader” phenotype varies among different tumor cell populations originally derived from different patients with individual intracellular and extracellular microenvironment (Liu et al, 2013) (Dang et al, 2011; Nguyen-Ngoc et al, 2012). The mechanisms of how collectively cell invading into the surroundings was triggered are still not clearly understood.

Clinically, breast cancers are heterogeneous, often containing multiple subpopulations of cancer cells with distinct molecular states (Almendro et al, 2013; McGranahan & Swanton, 2015), which are likely to be plastic and adaptive to microenvironment stimuli. The heterogeneity and plasticity in mammary epithelial cell may cause poor response to endocrine therapy (Ito et al, 2015) and high risk of metastasis (Kennecke et al, 2010), although a more favorable prognosis could be

expected in most of luminal breast cancers. Understanding how these subpopulations in luminal epithelial cells, though displaying less invasive than basal like BrCa cells, acquire phenotypic changes and enable an enhanced invasive ability to metastasis is of great importance.

Here, we investigated the switch of phenotypic states in subpopulations in epithelial breast cancer induced by the shift of CD44 levels during collective cell invasion. The phenotypes of cancer cell were showed to be dynamic and plastic(Friedl & Alexander, 2011; Matak et al, 2017), and a dynamic conversion between CSCs and Non-CSCs was also revealed(Iliopoulos et al, 2011). CD44, as a mesenchymal marker, was used to identify cancer stem cells, and also contributed to regulate tumor stemness and metastasis(Ahmed et al, 2016). Interconversions between CSCs and Non-CSCs were usually happened that illustrating the dynamic and plastic of cancer cell phenotypes(Friedl & Alexander, 2011; Matak et al, 2017)<sup>23</sup>. Switch in expression pattern of CD44 standard (CD44s) and variant isoforms (CD44v) has been shown to influence EMT process, but functional significances of CD44 variants are not yet fully understood(Brown et al, 2011; Preca et al, 2015). Nevertheless, the involvement of CD44 in leader cell formation during epithelial collective invasion has not been studied. We therefore observed the inducible and exchangeable acquirement of CD44 along with other leader cell features, and monitored the formation of leading tips during collective migration. Our results revealed that the plasticity and ability of epithelial cancer cells, CD44<sup>lo</sup>-to-CD44<sup>hi</sup> conversion, generate invasive leading cells, which might be contribute to the transition from in situ to invasive behavior of luminal-type BrCas.

## Results

### 1. CD44 is a featured molecular of leading cells during collective invasion in breast cancers

To investigate intratumor heterogeneity related to CD44 expression during progression in breast carcinoma, we performed immunohistochemistry analysis of 247 human breast tumors of different molecular subtypes, 30 of which containing the tumor-stromal interface of the invasive edge. We analyzed the tumors for the expression of CD44 in 41 luminal A, 92 luminal B, 44 HER2+, and 70 triple negative (TN), as defined by immunohistochemical analysis. Interestingly, at the tumor-stromal

interface in all four molecular types, multicellular groups of cells were revealed invading into the adjacent stroma (Fig. 1A). Remarkably, the level of CD44 was significantly upregulated in the edge of leading cells in the invading strands, but not in other cells in the inner region of the collective clusters (Fig. 1A). The frequency of CD44<sup>hi</sup> cancer cells is correlated with the clinical molecular subtypes, displaying the highest in triple negative and the lowest in luminal-like cases examined (Fig. 1B and Supplemental Table S1), which is consistent with previous studies (Park et al, 2010). No significant differences were detected when the level of CD44<sup>hi</sup> cell was compared between luminal-A and luminal-B cases ( $P = 0.9301$ ) (Fig. 1B and Supplemental Table S1).

Similar to clinical tumor tissues, in orthotopic xenografts in mice, CD44 was also enriched at the invading boundary in luminal-like primary breast tumor and lung metastases (Fig. 1C). Also, immunostaining assay further indicated that cancer cells with CD44 high levels were located at migrating tips during collective movements in luminal-like breast cancer cells (Fig. 1D and Fig.S1). Follower cells at the inner region during collective migration, however, presented only low levels of CD44. Taken together, the distinctively high levels of CD44 in the invading front observed indicate that the CD44<sup>hi</sup> state is one of molecular features of leader cells, and breast tumors are composed of subpopulations of cancer cells with distinct CD44 expressions.

## **2. Oncogenic properties of CD44<sup>lo</sup> and CD44<sup>hi</sup> cancer cells generated during the collective invasion in vivo**

Clinically, 20% of luminal breast cancers finally metastasis to distant organs<sup>20</sup>. To test the role of CD44<sup>hi</sup> status on the leading edge cells in luminal breast cancers, though with general lower-level of CD44 than basal cancers (Fig. 1B and Table S1), we next sought to determine the function of CD44<sup>lo</sup> and CD44<sup>hi</sup> cells arising in vivo in collective invasion. Purified subpopulations of CD44<sup>hi</sup> and CD44<sup>lo</sup> cells from luminal-like tumors were isolated by fluorescence activated cell sorting (FACS). A subsequent cell proliferation assay displayed that CD44<sup>hi</sup> cells are significantly more proliferative than CD44<sup>lo</sup> subsets (Fig. 2A). Compatible with this result, an EdU incorporation assay also revealed that the proliferation of CD44<sup>hi</sup> cells was higher than that of CD44<sup>lo</sup> cells (Fig. 2B). Furthermore, we investigated the invasiveness of each of the subpopulations using the Matrigel-coated transwell filter assay, and the CD44<sup>hi</sup> populations showed by far more invasive (Fig. 2C). We also determined

stem/progenitor activity by mammosphere formation assay. Results showed that CD44<sup>hi</sup> subpopulation is more capable of forming mammospheres compared with CD44<sup>lo</sup> subsets (Fig. 3D). Interestingly, the CD44<sup>lo</sup> population is able to form mammospheres too, although with less efficient.

To further determine how CD44<sup>lo</sup> and CD44<sup>hi</sup> cells act in vivo, CD44<sup>hi</sup> and CD44<sup>lo</sup> cells sorted from MMTV tumors were immediately injected orthotopically into NOD/SCID mice at serial dilutions from 200 cells to 5×10<sup>5</sup> cell. In most cases, CD44<sup>hi</sup> cells generated 3- to 23-fold larger tumors compared to their CD44<sup>lo</sup> counterparts and presented greater tumor-initiating ability (Fig. 2E). However, we observed that the CD44<sup>lo</sup> cells formed tumors even in relatively low numbers of cells, although with a smaller size and longer duration at causing tumors (Fig. 2E). These results illustrated that CD44<sup>hi</sup> cells arising in vivo behave much like the CSCs in that they exhibit higher tumor-initiating and tumor growth potential than their CD44<sup>lo</sup> counterparts. As CD44 is widely accepted as a CSC marker, the results raised the question of how CD44<sup>lo</sup> populations that apparently lacked of CSC features were able to generate tumors when transplanted into mice.

### **3. CD44<sup>hi</sup> cells led collective invasion of luminal breast cancer cells**

To observe the migration of CD44<sup>hi</sup> cells and CD44<sup>lo</sup> cells, we cultured two confluent epithelial monolayers within two confined wells separately and then remove the confinement, triggering cluster migration in the forward facing direction(Fig.3A) as modified from previous report(Das et al, 2015). The CD44<sup>hi</sup>/DsRed<sup>+</sup> and CD44<sup>lo</sup>/DsRed<sup>-</sup> cells were isolated from MCF7/naïve cells and MCF7/DsRed cells by FACS. The migration of CD44<sup>hi</sup> cells and CD44<sup>lo</sup> cells was monitored during collective migration using time-lapse microscopy. Result showed that the migration rate of CD44<sup>hi</sup> cells is significantly higher than that of CD44<sup>lo</sup> cells( Fig.3B; Supplemental Movie 1-2). Knockdown and overexpression experiments showed that CD44 promoted collective migration, with CD44 depletion expectedly inhibited the migration of CD44<sup>hi</sup> subpopulation and CD44 overexpression encouraged the motility of CD44<sup>lo</sup> subpopulation (Fig.S2).

Moreover, to investigate whether CD44<sup>hi</sup> subset leads the migration of CD44<sup>lo</sup> subset and thereby promotes collective invasion, confocal time-lapse imaging of the cocultures of CD44<sup>hi</sup>/DsRed<sup>+</sup> and CD44<sup>lo</sup>/DsRed<sup>-</sup> subsets was performed. Imaging result showed that the leading cell is always a CD44<sup>hi</sup> subset and that the CD44<sup>lo</sup> cells move behind (Fig.3, C and D; Supplemental Movie 3-5).

#### **4. The molecular phenotype was distinct between the CD44<sup>hi</sup> and CD44<sup>lo</sup> cancer cells generated during the collective invasion**

In order to gain further insights into the molecular mechanisms enabling the observed differences between the two cell subsets in breast cancer, gene-expression signatures of FACS-sorted epithelial cells were analyzed. CD44<sup>hi</sup> cells that had been grown as leader cells or CD44<sup>lo</sup> cell as follower cells from the primary breast tumors were collected. Then, RNA sequencing was conducted for detecting the acquired gene expression at the CD44<sup>hi</sup> leader cells and compared with other CD44<sup>lo</sup> follower cells. As a result, 3,826 genes were differentially expressed between the two subsets (R2-fold;  $p < 0.05$ ) (Fig.S3A, B). Unsupervised hierarchical clustering of differentially expressed genes was involved in three major pathways: focal adhesion, EMT and RAS signaling pathway.

To address the factors underlying the different phenotypes, we analyzed the most significantly changed genes bioinformatically. Gene Ontology (GO) and KEGG functional enrichment analysis revealed 10 distinct terms enriched in the CD44<sup>hi</sup> cells (Fig. S3C). According to the high invasiveness in CD44<sup>hi</sup> cells, several genes related to cell motility and modifiers of the extracellular microenvironment were among the most differentially expressed genes, with lower levels in the CD44<sup>lo</sup> cells, including MMP2, MMP9, Plau, CLDN-1, -3, -11, and -7. Moreover, the CD44<sup>hi</sup> cells expressed a number of mesenchymal markers not present in CD44<sup>lo</sup> cells (Fig. 3E), including reduction of epithelial markers including CDH1 (E-cadherin), JUP ( $\gamma$ -catenin), krt8, and krt18 and increase of mesenchymal markers including FN1 (fibronectin), Vim (vimentin), COX-2(Ptgs2), and ZEB2. This was confirmed by quantitative real-time PCR (qRT-PCR) (Fig. 3F). However, how the CD44-associated invasive gene set revealed in CD44<sup>hi</sup> subset coordinates with each other and contributes to collective invasion are unknown.

Next, we investigated the localization of some additional mesenchymal markers that might explain the sustained invasion of carcinoma cells (Fig.3G). Surprisingly, besides the difference of E-cadherin RNA expression between CD44<sup>hi</sup> and CD44<sup>lo</sup> cells, epithelial markers E-cadherin and ZO-1 are missing at the lamellipodia in leading edge cells but remaining\_membrane-localized in the rear end of leading cells at the contacting area with the follower cells, with a similar distribution of  $\beta$ -catenin and a reorganization of F-actin(Fig.3G, Fig. S4). The migrating leader cells displayed branched F-actin filaments at the leading edge cells with a concomitant



forward-moving lamellipodia. Together, our results suggest that a complete EMT program is not achieved in the invading leader cells, despite vigorous invasiveness of cells into the ECM.

Although tumor organoids culture-based studies have described a critical role for K14<sup>+</sup> and p63<sup>+</sup> cells in leading tumor cells during collective invasion (Cheung et al, 2013), we did not observe any differences in the expression of K14 or p63 gene between CD44<sup>hi</sup> MMTV-PyMT and CD44<sup>lo</sup> MMTV-PyMT primary breast tumor cells. Besides tumor cells, tumor organoids in these studies contain tumor cells and stromal cells too, which were located at the boundary of tumor-stromal interface (Corsa et al, 2016) and influenced cell invasiveness. What's different here is that we obtained CD44<sup>hi</sup> and CD44<sup>lo</sup> cell by FACS depending on EpCAM<sup>+</sup>/CD45<sup>-</sup>, in which stromal cells were excluded. Moreover, observations of luminal cell lines also revealed that CD44 was located at the migration edge and tumor-stromal boundary of xenografts (Fig.1C).

### **5. The CD44<sup>lo</sup> and CD44<sup>hi</sup> states can spontaneously switch to each other**

To understand why the CD44<sup>lo</sup> subsets at lower cell numbers including 2000 cells/site were able to generate tumors when xenografted into mice, we investigated the dynamic changes of CD44 in the sorted CD44-different subpopulations. In the purified CD44<sup>lo</sup> group, the proportion of CD44<sup>hi</sup> cells was found to increase with cell expansion (Fig.4A). Alternatively, in the purified CD44<sup>hi</sup> group, the proportion of CD44<sup>lo</sup> cells was also expectedly found to increase with collective cell growth. Significantly, over a period of time, the level of CD44 in purified CD44<sup>hi</sup> and CD44<sup>lo</sup> group was stabilized to the level typical of the unsorted parent cells, indicating that the two cell subsets can shift from CD44<sup>hi</sup> phenotype to the CD44<sup>lo</sup> state in a dynamic equilibrium that maintains the proportion of CD44<sup>hi</sup> cells located at the migrating tips (Fig.4A). The failure of the sorted CD44-different subsets to be stably propagated indicated that mesenchymal-like CD44<sup>hi</sup> leading cells do not represent a stable state within an individual tumor.

Alternatively, to observe the CD44-dependent switching directly, we generated MCF7 cells carrying a CD44 promoter biosensor in which YFP reports on CD44 gene expression. In these cancer cells, CD44<sup>lo/YFP<sup>+</sup></sup> and CD44<sup>hi/YFP<sup>-</sup></sup> subpopulations were sorted and cultured in vitro, separately. We observed the transition from CD44<sup>lo</sup> to CD44<sup>hi</sup> states of individual cells at the leading edge and the switch to CD44<sup>lo/YFP<sup>+</sup></sup> states in the CD44<sup>hi/YFP<sup>-</sup></sup> cells derived follower cells (Fig.4B, Supplemental Movie



6-7). Following phenotypic conversion, the newly generated CD44<sup>hi</sup>/YFP<sup>-</sup> cells dynamically extended and retracted membrane protrusions and led migration of trailing CD44<sup>lo</sup>/YFP<sup>+</sup> cells (Fig.4B, Supplemental Movie 6-7). Taken together, these data informed that CD44<sup>lo</sup> cells were nonetheless able to generate new CD44<sup>hi</sup> cells, albeit at a low frequency. The invasive CD44<sup>hi</sup> leader phenotype may convert directly from CD44<sup>hi</sup> to CD44<sup>lo</sup> states along with the collective cell migrating, not a stable lineage.

## **6. Microenvironmental stimuli enhanced the rate of interconversions between CD44<sup>lo</sup> and CD44<sup>hi</sup> states**

Extensive evidence indicated that the stromal microenvironmental signals could influence cancer cell plasticity and tumorigenicity (Friedl & Alexander, 2011). Among these signals, hypoxia and tumor associated microphage (TAM) stimulation have been shown to potently activate EMT program and cancer stemness (Chaffer et al, 2013; Gregory et al, 2011). Accordingly, we examined whether hypoxia or TAM stimulation could affect the spontaneous transitions between CD44<sup>lo</sup> and CD44<sup>hi</sup> states of luminal BrCa cells. Purified CD44<sup>lo</sup> and CD44<sup>hi</sup> epithelial cells were cultured under hypoxic condition or cocultured with TAMs, respectively. TAMs cells were obtained after THP-1 polarized to M2-phenotype macrophages, according to our previous report (Zhang et al, 2016). Results showed that the treatment of hypoxia to CD44<sup>lo</sup> cells results in a rapid generation of a CD44<sup>hi</sup> subpopulation, and similar results occur when isolated CD44<sup>lo</sup> cells are cocultured with TAMs (Fig. 4C). Conversely, CD44 expression was reduced in CD44<sup>hi</sup> cells treated with hypoxia (Fig. 4D). Additionally, coculture with TAMs results in the rapid increase of CD44 levels in both of isolated CD44<sup>lo</sup> cells and CD44<sup>hi</sup> cells (Fig. 4C). The results suggested that tumor microenvironment contributes to the formation of CD44<sup>hi</sup>-leader cells in the leading front. Moreover, breast cancer microenvironment has an enhanced inflammatory network, in which IL-6 is reported to be an important regulator (Yadav et al, 2011). However, when rhIL6 is added to the medium, the conversion between CD44<sup>lo</sup> and CD44<sup>hi</sup> was not influenced, suggesting that secreted IL6 is not indispensable for generating CD44<sup>hi</sup> leader cells from epithelial tumor (Fig. S5). Together, these data indicated that microenvironmental signals in individual tumor enhance the rate of interconversions between CD44<sup>lo</sup> and CD44<sup>hi</sup> subsets in luminal breast cancer and contribute to the generation of CD44<sup>hi</sup>-leader cells.

## 7. The expression patterns of CD44 in the invasive CD44<sup>hi</sup> leading-cells and the CD44<sup>lo</sup> following-cells

To further identify the molecular mechanisms responsible for the switch of CD44, we investigated the expression pattern of CD44 variants. This pattern was detected in a reverse transcription polymerase chain reaction (RT-PCR) assay by a primer pair spanning the entire variant region (Banky et al, 2012; Yang et al, 2015). Our results showed that there is no difference of CD44s expression between CD44<sup>hi</sup> and CD44<sup>lo</sup> cells, whereas, the mRNA levels of CD44 variants were higher in the CD44<sup>hi</sup> cells compared with those CD44<sup>lo</sup> cells (Fig.5, A and C). When co-cultured with TAMs, expression of the CD44v was strongly up-regulated in both of the CD44<sup>hi</sup> and CD44<sup>lo</sup> subsets (Fig.5, B and D). Our results revealed that CD44 alternative splicing is differentially regulated during collective invasion, resulting in a switch in expression from the predominant isoform, CD44s, to the enriched variable exon-containing CD44v isoforms.

## 8. Changes of CSC population during the CD44 switching

It is widely accepted that CSCs are contained exclusively in the CD44<sup>hi</sup> cancer cells (Al-Hajj et al, 2003; Visvader & Lindeman, 2012). Therefore, we observed the localization of CSC cells in migrating cohesive clusters of luminal cancer cells using the expression of CD44 and CD24. Results showed that only a small CD44<sup>+</sup>/CD24<sup>-</sup> subclone was detected in migrating front cells, but not all of leading cells were in CD44<sup>+</sup>/CD24<sup>-</sup> state (Fig.5E). Whereas, a subpopulation of cancer cells with CD44<sup>+</sup>/CD24<sup>+</sup> positive accounted for the dominant population of the leading cells. The result illustrated that the nature of CD44<sup>hi</sup>/leading cells, residing in the migrating tips that produce higher tumor-initiating and tumor motility potential than their CD44<sup>lo</sup> follower counterparts, is not depending on the CSC-like characteristics.

Furthermore, to test whether the changes of CD44 levels on its own would allow enrichment of CSCs from BrCa cells, we analyzed the changes of CSCs proportion during CD44 switching in collective expansion of luminal BrCa using the Hoechst 33342 dye exclusion method (side population: SP). Accordingly, the proportion of CSC in luminal CD44<sup>hi</sup> or CD44<sup>lo</sup> cancer cells was measured by flow cytometry (FCM). We found that luminal CD44<sup>hi</sup> subpopulations contained CSC cells with a proportion of CSC cells (less than 5.5%). Meanwhile, luminal CD44<sup>lo</sup> subsets also comprised a small proportion of CSC cells (less than 2.2%) (Fig.5F). This result

implied that the switching of CD44 might not be corresponding to conversions of the stemness properties of non-CSCs-to-CSCs.

## **9. Dynamic colocalization of merlin/ezrin and CD44 enabled collective migration**

Of known intracellular CD44 binding partners, ezrin and merlin were reported important in contractile transduction and assembly of cortical actin during cell migration(Fehon et al, 2010). The distribution change of ezrin or merlin in response to various signals could control the lateral cell mobility(Chiasson-MacKenzie et al, 2015). We then investigate how the binding partners are involved in enabling the observed collective cell movements. A confluent epithelial BrCa monolayer was cultured to generate sheet migration by scratch-wounding. In the early time points of the sheet migration, the expression of ezrin was very low at the wound edge, which was coinciding with that of CD44 (Fig.6 A, left, Fig.S5). In contrast, after 72h of migration, ezrin was enriched in the leading edge cells, displaying co-localization with increased CD44 and suggesting complex formation between ezrin and CD44(Fig.6A, right, Fig.S6). Therefore, the increased CD44-Ezrin complexes at the leader cells presumably stabilize directional cell motility. Simultaneously, within the first few minutes of wound, merlin mostly presented at cohesive cell-cell boundaries (Fig.6B, left, Fig.S6). Interestingly, after collectively migrating for 72h, in the leading cells, merlin molecules transferred to the cytoplasm, whereas remains stable in stationary follower cells (Fig. 6B, right, Fig.S6). The spatial relocation of merlin in the cell-sheets suggested that the migration-induced cellular pulling-force triggered by leader cells resulted in a release of cortical merlin. Merlin is known to control lamellipodium formation(Das et al, 2015). The spatial location of merlin and formation of CD44-Ezrin complexes, accompanied with CD44 switch in the moving cell-sheets, presumably stabilizes the formation of invasive lamellipodium. Merlin/ezrin and CD44 might form a molecular switch that supports collective cell migration.

## **10. The shift of CD44<sup>hi</sup> and CD44<sup>lo</sup> states within tumors in mouse xenografts**

To further address the reciprocal conversion between sorted CD44<sup>hi</sup> and CD44<sup>lo</sup> subsets, we performed a mixed xenograft experiment (Fig.7A) in SCID mice involving co-injection of purified CD44<sup>hi</sup> and CD44<sup>lo</sup> cells in a mixed population from MMTV-PyMT mice. Initial cells from purified CD44<sup>hi</sup> cancer cells, CD44<sup>lo</sup> cancer cells, or the mixture group were xenografted orthotopically, separately. Accordingly, tumor size was measured with calipers. The tumors that arose from

different subpopulations as described above were digested and analyzed for CD44 expression by flow cytometry (FCM). Results showed that CD44<sup>lo</sup> BrCa cells displayed relative weaker tumor-initiating ability, with the smallest tumors among the three groups (Fig.7B). However, tumors derived from each subpopulation comprised almost the same proportion of EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> (p=0.4068) and EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> cells (p=0.4891) (Fig.7, C and D). This suggested that luminal CD44<sup>lo</sup> BrCa cells generate new CD44<sup>hi</sup> cells in vivo, therefore initiating tumor formation, although at a low speed. In comparison with tumors generated by CD44<sup>lo</sup> subset alone, the tumors derived from coinjection group were obviously larger. Interestingly, although only half the amount of CD44<sup>hi</sup> cells is used in the mixed group (Fig.7E), in comparison with tumors generated by CD44<sup>hi</sup> subset alone, there is no obvious difference in tumor size between the CD44<sup>hi</sup>-derived tumors and the mixed group-derived tumors. These findings indicate that luminal CD44<sup>hi</sup> to CD44<sup>lo</sup> conversion happened in vivo so that they achieve final equilibrium state of CD44 levels.

## Discussion

Clinically, luminal breast cancers can collectively invade as cohesive cell clusters, with about 20% of cases that eventually undergo metastasis(Kennecke et al, 2010). The molecular basis of the transition from in situ to invasive state in breast cancer remained elusive till now. Recently, the plasticity and continuous evolving of cell populations within the same tumor bring more challenges for explanation of the mechanisms involved(Friedl & Alexander, 2011).

**CD44<sup>lo</sup>-to-CD44<sup>hi</sup> Conversion Generates Invasive Leader Cells.** The present work reveals that CD44<sup>hi</sup> cancer cells are the leading subpopulation in collective invading luminal cancer, which could efficiently lead the collective invasion of CD44<sup>lo</sup>/follower cells. Indeed, analysis of the gene expression signature of CD44<sup>hi</sup>/leading subpopulation identified a cohort of hybrid epithelial/mesenchymal genes and key functional co-regulators of collective invasion, which was distinct from CD44<sup>lo</sup>/following cells. Furthermore, luminal CD44<sup>hi</sup> subpopulation was readily switching to CD44<sup>lo</sup> state along with collective cell movements, which is spontaneous and also sensitive to tumor microenvironment. Meanwhile, converting to CD44<sup>hi</sup> state was presented in luminal CD44<sup>lo</sup> subsets. Thus, the CD44<sup>hi</sup> leader cells featuring an intermediate EMT are not a stable subpopulation in breast tumor (Fig.7F). The ability of epithelial cancer cells to acquire an invasive mesenchymal-like phenotype

associated with the CD44 switching from a CD44<sup>lo</sup> to a CD44<sup>hi</sup> state, accompanied with a shift of CD44s-to-CD44v, suggests that invasion, metastasis, and patient outcome may be influenced by the inducible interconversions between CD44<sup>lo</sup> and CD44<sup>hi</sup> state. Collectively, the transition from in situ to invasive behavior of luminal-type BrCas could be considered to be due to this plasticity and ability to generate CD44<sup>hi</sup> carcinoma cells with enhanced invasion-initiating powers.

**CD44 Is Spatially Distributed In Invasive Cell Clusters.** CD44 is revealed to contribute to promote tumor progress by cooperating with coreceptors(Godar et al, 2008; Ponta et al, 2003). In our study, we found that the differences of CD44 levels appear to be closely related to the differences in cell molecular subtypes, with generally lower level in luminal BrCa than in basal BrCa, corresponding to its less aggressive and better prognosis than basal BrCa in clinical cases(Roswall et al, 2018). Additionally, in luminal BrCas, CD44 was enriched at the leading front of the invading cell strands, whereas in the inner region of the collective clusters it was maintained at a lower level. The CD44<sup>hi</sup> cells residing have a strong impact on the overall invasive behavior of these cohesive tumor cells. Accordingly, as a tumor-promoting protein(Ahmed et al, 2016) and a useful CSC marker(Al-Hajj et al, 2003; Prince et al, 2007), the spatial distribution of CD44 in invasive cell clusters, appears to be responsible for the profound differences in cell invasive ability among heterogeneous subpopulations.

**CD44<sup>lo</sup>-to-CD44<sup>hi</sup> Conversion Is Accompanied by Switch of Molecular Phenotypes.** Our further study showed that the invasive phenotype of CD44<sup>hi</sup>/leading cells is obtainable and dependent upon the CD44<sup>lo</sup>-to-CD44<sup>hi</sup> conversion, which is in line with the notion that cancer cells may undergo reversible switching of phenotype under deliberate cell regulating program. As suggested before, switch of phenotypes between subpopulations modulates the behavior of cancer cells in solid tumors(Roswall et al, 2018). In particular, there have been multiple observations about phenotype switching between subpopulations of breast cancer cells, that caused chemotherapy failure and tumor metastatic dissemination(Hanahan & Coussens, 2012; Yu et al, 2013). Evidences revealed that CSC-like cells may arise from non-CSCs and vice versa, and the molecular subtype of breast cancer can convert from basal-like phenotype to luminal-like state, which may influence the sensitivity to chemotherapy (Iliopoulos et al, 2011; Yang et al, 2012). In this study, results indicate that the partial-mesenchymal CD44<sup>hi</sup> subpopulations contain a small subclone of CSC cells;

however, the following cells also have a small subset of CSC-like cells. Therefore, the CD44 switching is not equivalent to the shift of non-CSC-to-CSC state. Whether traits attributed to CSC cells in previous studies, such as resistance to chemotherapy, are conferred to CD44<sup>hi</sup>/leading cells or CD44<sup>lo</sup>/following cells remain to be determined. However, although we cannot rule out the possibility that CSCs may play a critical role in attaining invasive phenotype of leader cells, we demonstrated that CD44 switching plays a critical role in driving formation of leading cells.

Furthermore, our results showed that the isoform switch from CD44s to CD44v occurred during the collective invasion. Knowledge about the precise function of CD44s and CD44v is as yet incomplete, rather controversial. It is reported that CD44v is related to metastatic properties whereas CD44s confers nonmetastatic properties (Morath et al, 2018; Yae et al, 2012). CD44v isoforms are also associated with stem-like properties in many different cancers (Banky et al, 2012; Guo & Frenette, 2014; Ishimoto et al, 2011). However, cancer cells with CD44s were showed to be more invasive and stronger drug resistance in pancreatic cancers (Zhao et al, 2016). The isoform switching from CD44v to CD44s was reported to contribute to EMT and breast cancer progression in immortalized normal cells where CD44s, not CD44v, has been manifested to be crucial for successful complete EMT (Brown et al, 2011). Taken together, although being disputative, the isoform switching of CD44 mediated by alternative splicing has important impact on malignant behaviors of cancer cells. Our results thus revealed that CD44 alternative splicing happens in response to the stimuli of TAMs, resulting in a shift in expression from CD44s to CD44v isoforms. Through isoform switching, the obtained high levels of CD44 isoforms might be responsible for the distinct cellular functions of cell subpopulations, including invasion and tumor-initiating ability.

**Leading Cells Have CD44-Dependent Invasion Program.** We observed large, CD44-dependent differences between CD44<sup>hi</sup> and CD44<sup>lo</sup> subpopulations in the same luminal breast tumor. In accordance with the ability of CD44<sup>hi</sup> cancer cells in driving collective invasion, we detected a specific molecular program in CD44<sup>hi</sup> subpopulation. Suppression of epithelial markers (CDH1 and JUP) and upregulation of mesenchymal markers (Vim, FN1, Ptgs2, and ZEB2) occurred in the CD44<sup>hi</sup>/leading cells. These changes are reported to be sufficient to activate complete EMT (Drees et al, 2005). However, at the same time, upregulation of tight junction proteins claudin 1, 3, 4 and 7 was also observed in the leader cells, which is opposite



to the classic EMT program(Ribeiro & Paredes, 2014) . Moreover, krt8 and krt18, typically expressed in epithelial cells, were reduced in CD44<sup>hi</sup>/leader cells compared with that in CD44<sup>lo</sup>/follower cells. Thus, the CD44<sup>hi</sup>/leading cells display a hybrid phenotype with both epithelial and mesenchymal features, implying that it did not undergo a complete EMT; instead, undergo partial EMT, an intermediate state. In agreement with this result, loss of krt8/krt18 in epithelial cancer cells was shown to increase cancer invasiveness without inducing EMT(Fortier et al, 2013). The hybrid E/M state has been described as a metastable phenotype(Lee et al, 2006), generating more plastic and metastable cancer cells(Klymkowsky & Savagner, 2009). Therefore, all these alterations might improve the formation of invasive tips, leading to an increase in cell motility and invasion capacity, and orchestrate the collective movement. Although expression of CD44 is reported to be regulated by p53, p63(Godar et al, 2008), and ESPR1/2(Brown et al, 2011), we did not observe any difference of those genes expression during the switch of CD44 in luminal BrCa. The precise mechanisms related need further exploration.

In addition, MMP2, MMP9 and Plau expression were increased in CD44<sup>hi</sup>/leader cells, in comparison with CD44<sup>lo</sup>/follower cells, which were demonstrated to accelerate invasion through remodeling of the ECM. Furthermore, analysis on the distribution features showed that CD44<sup>hi</sup>/leading cells presenting specific characteristics with some epithelial markers (E-cadherin,  $\beta$ -catenin, and ZO-1) were missed at the leading edge of lamellipodia, but remained at the contacting area with the following cells. Taken together, our data suggest that leader cells could possess all the gene expression required for enhanced cell invasion capacity but remain cell-cell junction with the follower cells. All these hybrid traits exhibited a transient EMT state, with an increased plasticity and invasiveness, which might be an additional explanation why EMT is dispensable for metastasis(Zheng et al, 2015).

**Merlin/Ezrin and CD44 Coordinately Regulate Collective Invasion.** Observations have suggested that, through forming coreceptor complexes with various receptors and signaling kinases, CD44 drives numerous signaling pathways in tumor-promoting effects (Godar et al, 2008) (Ponta et al, 2003). Also, CD44 serves as a picket attaching the actin cytoskeleton to the plasma membrane, connecting cyto- and pericellular signaling to modulate receptors mobility(Freeman et al, 2018). In this study, confocal microscopy studies confirmed the co-localization of enriched CD44 and ezrin at the front ends of a migrating leader cell, while merlin relocating to the



cytoplasm. Meanwhile, in the follower cells, merlin mostly concentrated at cell-cell contact sites together with E-cadherin and  $\beta$ -catenin. Ezrin was revealed to interact with both of CD44 and F-actin, forming a bridge between CD44 and the actin cytoskeleton, mediating changes in cell morphology and migration (Donatello et al, 2012). Merlin was reported to act as a mechanochemical transducer, coordinating lamellipodium formation and collective movement of cell clusters during wound healing (Das et al, 2015). Therefore, the reciprocity of Ezrin-merlin-CD44 association in migrating cell clusters may help maintain collective cell movement during breast cancer cell invasion. However, further studies are needed to clarify the mechanism.

## **Conclusion**

In summary, the study indicated the potential values of CD44 switching and shifts of CD44s/v in explaining the mechanism of metastasis and predicting the progression from luminal breast cancers in situ to invasive carcinoma. This inducible acquirement of CD44 cast some challenges on the molecular targeting therapy, as has been demonstrated in non-CSC and CSC conversions. In addition, spatial localization of distinct phenotypic cells in solid tumors might lead to the generation of distinct niches around the invading cell clusters that favor the cohesive cell invasion as a whole. Therefore, in the future, a deeper understanding of the intratumor plasticity and their consequences might provide new insights for the mechanisms underlying collective invasion and metastasis.

## **Experimental Procedures**

### **1. Antibodies and Immunofluorescence**

For immunofluorescence, cells were fixed in 4% paraformaldehyde, permeabilized with Triton-X-100/PBS, and blocked in 1×PBS + 1% BSA. Primary antibodies were incubated overnight at 4°C, 1× PBS + 1% BSA. Secondary antibodies with Alexa Fluor 488 or 647 (Abcam) were incubated 1 h at room temperature. Images were analyzed under confocal microscopy (Nikon A1, Tokyo, Japan). Primary antibodies were CD44 (Abcam, ab119348), E-cadherin (Abcam, ab40772),  $\beta$ -catenin (Abcam, ab32572), Merlin (Bioworld Technology, BS3663), Ezrin (Cell Signal, 3145s), and phalloidin (Abcam, ab143533).

### **2. Cell culture and Mice**

Human breast cancer cell lines MCF-7, T47D, and ZR-75-30 were bought from American Type Culture Collection (ATCC) and cultured in MEM, RPMI-1640 medium, or Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10%

fetal calf serum, 100 U/ml penicillin, and 100 mg/ml streptomycin. All cells were grown to 85% confluency for experiments.

In this report, we explore the role of CD44 in tumor progression using the mouse mammary tumor virus LTR-driven polyoma middle T antigen (MMTV-PyMT) transgenic mouse, which has been proved to be a reliable model of human luminal breast cancer (Guy et al., 1992; Herschkowitz et al., 2007; Lin et al., 2003). MMTV-PyMT (FVB/n) mice were from Jackson Labs. All protocols involving mice were evaluated and approved by our Institutional Animal Care and Use Committee and performed under veterinary supervision.

### **3. Immunohistochemical staining of tumor samples**

Formalin-fixed paraffin-embedded human breast cancer tissues were obtained from US Biomax (BR1504, BR1505) and Shanghai Superchip Biotech (HBre-Duc052Bch, HBre-Duc068Bch-01, OD-CT-RpBre03-004). For xenografted tumors arising from MCF7 cells, tumors were extracted and formalin-fixed followed by paraffin-embedding. For immunostaining, paraffin sections were fixed in acetone, air dried, and then stored at -80°C. After abolishment of endogenous peroxidase activity (0.3% H<sub>2</sub>O<sub>2</sub>, 20 minutes), tissues were rehydrated with PBS and blocked with 3% BSA in PBS for 30 minutes. Then, sections were treated with primary antibodies specific for CD44 for 60 minutes at room temperature. Detection was achieved with compatible-conjugated secondary antibody (Invitrogen, CA, USA) and Horse radish-peroxidase-conjugated (HRP-conjugated) ABC amplification system.

**4. CCK-8 assay** The proliferation of CD44<sup>hi</sup> and CD44<sup>lo</sup> cells was determined by CCK-8 kit (Dojindo, Japan). Approximately purified  $3.5 \times 10^3$  cells in 100 ml were incubated in triplicate in 96-well plates. At 0, 24, 48, 72 and 96 h, the CCK-8 reagent (10 ml) was added to each well and incubated at 37°C for 2 h. The optical density at 450 nm was measured using an automatic microplate reader (Synergy4; BioTek, Winooski, VT, USA).

**5. EdU assay** The cell proliferation was analyzed by EdU (5-ethynyl-20-deoxyuridine) assay using Cell-Light EdU DNA Cell Proliferation Kit (RiboBio, Shanghai, China). CD44<sup>hi</sup> and CD44<sup>lo</sup> cells ( $1 \times 10^4$ ) were seeded in 96-well plates. After incubation for 48 h, 50 mM EdU was added and incubated at 37°C for another 2 h. Then, cells were fixed with 4% paraformaldehyde and stained with Apollo Dye Solution for proliferating cells. Finally, Hoechst 33342 was added to stain nucleic acids in all cells. The cell proliferation rate was calculated according to the manufacturer's instructions.

Images were taken using a fluorescence microscope (Nikon).

**6. Mammosphere formation assay** Sphere formation assays were performed as described previously (Phillips et al, 2006). CD44<sup>hi</sup> or CD44<sup>lo</sup> cells from MCF7 or MMTV PyMT tumors were purified by FACS. To measure the mammosphere formation potential of the sorted cells,  $1 \times 10^3$  cells were plated in ultra-low-attachment 24-well dishes containing mammosphere growth medium (MEGM with all required growth factors as described above for the growth of HMLER cells plus B-27 supplements, without bovine pituitary extract (BPE]) from Lonza. After 12 days of culture at 37°C, the resulting mammospheres were counted.

**7. In vivo tumor initiation assay** CD44<sup>hi</sup> or CD44<sup>lo</sup> cells from MMTV PyMT tumors were FACS sorted using Astrios System (Beckman). Purified cells were transplanted subcutaneously into NOD/SCID mice at concentrations of 100,000, 20,000, and 2000 cells per site to analysis of the tumor initiation potential. Cells were harvested in a suspension of 50% Matrigel (BD Biosciences, cat. no. 356231) in DME/F12 medium to a final volume of 50µl, then orthotopically injected into 4- to 6-week-old NOD/SCID mice. Tumor volume was measured at the time of sacrifice and calculated by the ellipsoid volume. 3-8 mice were used in each group. After 3-4 weeks the tumors were identified by palpation. Mice were killed after tumors reached a diameter of 1.5 cm, in accordance with institutional guidelines.

**8. Live imaging of cell migration assay** Live imaging of cells was conducted using a Nikon Living-Cell Observer system with a confocal microscope (Nikon A1, Tokyo, Japan). In general, images were collected at 30-min or 1h intervals with exposure times of ~250 ms. Some of the movies were collected. Temperature was held at 37 °C and CO<sub>2</sub> at 5%.

Migration experiments were performed according to the method described by Das et al.(Das et al, 2015). CD44<sup>hi</sup> and CD44<sup>lo</sup> cells from MCF7 cells were cultured on each dish separately with one silicone culture-insert (ibid) press-bonded to create confined wells (each well 3.25mm×7mm; growth area per well 22 mm<sup>2</sup>). When cells formed 100% confluent monolayers, the insert was removed to generate cluster migration in the forward facing direction with 0.5mm wound-gap.  $3 \times 10^4$  cells were seeded into each well with 80 µl cell culture medium and incubated at 37 °C in a 5% CO<sub>2</sub> humidified incubator. The migration area of each cell= Cell area after 24 or 72 hours-Cell area at T0.

**9. RNA Sequencing** Tumor tissue was enzymatically digested using mouse tumor

dissociation kit (Miltenyi, Cat.130-096-730) before FACS analysis. Cells sorted from MMTV tumors with CD44<sup>hi</sup>/EpCAM<sup>+</sup>/CD45<sup>-</sup> or CD44<sup>lo</sup>/EpCAM<sup>+</sup>/CD45<sup>-</sup> features were collected. Total RNA was isolated using the Phenol-chloroform extraction and used for Illumina RNA sequencing analysis (Illumina HiSeq X-Ten). Transcriptomic libraries were constructed using Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, United States). Reads with high-quality were aligned to the mouse reference mRNA (GRCh38) using Bowtie2. For gene ontology analysis, we used Gene Ontology (GO) and KEGG for gene functional and signaling pathway annotation. Hypergeometric test and the following FDR correction were used for enriched GO term or KEGG pathway analysis. After FDR corrected, significant enrichment was valued by the P-value cutoff of 0.05. The expression levels for each of the genes were normalized to fragments per kilobase of exon model per million mapped reads (FPKM) using RNA-seq by Expectation Maximization (RSEM). The hierarchical clustering was analyzed using Pearson correlation distance measure and average linkage method.

## 10. Quantitative real-time PCR

Real-time PCR was performed to validate gene expression in CD44<sup>hi</sup> and CD44<sup>lo</sup> cancer cells. Total RNA was extracted by RNeasy mini kit (Takara) according to the manufacturer's instructions. Purified RNA was treated with RNase-free DNaseI to ensure complete degradation of potential genomic DNA contamination, followed by the addition of 25mM EDTA and heat denaturation of the enzyme. RNA samples were reverse transcribed to cDNA with random hexamer primers using a SuperScript III First-Strand cDNA Synthesis Kit (Takara). Real-time PCR primers were designed with RealTime Design software (<http://www.bio-searchtech.com>) and were custom made by Invitrogen.

**10. Analysis of CD44 expression** Cells were harvested and washed with Hanks' balanced salt solution (HBSS) supplemented with 2% bovine serum albumin, (pH 7.4). A single cell suspension (10<sup>6</sup> cells/ml) was incubated with CD44-PE antibody or isotope control antibodies on ice for 1 h. Cells were washed three times and analyzed using a flow cytometer (Beckman-Coulter, Brea, USA), and at least 8,000 cells were analyzed per sample. All experiments were performed at three times.

**11. Analysis of CD44 expression pattern** Cells were harvested and total RNA was purified using TRIzol reagent. For reverse transcription, a random nonamer-oligo dT combination was used as the primer. To determine the CD44 expression pattern, we

use a human-specific primer pair spanning the entire variant region (forward primer: 5'-AGT CAC AGA CCT GCC CAA TGC CTT T, reverse primer: 5'-TTTGCT CCA CCT TCT TGA CTC CCA TG-3'(Banky et al, 2012; Yang et al, 2015). All CD44 variants are theoretically amplified. The PCR reaction mixture was obtained from Takara. The PCR cycling procedures were described as our previous report(Yang et al, 2015).

## 12. The shift of CD44 in vivo

To determine the generation of CD44<sup>hi</sup> cells from CD44<sup>lo</sup> cells and vice versa in vivo,  $1 \times 10^6$  CD44<sup>hi</sup>, CD44<sup>lo</sup>, or the mixer cells (CD44<sup>hi</sup>:CD44<sup>lo</sup>=1:1) obtained from MMTV PyMT tumors, were transplanted orthotopically into NOD/SCID mice. The tumors were removed 4 weeks after transplantation, and digested into single-cell suspensions using mouse tumor dissociation kit (Miltenyi, Cat.130-096-730). The cell suspension was analyzed for CD44 expression using an Astrios flow cytometer.

**Statistical analysis** Statistical analysis was performed using GraphPad InStat software (GraphPad Software, Inc.). Nonparametric Mann-Whitney tests were performed to measure the differences in CD44 expression, tumor formation, mammosphere-forming capacity, and invasiveness between CD44<sup>hi</sup> and CD44<sup>lo</sup> cells. The significance of differences among groups was determined by one-way ANOVA t test, or Fisher exact test accordingly. Statistically significance was considered as  $P < 0.05$ .

## Author Contributions

Conceptualization, CX.Y., ML. C., F.G.; Formal Analysis, T.M., R. J., J.V.; Investigation, YW.L., YQ.H., Y.D., GL. Zh.; Resources, CX.Y., YW.L., YQ.H., Y.D., GL. Zh., F.G; Funding Acquisition, CX.Y., F.G.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (81672843, 81572821, 81502490, 81502491, 81402419), Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (20171924), the Natural Science Foundation of Shanghai Municipality (14YF1412200), and the Shanghai Shen-Kang Hospital Development Center (SHDC22014004).

## References

Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, Yu M, Pely A, Engstrom A,

Zhu H, Brannigan BW, Kapur R, Stott SL, Shioda T, Ramaswamy S, Ting DT, Lin CP, Toner M, Haber DA, Maheswaran S (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**: 1110-1122

Ahmed M, Sottnik JL, Dancik GM, Sahu D, Hansel DE, Theodorescu D, Schwartz MA (2016) An Osteopontin/CD44 Axis in RhoGDI2-Mediated Metastasis Suppression. *Cancer cell* **30**: 432-443

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 3983-3988

Almendo V, Marusyk A, Polyak K (2013) Cellular heterogeneity and molecular evolution in cancer. *Annual review of pathology* **8**: 277-302

Banky B, Raso-Barnett L, Barbai T, Timar J, Becsagh P, Raso E (2012) Characteristics of CD44 alternative splice pattern in the course of human colorectal adenocarcinoma progression. *Molecular cancer* **11**: 83

Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, Cheng C (2011) CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *The Journal of clinical investigation* **121**: 1064-1074

Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, D'Alessio AC, Young RA, Weinberg RA (2013) Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* **154**: 61-74

Chapnick DA, Liu X (2014) Leader cell positioning drives wound-directed collective migration in TGFbeta-stimulated epithelial sheets. *Molecular biology of the cell* **25**: 1586-1593

Cheung KJ, Gabrielson E, Werb Z, Ewald AJ (2013) Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155**: 1639-1651

Cheung KJ, Padmanaban V, Silvestri V, Schipper K, Cohen JD, Fairchild AN, Gorin MA, Verdone JE, Pienta KJ, Bader JS, Ewald AJ (2016) Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proceedings of the National Academy of Sciences of the United States of America* **113**: E854-863

Chiasson-MacKenzie C, Morris ZS, Baca Q, Morris B, Coker JK, Mirchev R, Jensen AE, Carey T, Stott SL, Golan DE, McClatchey AI (2015) NF2/Merlin mediates contact-dependent inhibition of EGFR mobility and internalization via cortical actomyosin. *The Journal of cell biology* **211**: 391-405

Corsa CA, Brenot A, Grither WR, Van Hove S, Loza AJ, Zhang K, Ponik SM, Liu Y, DeNardo DG, Eliceiri KW, Keely PJ, Longmore GD (2016) The Action of Discoidin Domain Receptor 2 in Basal Tumor Cells and Stromal Cancer-Associated Fibroblasts Is Critical for Breast Cancer Metastasis. *Cell reports* **15**:

2510-2523

Dang TT, Precht AM, Pearson GW (2011) Breast cancer subtype-specific interactions with the microenvironment dictate mechanisms of invasion. *Cancer research* **71**: 6857-6866

Das T, Safferling K, Rausch S, Grabe N, Boehm H, Spatz JP (2015) A molecular mechanotransduction pathway regulates collective migration of epithelial cells. *Nature cell biology* **17**: 276-287

Donatello S, Babina IS, Hazelwood LD, Hill AD, Nabi IR, Hopkins AM (2012) Lipid raft association restricts CD44-ezrin interaction and promotion of breast cancer cell migration. *The American journal of pathology* **181**: 2172-2187

Drees F, Pokutta S, Yamada S, Nelson WJ, Weis WI (2005) Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly. *Cell* **123**: 903-915

Fehon RG, McClatchey AI, Bretscher A (2010) Organizing the cell cortex: the role of ERM proteins. *Nature reviews Molecular cell biology* **11**: 276-287

Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, Choi H, El Rayes T, Ryu S, Troeger J, Schwabe RF, Vahdat LT, Altorki NK, Mittal V, Gao D (2015) Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**: 472-476

Fortier AM, Asselin E, Cadrin M (2013) Keratin 8 and 18 loss in epithelial cancer cells increases collective cell migration and cisplatin sensitivity through claudin1 up-regulation. *The Journal of biological chemistry* **288**: 11555-11571

Freeman SA, Vega A, Riedl M, Collins RF, Ostrowski PP, Woods EC, Bertozzi CR, Tammi MI, Lidke DS, Johnson P, Mayor S, Jaqaman K, Grinstein S (2018) Transmembrane Pickets Connect Cyto- and Pericellular Skeletons Forming Barriers to Receptor Engagement. *Cell* **172**: 305-317.e310

Friedl P, Alexander S (2011) Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147**: 992-1009

Friedl P, Locker J, Sahai E, Segall JE (2012) Classifying collective cancer cell invasion. *Nature cell biology* **14**: 777-783

Godar S, Ince TA, Bell GW, Feldser D, Donaher JL, Bergh J, Liu A, Miu K, Watnick RS, Reinhardt F, McAllister SS, Jacks T, Weinberg RA (2008) Growth-inhibitory and tumor-suppressive functions of p53 depend on its repression of CD44 expression. *Cell* **134**: 62-73

Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S, Morris M, Wyatt L, Farshid G, Lim YY, Lindeman GJ, Shannon MF, Drew PA, Khew-Goodall Y, Goodall GJ (2011) An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Molecular biology of the cell* **22**: 1686-1698



Guo W, Frenette PS (2014) Alternative CD44 splicing in intestinal stem cells and tumorigenesis. *Oncogene* **33**: 537-538

Hanahan D, Coussens LM (2012) Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer cell* **21**: 309-322

Iliopoulos D, Hirsch HA, Wang G, Struhl K (2011) Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 1397-1402

Ishimoto T, Nagano O, Yae T, Tamada M, Motohara T, Oshima H, Oshima M, Ikeda T, Asaba R, Yagi H, Masuko T, Shimizu T, Ishikawa T, Kai K, Takahashi E, Imamura Y, Baba Y, Ohmura M, Suematsu M, Baba H, Saya H (2011) CD44 variant regulates redox status in cancer cells by stabilizing the xCT subunit of system xc(-) and thereby promotes tumor growth. *Cancer cell* **19**: 387-400

Ito T, Sato N, Yamaguchi Y, Tazawa C, Moriya T, Hirakawa H, Hayashi S (2015) Differences in stemness properties associated with the heterogeneity of luminal-type breast cancer. *Clinical breast cancer* **15**: e93-103

Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K (2010) Metastatic behavior of breast cancer subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**: 3271-3277

Klymkowsky MW, Savagner P (2009) Epithelial-mesenchymal transition: a cancer researcher's conceptual friend and foe. *The American journal of pathology* **174**: 1588-1593

Lee JM, Dedhar S, Kalluri R, Thompson EW (2006) The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *The Journal of cell biology* **172**: 973-981

Liu L, Duclos G, Sun B, Lee J, Wu A, Kam Y, Sontag ED, Stone HA, Sturm JC, Gatenby RA, Austin RH (2013) Minimization of thermodynamic costs in cancer cell invasion. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 1686-1691

Matak A, Lahiri P, Ford E, Pabst D, Kashofer K, Stellas D, Thanos D, Zatloukal K (2017) Stochastic phenotype switching leads to intratumor heterogeneity in human liver cancer. *Hepatology*

McGranahan N, Swanton C (2015) Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer cell* **27**: 15-26

Morath I, Jung C, Leveque R, Linfeng C, Toillon RA, Warth A, Orian-Rousseau V (2018) Differential recruitment of CD44 isoforms by ErbB ligands reveals an involvement of CD44 in breast cancer. *Oncogene*

Nguyen-Ngoc KV, Cheung KJ, Brenot A, Shamir ER, Gray RS, Hines WC, Yaswen P, Werb Z, Ewald AJ (2012) ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E2595-2604

Park SY, Gonen M, Kim HJ, Michor F, Polyak K (2010) Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *The Journal of clinical investigation* **120**: 636-644

Ponta H, Sherman L, Herrlich PA (2003) CD44: from adhesion molecules to signalling regulators. *Nature reviews Molecular cell biology* **4**: 33-45

Preca BT, Bajdak K, Mock K, Sundararajan V, Pfannstiel J, Maurer J, Wellner U, Hopt UT, Brummer T, Brabletz S, Brabletz T, Stemmler MP (2015) A self-enforcing CD44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. *International journal of cancer* **137**: 2566-2577

Prince ME, Sivanandan R, Kaczorowski A, Wolf GT, Kaplan MJ, Dalerba P, Weissman IL, Clarke MF, Ailles LE (2007) Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 973-978

Reffay M, Parrini MC, Cochet-Escartin O, Ladoux B, Buguin A, Coscoy S, Amblard F, Camonis J, Silberzan P (2014) Interplay of RhoA and mechanical forces in collective cell migration driven by leader cells. *Nature cell biology* **16**: 217-223

Ribeiro AS, Paredes J (2014) P-Cadherin Linking Breast Cancer Stem Cells and Invasion: A Promising Marker to Identify an "Intermediate/Metastable" EMT State. *Frontiers in oncology* **4**: 371

Roswall P, Bocci M, Bartoschek M, Li H, Kristiansen G, Jansson S, Lehn S, Sjolund J, Reid S, Larsson C, Eriksson P, Anderberg C, Cortez E, Saal LH, Orsmark-Pietras C, Cordero E, Haller BK, Hakkinen J, Burvenich IJG, Lim E, Orimo A, Hoglund M, Ryden L, Moch H, Scott AM, Eriksson U, Pietras K (2018) Microenvironmental control of breast cancer subtype elicited through paracrine platelet-derived growth factor-CC signaling. *Nature medicine* **24**: 463-473

Shamir ER, Pappalardo E, Jorgens DM, Coutinho K, Tsai WT, Aziz K, Auer M, Tran PT, Bader JS, Ewald AJ (2014) Twist1-induced dissemination preserves epithelial identity and requires E-cadherin. *The Journal of cell biology* **204**: 839-856

Visvader JE, Lindeman GJ (2012) Cancer stem cells: current status and evolving complexities. *Cell stem cell* **10**: 717-728

Yadav A, Kumar B, Datta J, Teknos TN, Kumar P (2011) IL-6 promotes head and neck tumor metastasis by inducing epithelial-mesenchymal transition via the JAK-STAT3-SNAIL signaling pathway. *Molecular cancer research : MCR* **9**: 1658-1667

Yae T, Tsuchihashi K, Ishimoto T, Motohara T, Yoshikawa M, Yoshida GJ, Wada T, Masuko T, Mogushi K, Tanaka H, Osawa T, Kanki Y, Minami T, Aburatani H, Ohmura M, Kubo A, Suematsu M, Takahashi K, Saya H, Nagano O (2012) Alternative splicing of CD44 mRNA by ESRP1 enhances lung colonization of metastatic cancer cell. *Nature communications* **3**: 883

Yang C, He Y, Zhang H, Liu Y, Wang W, Du Y, Gao F (2015) Selective killing of breast cancer cells expressing activated CD44 using CD44 ligand-coated nanoparticles in vitro and in vivo. *Oncotarget* **6**: 15283-15296

Yang G, Quan Y, Wang W, Fu Q, Wu J, Mei T, Li J, Tang Y, Luo C, Ouyang Q, Chen S, Wu L, Hei TK, Wang Y (2012) Dynamic equilibrium between cancer stem cells and non-stem cancer cells in human SW620 and MCF-7 cancer cell populations. *British journal of cancer* **106**: 1512-1519

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, Concanon KF, Donaldson MC, Sequist LV, Brachtel E, Sgroi D, Baselga J, Ramaswamy S, Toner M, Haber DA, Maheswaran S (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**: 580-584

Zajac O, Raingeaud J, Libanje F, Lefebvre C, Sabino D, Martins I, Roy P, Benatar C, Canet-Jourdan C, Azorin P, Polrot M, Gonin P (2018) Tumour spheres with inverted polarity drive the formation of peritoneal metastases in patients with hypermethylated colorectal carcinomas. *Nature cell biology* **20**: 296-306

Zhang G, Guo L, Yang C, Liu Y, He Y, Du Y, Wang W, Gao F (2016) A novel role of breast cancer-derived hyaluronan on inducement of M2-like tumor-associated macrophages formation. *Oncoimmunology* **5**: e1172154

Zhao S, Chen C, Chang K, Karnad A, Jagirdar J, Kumar AP, Freeman JW (2016) CD44 Expression Level and Isoform Contributes to Pancreatic Cancer Cell Plasticity, Invasiveness, and Response to Therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research*

Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, Wu CC, LeBleu VS, Kalluri R (2015) Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**: 525-530

## Figures and legends:

**Fig.1 CD44<sup>hi</sup> cancer cells are enriched at the leading edge in collective invasive tumor cells and in lung metastases.** (A) Immunohistochemical expression of CD44 in mammary tumor sections based on molecular subtypes. CD44 with homogenous membranous staining is enriched in the leading breast carcinoma cells. Arrow, leading CD44<sup>hi</sup> cancer cell with forked protrusions. (B) The expression of CD44 in different molecular subtypes of breast cancers. (C) CD44<sup>hi</sup> cancer cells are enriched at the invasive border in primary tumor and in lung metastases from MMTV-PyMT mice. (D) Expression of CD44 in luminal breast cancer cells during collective migration. Scale bars, 25  $\mu$ m.

**Fig.2 The CD44<sup>hi</sup> and CD44<sup>lo</sup> cells generated during the collective invasion have distinct functions.** (A) Proliferation of FACS-sorted EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> or EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> cancer cells from MMTV/PyMT mice assessed using a CCK-8 kit at the indicated days. Data are the means  $\pm$  s.e.m. of three experiments. (B) DNA synthesis assessed using an EdU (5-ethynyl-2'-deoxyuridine) assay in sorted EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> and EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> cancer cells. Cells were fluorescently stained with EdU (red). Nuclei were stained with DAPI (blue). Micrographs represent at least three experiments. Scale bar, 200  $\mu$ m. Quantitative EdU assay data are the means  $\pm$  s.e.m. of three experiments. \*p < 0.05, \*\*p < 0.01 (Student's t-test). (C) Invasion assay were performed using sorted EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> or EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> cells and allowed to migrate for 24 h. Representative images are shown, scale bar=200  $\mu$ m (n=2 independent cell cultures and experiments, 3 technical replicates of each). Throughout figure, all graphs are shown as mean  $\pm$  s.e.m. \*p < 0.05; \*\*p < 0.01 by two-tailed Student's t-test. (D) Mammosphere-forming capacity of CD44<sup>hi</sup> and CD44<sup>lo</sup> populations. CD44<sup>hi</sup> or CD44<sup>lo</sup> cells ( $1 \times 10^3$ ) were sorted into each well of 24-well ultra-low attachment dishes containing mammosphere growth medium. Cells were cultured for 12 days, and the photos were taken using a light microscope. Scale bars: 100  $\mu$ m. Numbers of mammospheres per well were quantified. \*p < 0.01 compared with CD44<sup>lo</sup>. The experiment was performed in triplicate. (E) Assessment of the tumorigenicity of the EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> or EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> cells (sorted from MMTV/PyMT mice) orthotopically injected into NOD-SCID mice at several dilutions with  $200-5 \times 10^5$  cell, as determined by tumor volume, is shown. Numbers indicate the frequency of tumor formation. Dot plots indicate individual tumor volumes (Mean indicated by solid line). \*p < 0.05 compared with CD44<sup>lo</sup> cells-derived tumors.

**Fig.3 CD44<sup>hi</sup> cancer cells with the invasive phenotype lead the luminal CD44<sup>lo</sup> cancer cells during the collective invasion.**

(A) Schematic representation of the in vitro collective migrating model in an epithelial monolayer. The FACS-sorted CD44<sup>hi</sup> cells and CD44<sup>lo</sup> cancer cells were cultured within two confined wells separately. After confluency, the confinement was removed to trigger epithelial cluster migration. The space between two cell monolayers was 500µm wide. (B) Phase contrast microscopy images of the forward facing migration of CD44<sup>hi</sup> cells and CD44<sup>lo</sup> epithelial monolayers. Scale bars, 50µm. (C) Schematic representation of the collective migration with the mixed two types of cells. (D) CD44<sup>hi</sup>/Dsred<sup>+</sup> cells migrate collectively and lead trailing CD44<sup>lo</sup>/Dsred<sup>-</sup> cells. Luminal CD44<sup>lo</sup> or CD44<sup>hi</sup> cancer cells obtained from MCF7 cells were co-cultured in vitro. The collective expansion was observed by time-lapse microscopy. Scale bars, 50µm. (E) Heat maps of unsupervised hierarchical clustering data are shown depicting differentially expressed genes in EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> and EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> subsets (blue and yellow indicate higher and lower expression, respectively) with fold changes  $\geq 2$  and FDR  $\leq 0.05$ . (F) Differential gene expression verified by quantitative RT-PCR in FACS-purified CD44<sup>hi</sup> and CD44<sup>lo</sup> subpopulations from primary breast cancer cells. GAPDH was used as a loading control. (G) Confocal images of localization of E-cadherin,  $\beta$ -catenin, ZO-1, and F-actin at the leading edge and centre of the migrating clusters. Data are representative of at least three experiments. Scale bar, 25 µm.

**Fig.4 Inducible shift of CD44 expression occurs during collective invasion.**

(A) Flow cytometric analysis of the changes of CD44 expression in CD44<sup>hi</sup> and CD44<sup>lo</sup> subpopulations during cell expansion in vitro. Purified luminal CD44<sup>lo</sup> or CD44<sup>hi</sup> MCF7 cells were cultured in vitro and analyzed by FACS for proportion of CD44<sup>lo</sup> and CD44<sup>hi</sup> components. Luminal BrCa cells may convert directly from CD44<sup>hi</sup> to CD44<sup>lo</sup> states and vice versa. (B) Time-lapse microscopy of CD44 switching in CD44<sup>hi</sup> and CD44<sup>lo</sup> fractions sorted from MCF7 cells, expressing YFP under the control of the CD44 promoter (CD44 promoter-YFP). (C) Flow cytometric analysis of changes of CD44 levels after stimulated with TAM. TAM cells were obtained after THP-1 cells were stimulated to have M2-phenotypes. Purified CD44<sup>hi</sup> or CD44<sup>lo</sup> cells obtained from MCF7 cells were incubated in the absence or presence of TAM cells for 0-7days using a Transwell coculture system, then examined for CD44 levels by FACS. (D) Flow cytometric analysis of changes of CD44 levels after treatment with hypoxia. Purified CD44<sup>hi</sup> or CD44<sup>lo</sup> MCF7 cells were incubated in hypoxia and subjected to flow cytometric analysis. Scale bars, 50µm.

**Fig.5 CD44 expression pattern with CD44-dependent phenotype and characterization of CSC-like cells during collective migration.** (A) Reverse transcription PCR (RT-PCR) analysis of CD44 expression pattern in the CD44<sup>lo</sup> or CD44<sup>hi</sup> subpopulation generated from collective migration. Using a primer pair spanning the entire variant region, the CD44 variants expression was detected. (B) CD44v expression was boosted by TAM stimulation. Co-culture system of TAM cells and FACS-purified CD44<sup>lo</sup> or CD44<sup>hi</sup> MCF7 cells were used as described before. After co-incubation for 3 days, cells were harvested and analyzed by RT-PCR for the expression pattern of CD44. (C) Quantitative analysis of CD44s in the CD44<sup>lo</sup> or CD44<sup>hi</sup> subpopulation determined in (A) and its changes induced by TAM treatment in (B). (D) Quantitative analysis of CD44v in cell subpopulations before and after TAM stimulation in (B). Relative expression levels of CD44s or CD44v were normalized to GAPDH. Error bars indicate SD; n = 4. (E) Immunocytochemistry analysis of the expression of cell surface CD44 and CD24 in collective migrating cell clusters. MCF7 cells were cultured in the wound chamber (ibid). After a confluent monolayer was formed, the culture-insert was removed to trigger a sheet migration of MCF7 cells. Then, cells were fixed for immunofluorescence assay at 72 h, after the wound was made. Representative images were taken under a confocal microscope. Scale bars, 100  $\mu$ m. (F) Hoechst 33342 efflux assay of SP and non-SP cells in CD44<sup>hi</sup> and CD44<sup>lo</sup> subpopulations. CD44<sup>hi</sup> or CD44<sup>lo</sup> cancer cells (upper) were separately incubated with Hoechst 33342 dye and then analyzed by flow cytometry.

**Fig.6 Co-localization of CD44 and Merlin or Ezrin in stationary and migrating cells during collective migration.** (A) Co-localization of CD44 and Ezrin in stationary (left) and moving (right) monolayer. EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>hi</sup> or EpCAM<sup>+</sup>/CD45<sup>-</sup>/CD44<sup>lo</sup> cancer cells were purified from MMTV/PyMT mice, and then cultured in the wound chamber (ibid). A sheet migration of epithelial cells was generated as described before. Then, cells were fixed for Ezrin and CD44 expression at 0h and 24h post wounding by immunofluorescence assay. Representative images were presented. Scale bars, 50  $\mu$ m. (B) Localization of CD44 and merlin in stationary (left) and migrating (right) monolayer. A sheet migration of epithelial cells was made as previously described. Merlin and CD44 expression at 0h and 24h post wounding were determined by immunofluorescence assay. Scale bars, 50 $\mu$ m.

**Fig.7 Conversion of CD44<sup>hi</sup> and CD44<sup>lo</sup> cells within tumors in mouse xenografts.** (A) Scheme of mixed xenograft experiment. The CD44<sup>hi</sup> cells were combined (or not) with CD44<sup>lo</sup> cells from MCF7, injected into NSG-SCID mice. (B) Tumorigenicity of FACS-sorted luminal-like CD44<sup>lo</sup> or CD44<sup>hi</sup> BrCa cells following orthotopically injected into SCID mice (n $\geq$ 6/group). (C, D) Quantification of CD44<sup>hi</sup> or CD44<sup>lo</sup> cells generated from purified-CD44<sup>lo</sup>-, CD44<sup>hi</sup>-, or mixed group-derived tumors in (A). Cells were obtained from excised tumors 21 d after injection and analyzed by FCM. (E) Weights of tumors generated from luminal CD44<sup>lo</sup>, CD44<sup>hi</sup>, or mixed cell populations and representative tumor image. (F) Scheme of CD44 switching in leading collective cancer cell invasion. A collective invasion model shows phenotypic interconversion between two subpopulations, a highly invasive CD44<sup>hi</sup>/leading cell (red membrane) at the front of the invading cell clusters and CD44<sup>lo</sup>/following cells frequently occurred during collective invasion, which is spontaneous and sensitive to tumor microenvironment.

## Supplemental Figures Legends

**Fig. S1 Collective migration of breast cancer cells in 2-D monolayer culture.** Confocal XYZ vertical images showed that CD44 maintained in the leading edge of collective cell clusters.

**Fig. S2 The role of CD44 in mediating collective migration in monolayer sheets.** Purified CD44<sup>hi</sup> or CD44<sup>lo</sup> cells obtained from MCF7 cells were transfected with Sh-CD44 virus or CD44 virus, separately, then examined for the wound closure. Cells were cultured in 6-well plates until they formed 100% confluent monolayers. Then a 100 mm “wound” scratches were made using a sterile pipette tip. The cells was incubated with MEM containing 1% FBS for different periods. Then the pictures were captured.

**Fig. S3 Gene expression of purified CD44<sup>hi</sup>/leading and CD44<sup>lo</sup>/following cells.** (A) Scatter plot showing ,expression differences greater in follower cells(CD44<sup>lo</sup>) (y-axis) versus leader cells (CD44<sup>hi</sup>) (x-axis). (B) Unsupervised hierarchical clustering of 1422 significantly differentially expressed genes in leader (purple, N=2782) and follower (yellow,N=1039) cells. The hierarchical clustering was performed using Pearson correlation distance measure and average linkage method. The log<sub>2</sub> transformed gene levels are represented after quantile normalization. (C)Bioinformatic analysis of gene expression differences in CD44<sup>hi</sup>/leading and CD44<sup>lo</sup>/following cells.

**Fig.S4 Confocal images of co-localization of CD44 and E-cadherin, or  $\beta$ -catenin, during collective migration of MCF7 cells.**

Co-localization of CD44 and E-cadherin, or  $\beta$ -catenin, in migrating monolayer. MCF7 cells or EpCAM<sup>+</sup>/CD45<sup>-</sup> primary cancer cells purified from MMTV/PyMT mice, were cultured in the wound chamber (ibid). A sheet migration of epithelial cells was generated as described before. Then, cells were fixed for CD44 and E-cadherin, or  $\beta$ -catenin expression at 24h post wounding by immunofluorescence assay. Representative images were presented. Scale bar, 50  $\mu$ m.

**Fig. S5 Flow cytometric analysis of changes of CD44 levels after IL6 stimulation**

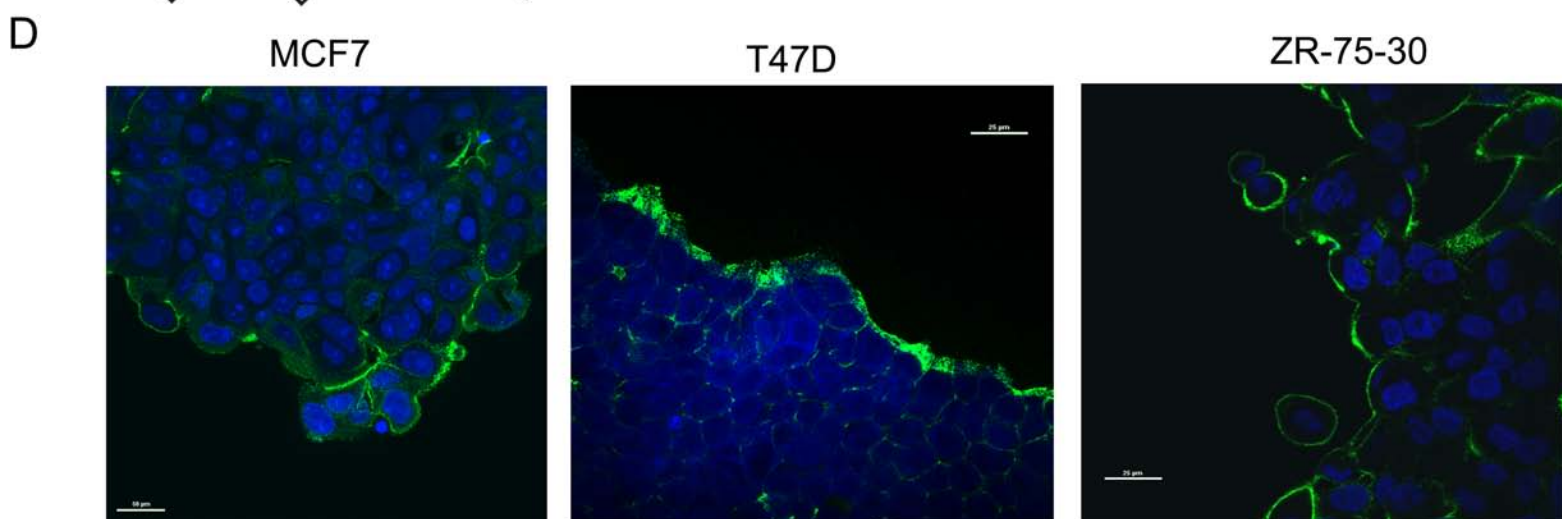
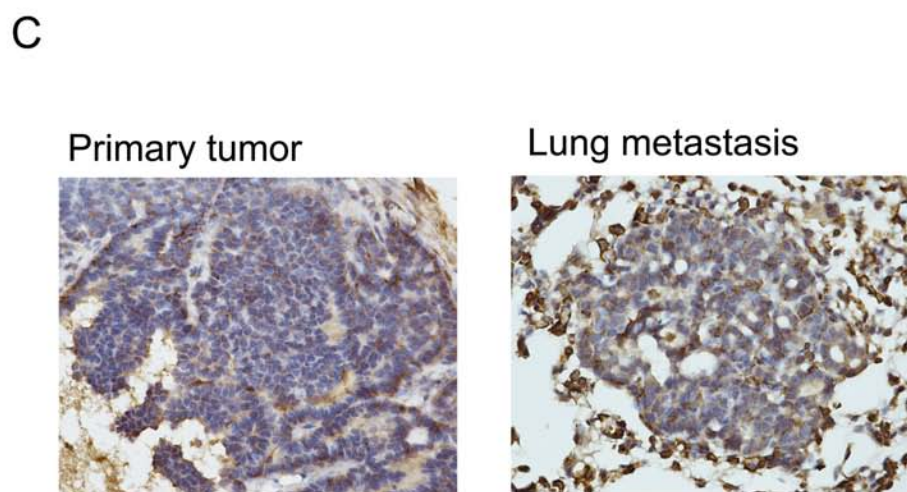
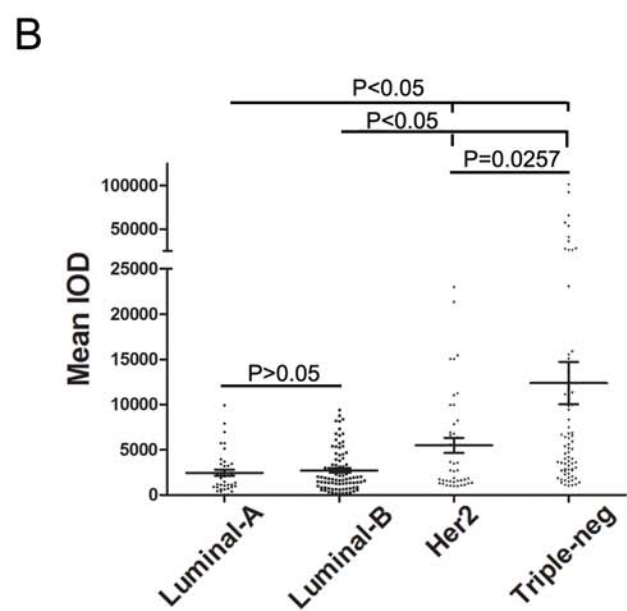
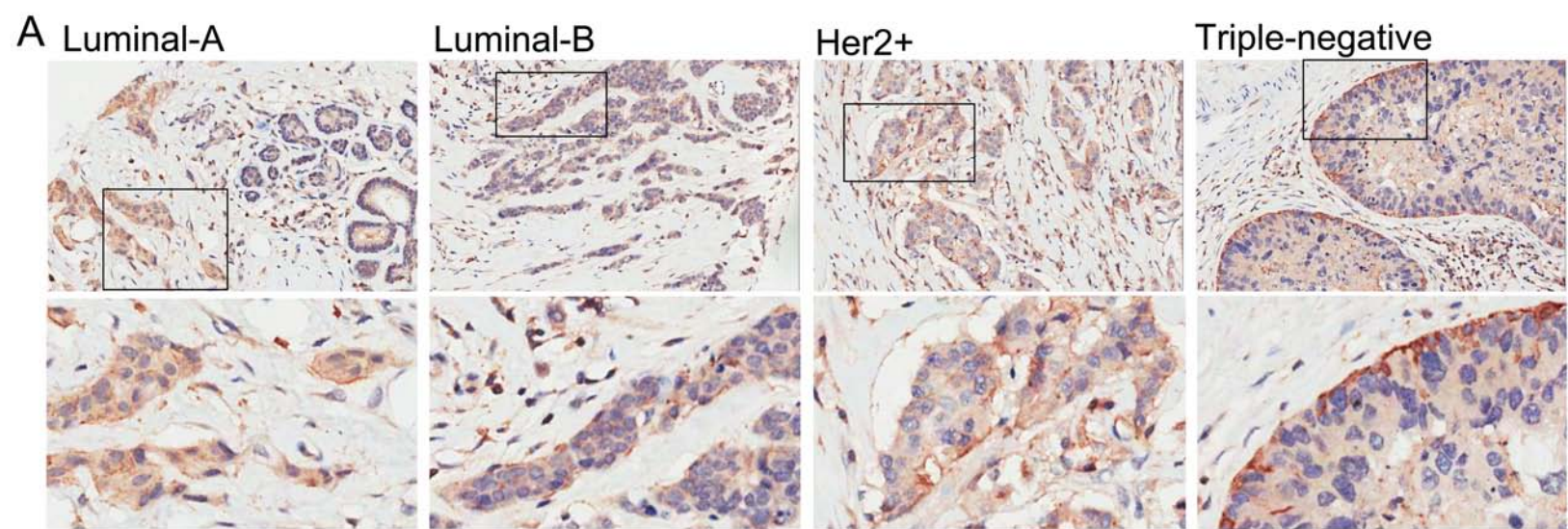
Flow cytometric analysis of changes of CD44 levels after stimulated with rhIL-6. Purified CD44<sup>hi</sup> or CD44<sup>lo</sup> cells obtained from MCF7 cells were incubated in the absence or presence of rhIL-6 (50ng/ml) for 0-7days, then examined for CD44 levels by FACS.

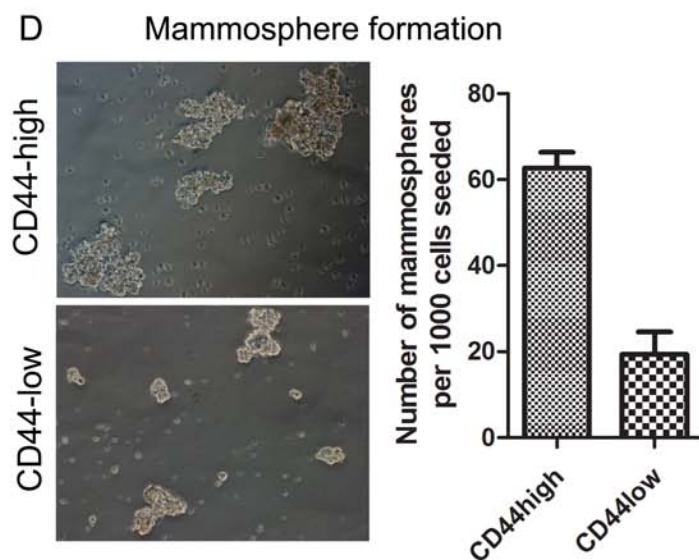
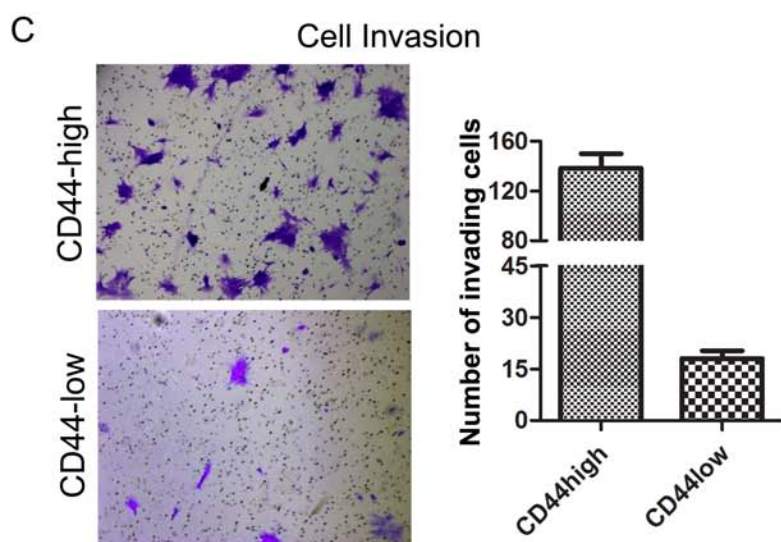
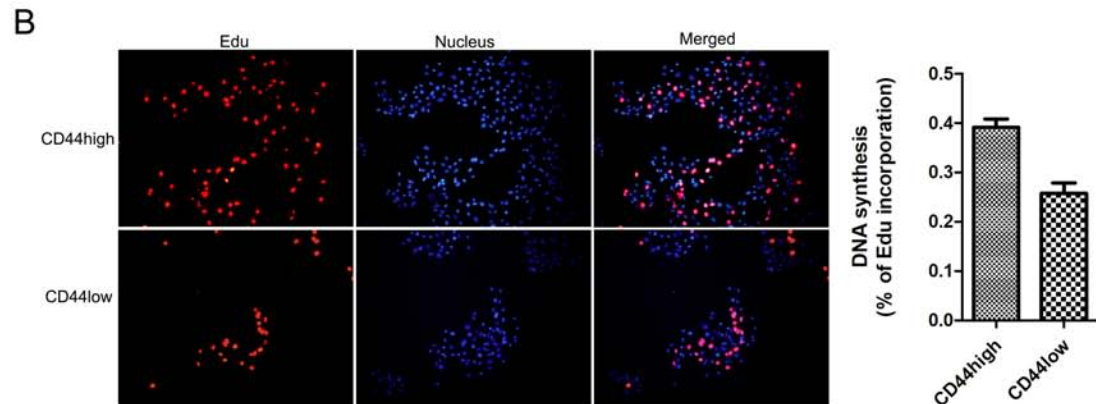
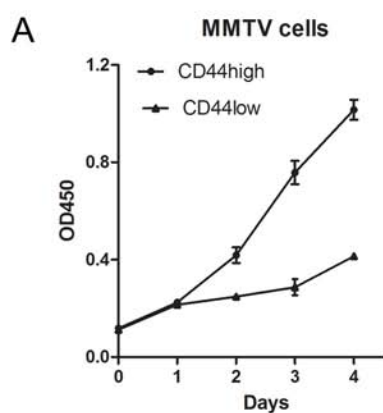
**Fig.S6 Co-localization of CD44 and Merlin or Ezrin in stationary and migrating cells during collective migration of MCF7 cells.** (A) Co-localization of CD44 and Merlin in stationary (upper) and moving (Lower) monolayer. CD44<sup>hi</sup> or CD44<sup>lo</sup> MCF7 cells were cultured in the wound chamber (ibid). A sheet migration of epithelial cells was generated and fixed for Merlin and CD44 expression at 0h and 72h post wounding by immunofluorescence assay. Representative images were presented. Scale bars, 50  $\mu$ m. (B) Localization of CD44 and ezrin in stationary (upper) and migrating ((Lower) monolayer. ezrin and CD44 expression at 0h and 72h post wounding were determined by immunofluorescence assay. Scale bars, 50 $\mu$ m.



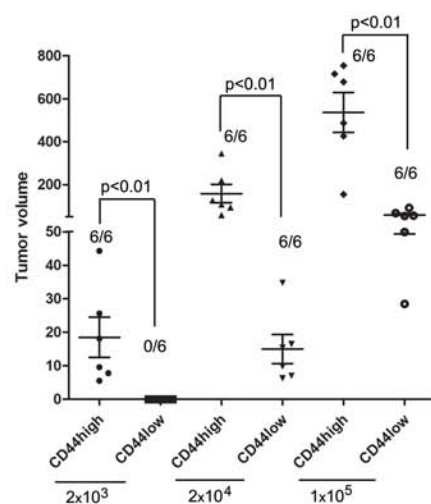
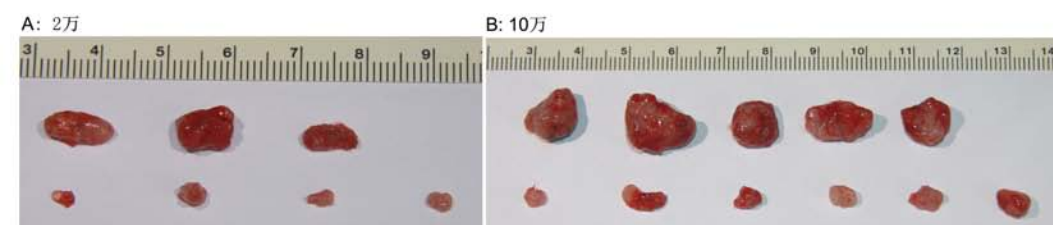
## Supplemental Videos Legends

- Movie 1: The migration rate of CD44<sup>hi</sup> cells and CD44<sup>lo</sup> cells from 0h to 17h
- Movie 2: The migration rate of CD44<sup>hi</sup> cells and CD44<sup>lo</sup> cells from 42h to 72h
- Movie 3: CD44<sup>hi/DsRed-</sup> subset leads the migration of CD44<sup>lo/DsRed+</sup> subset purified from MCF7 cells from 24h to 38h
- Movie 4: CD44<sup>hi/DsRed-</sup> subset leads the migration of CD44<sup>lo/DsRed+</sup> subset purified from MCF7 cells from 39h to 55h
- Movie 5: CD44<sup>hi/DsRed+</sup> subset leads the migration of CD44<sup>lo/DsRed-</sup> subset purified from T47D cells from 48h to 72h
- Movie 6: The CD44<sup>lo</sup>/following cells sorted from MCF7 cells, carrying a CD44 promoter biosensor in which YFP reports on CD44 gene expression, can spontaneously switch to CD44<sup>hi</sup> state according to collective migration.
- Movie 7: The CD44<sup>hi</sup>/following cells sorted from MCF7 cells, carrying a CD44 promoter biosensor in which YFP reports on CD44 gene expression, can spontaneously switch to CD44<sup>lo</sup> state according to collective migration.

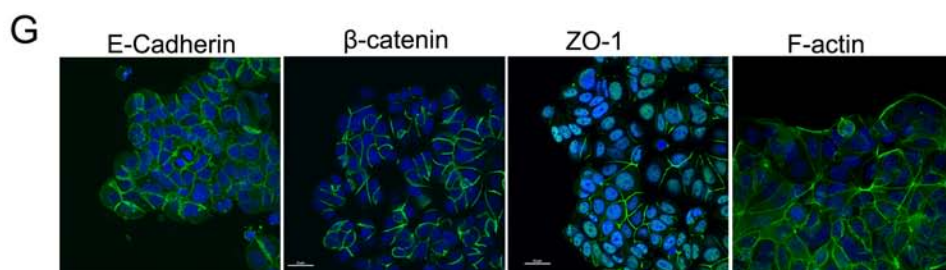
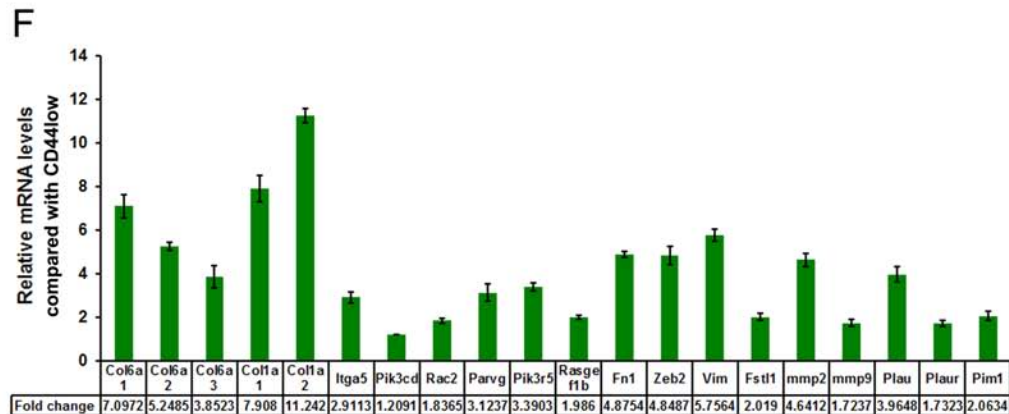
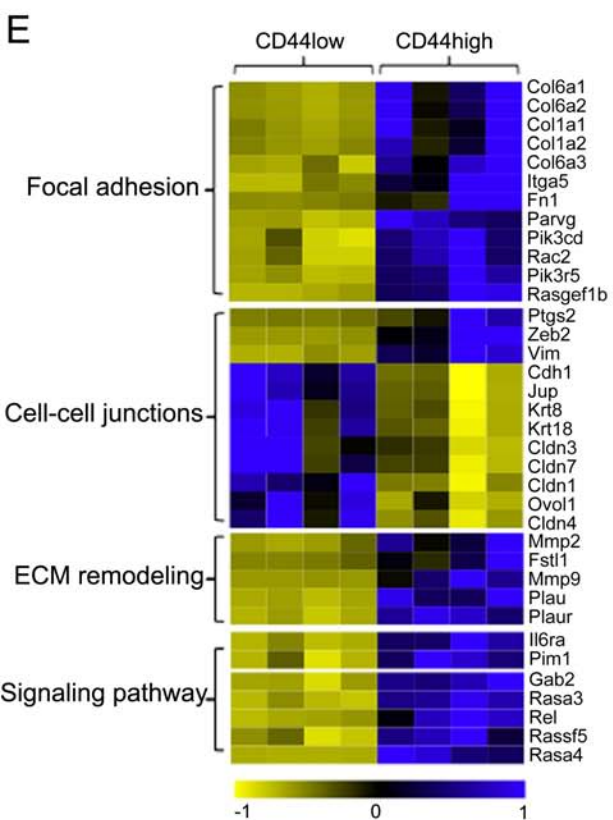
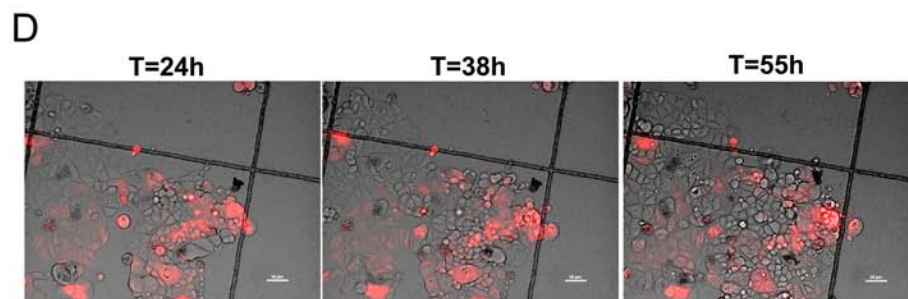
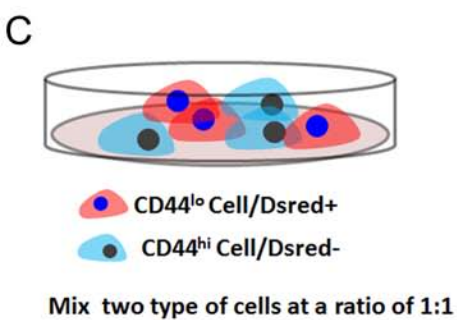
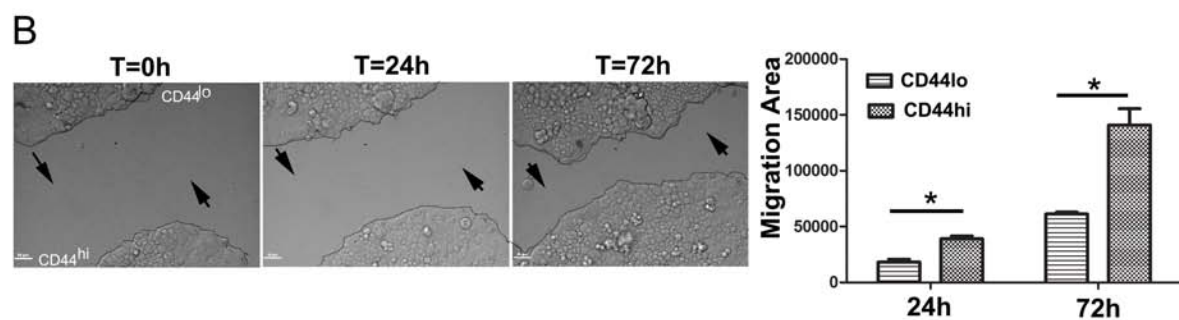
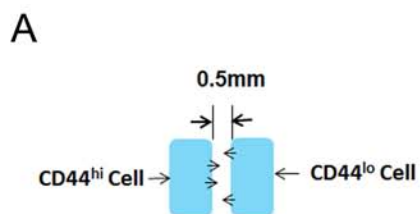


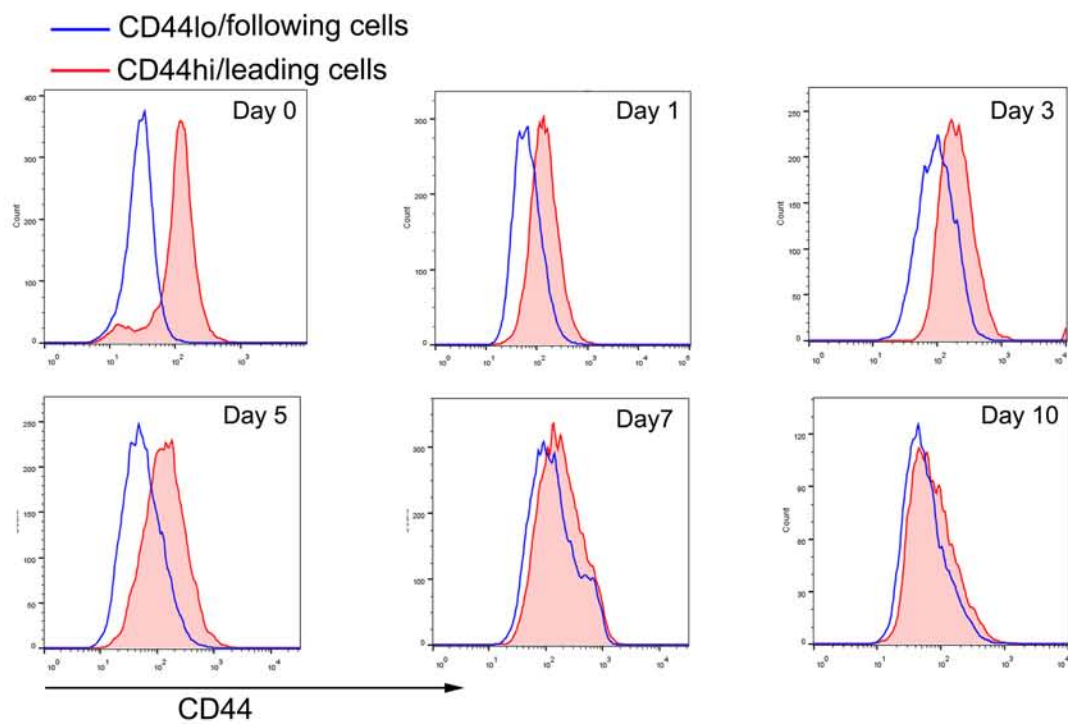
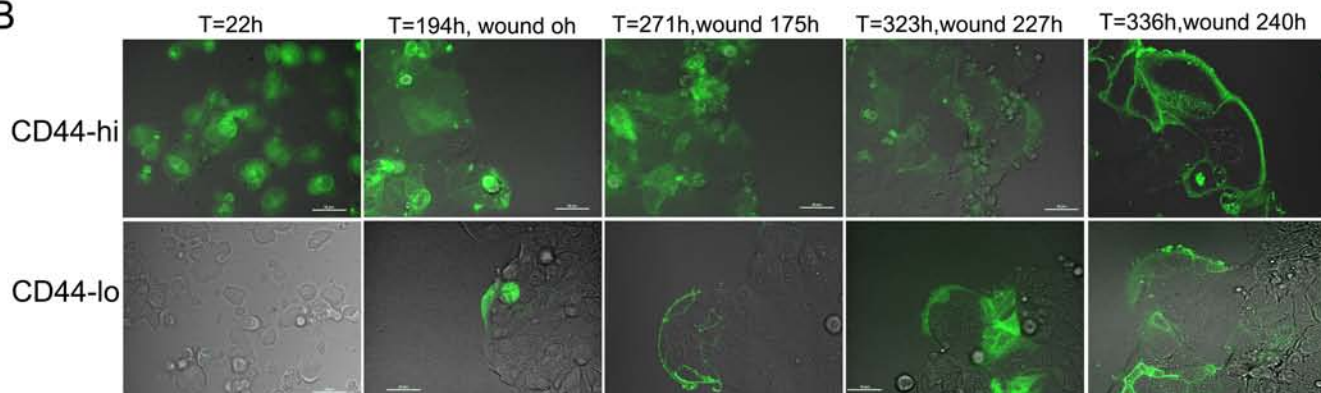
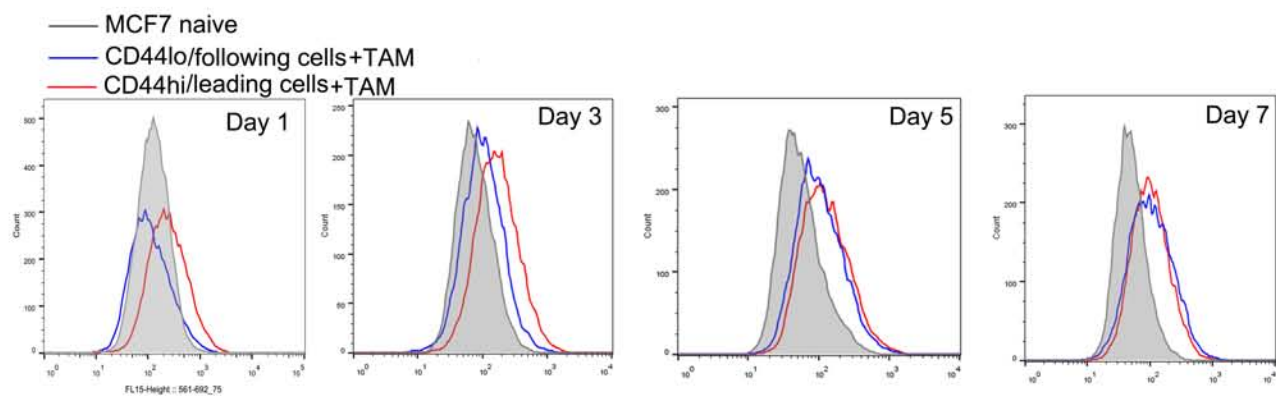
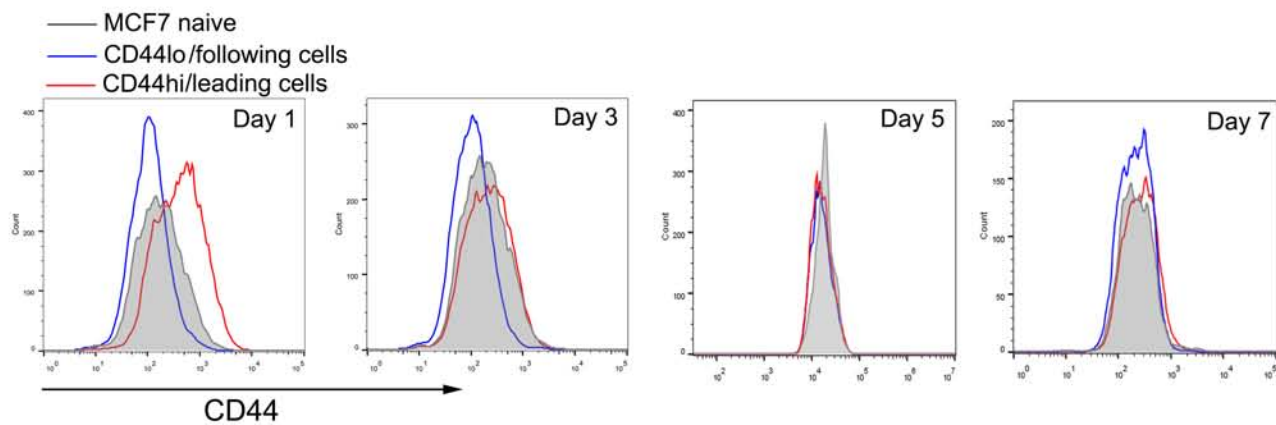


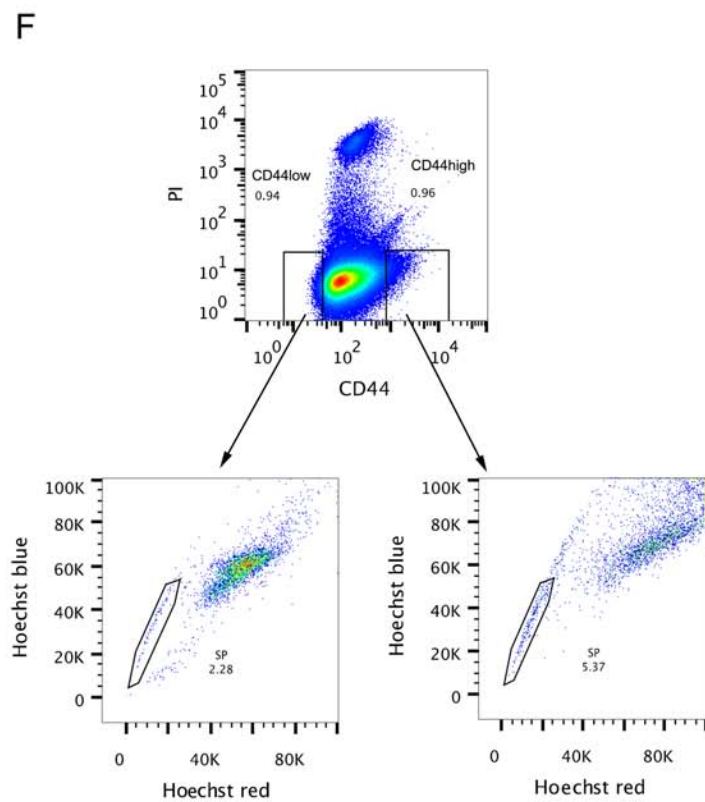
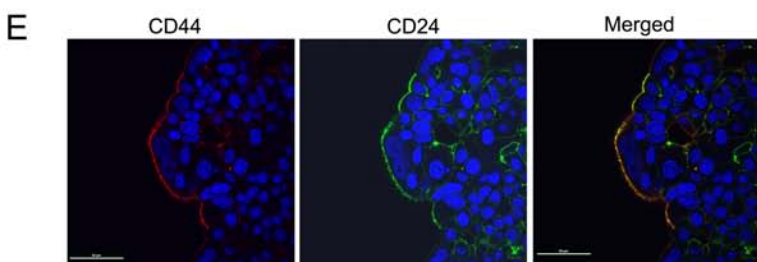
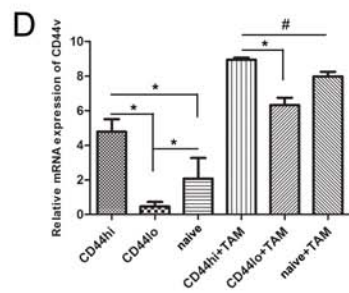
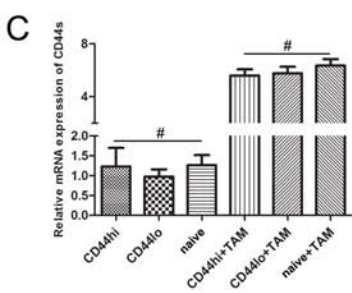
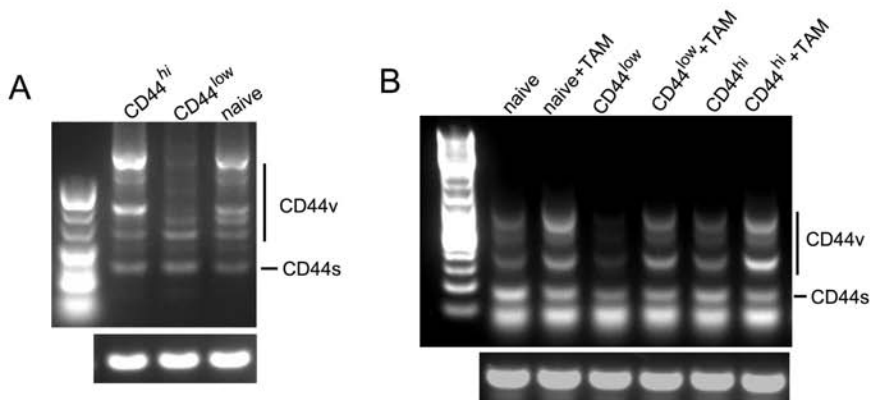
**E Tumor initiation**







**A****B****C****D**

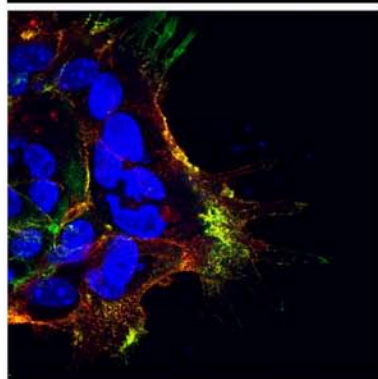
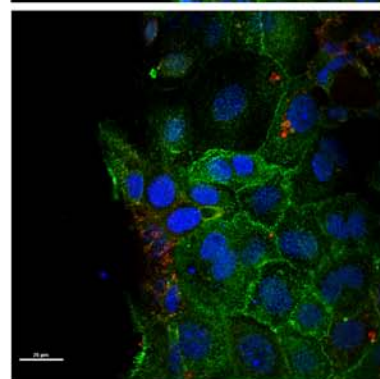
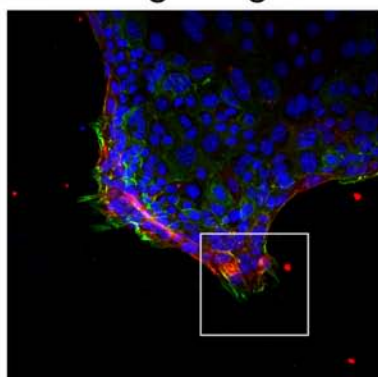
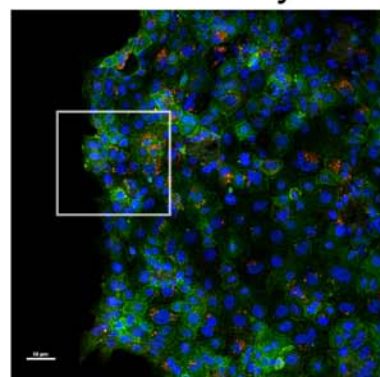




A. Ezrin/CD44

Stationary

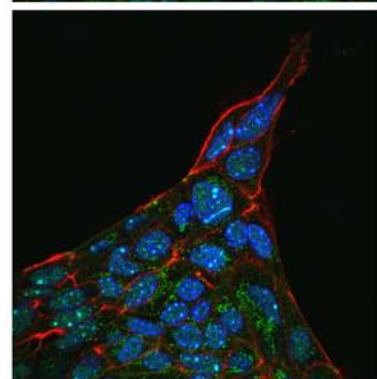
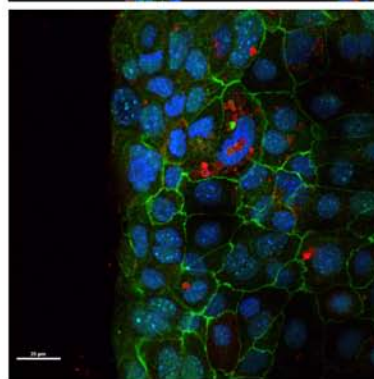
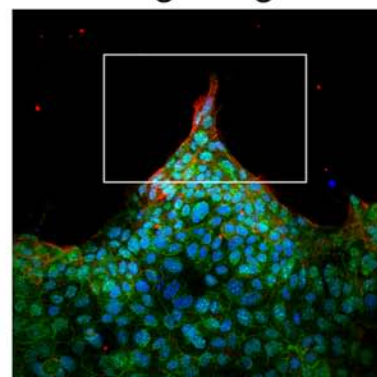
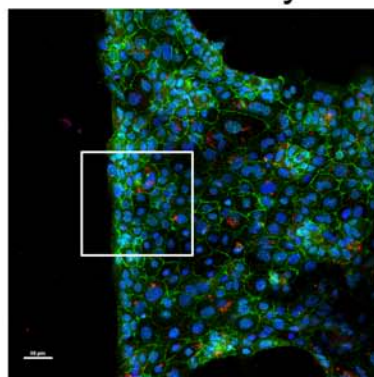
Migrating



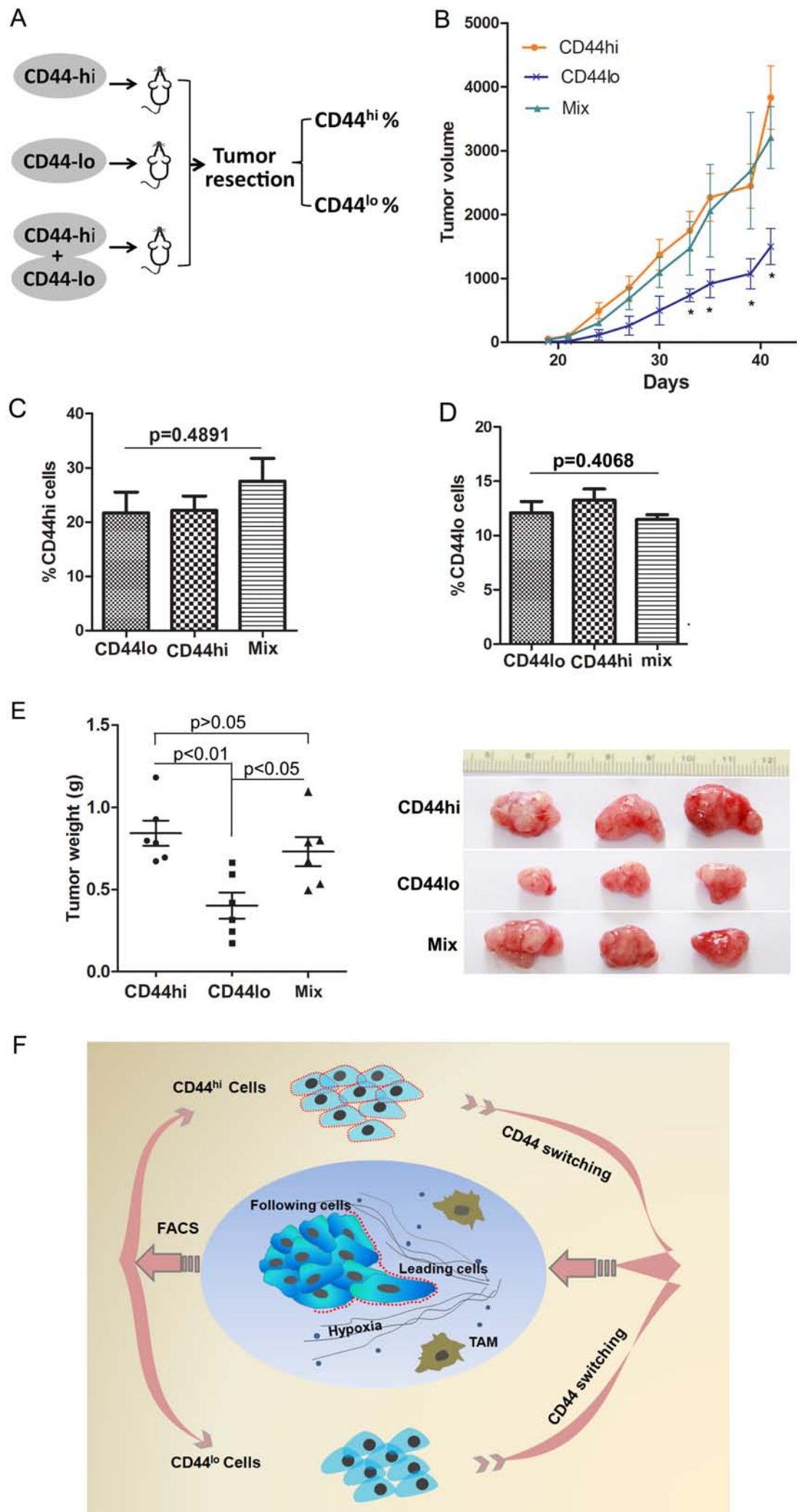
B. Merlin/CD44

Stationary

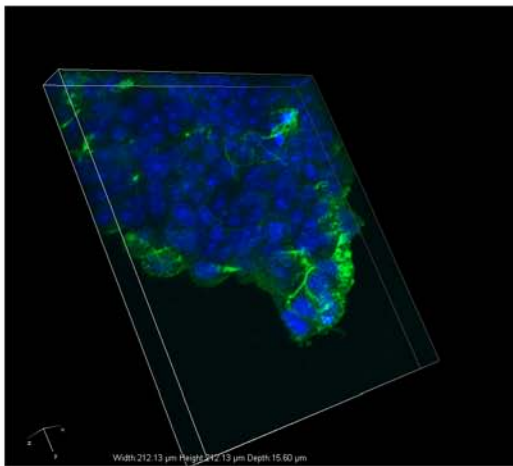
Migrating



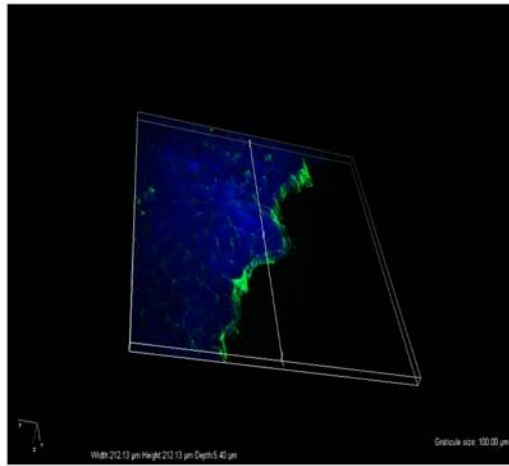




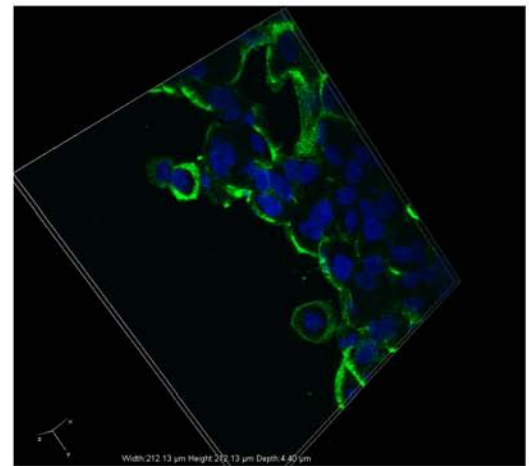
MCF7

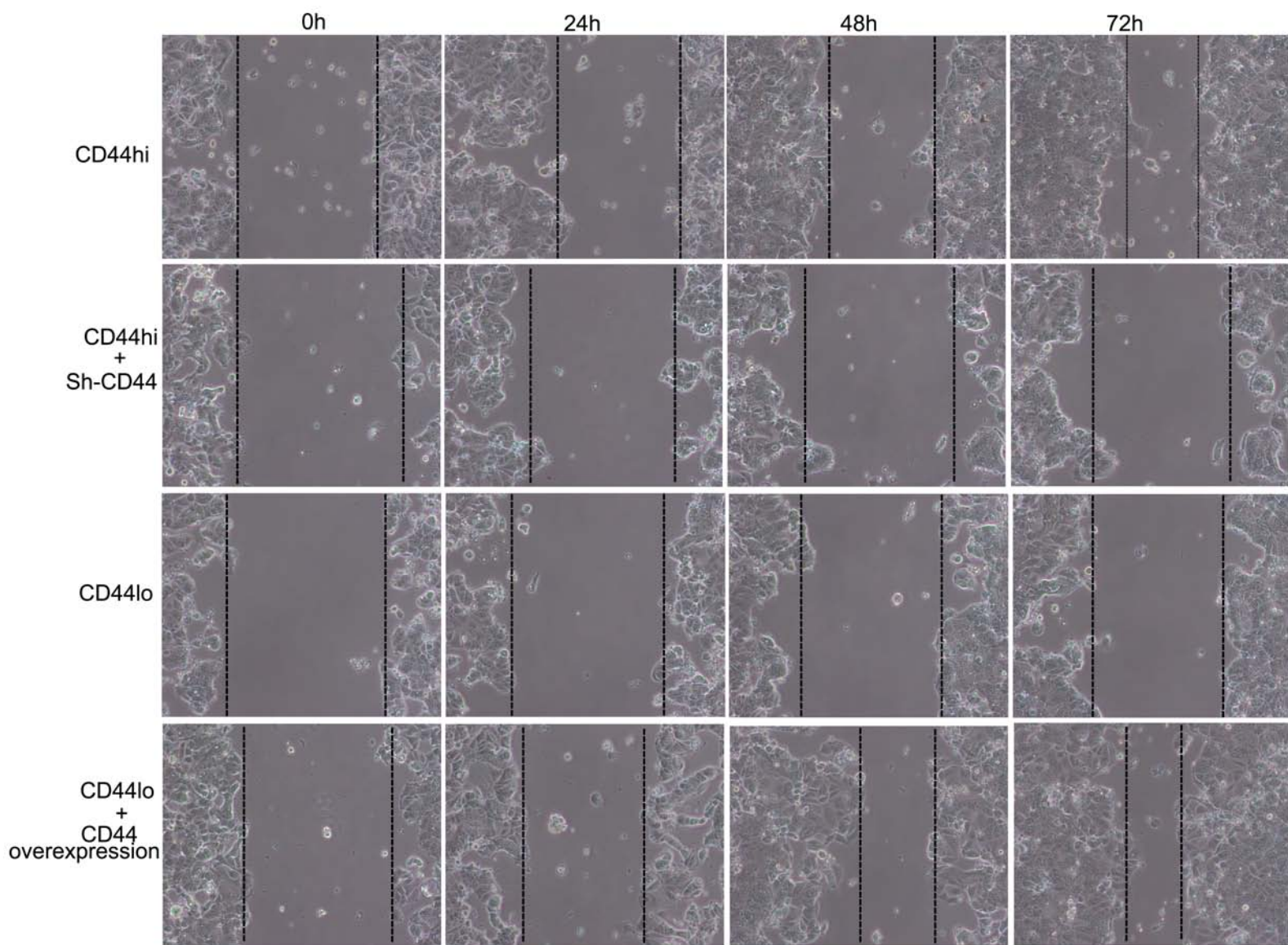


T47D



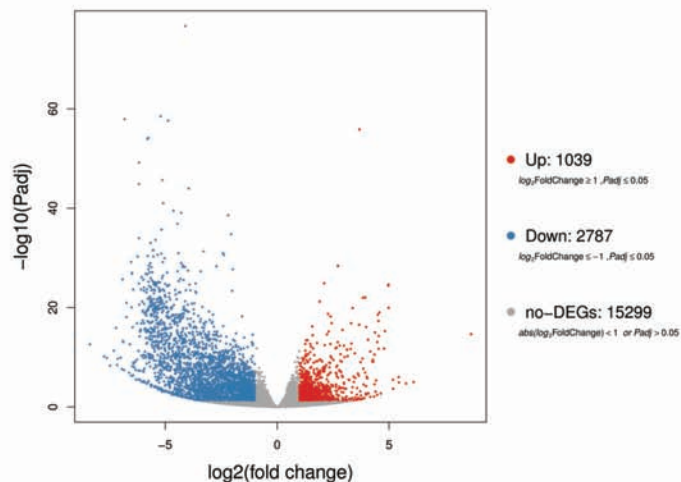
ZR-75-30



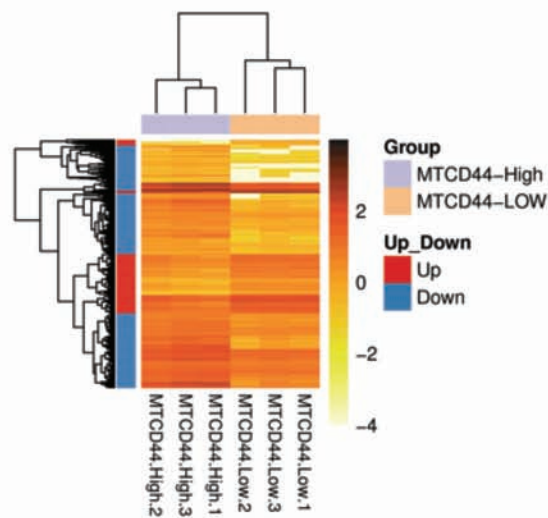


A

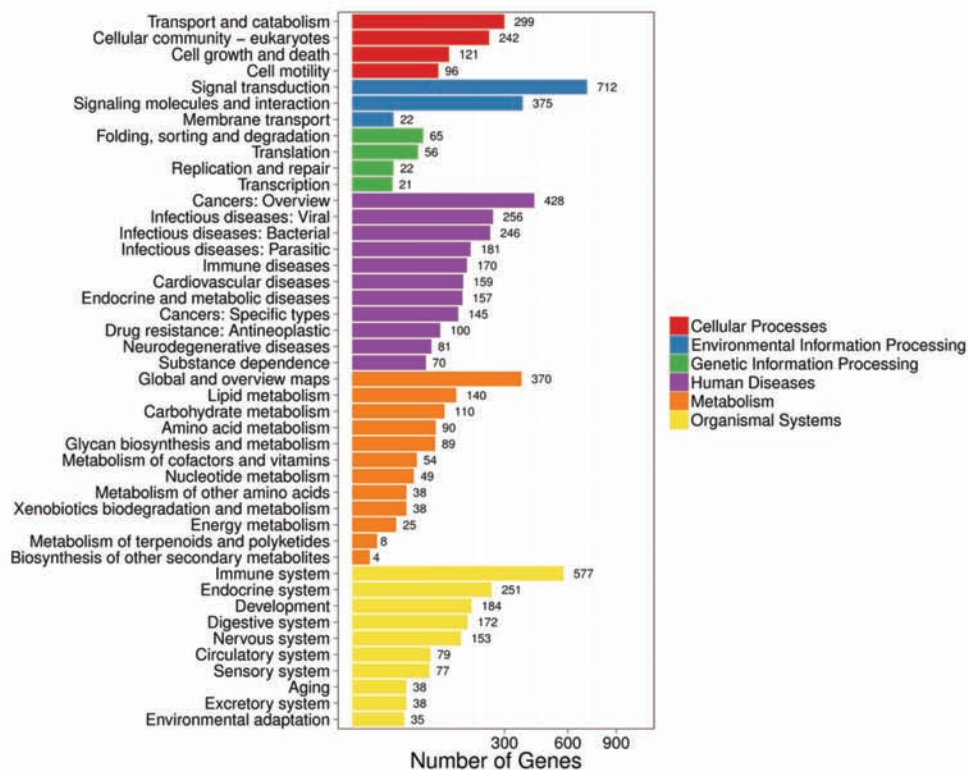
Volcano plot for MTCD44-High-VS-MTCD44-LOW.DEseq2\_Method



B

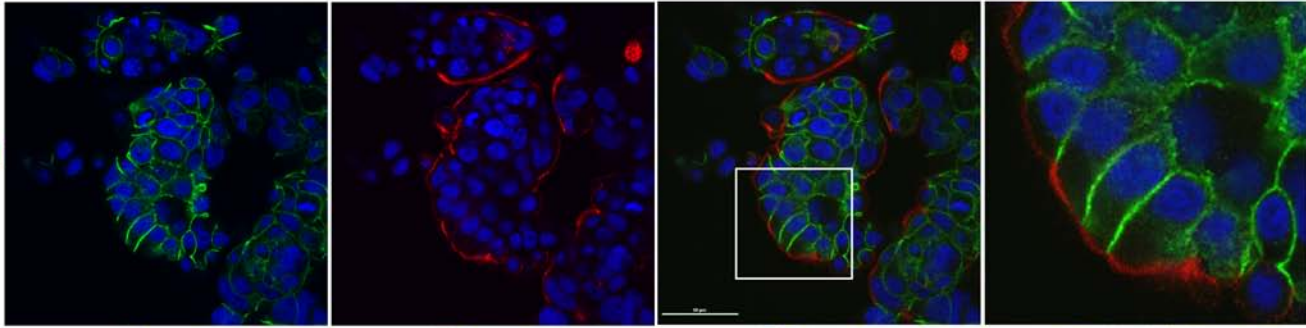


C

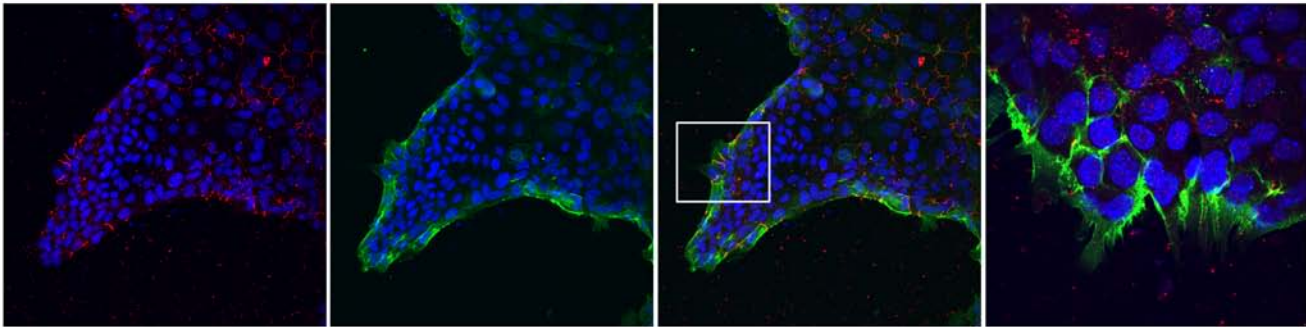




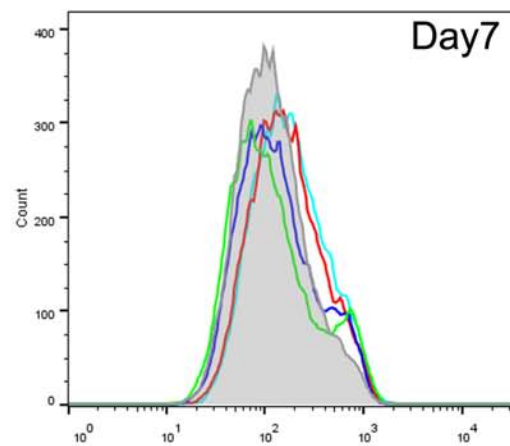
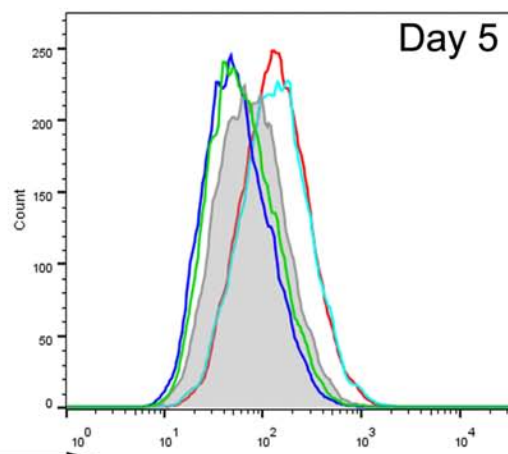
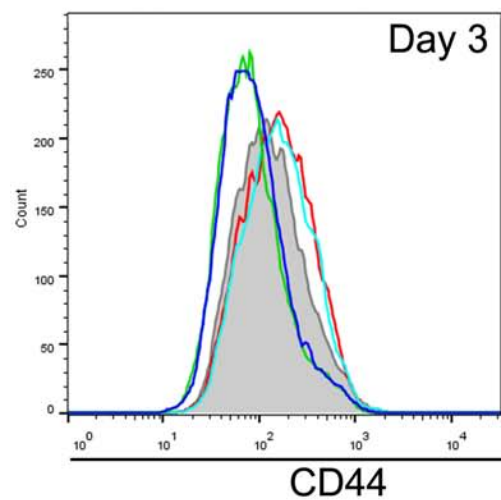
CD44/beta-Catenin



CD44/E-Cadherin



- MCF7 naive
- CD44<sup>lo</sup>/following cells
- CD44<sup>lo</sup>/following cells+IL6
- CD44<sup>lo</sup>/leading cells
- CD44<sup>lo</sup>/leading cells+IL6



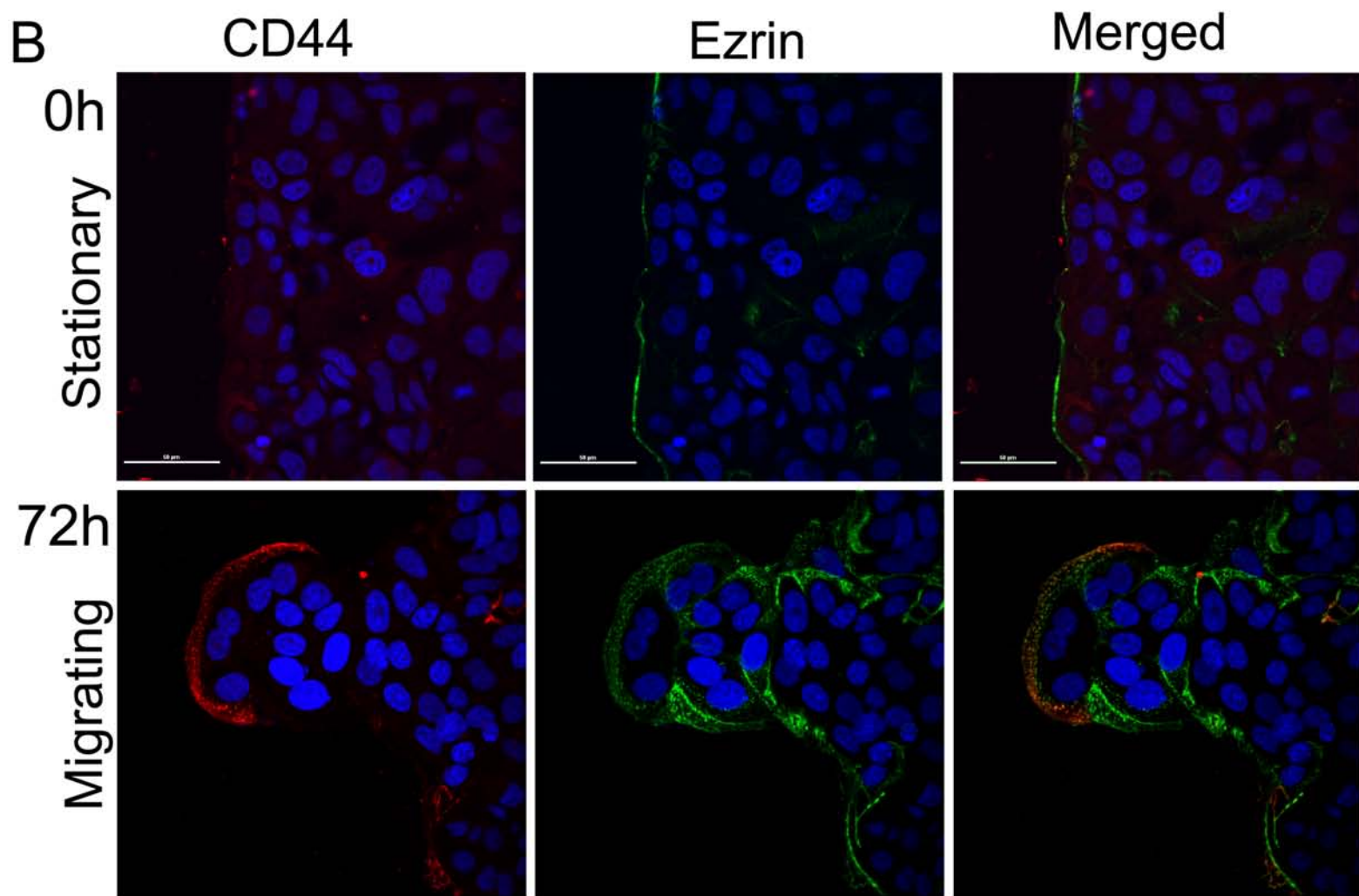
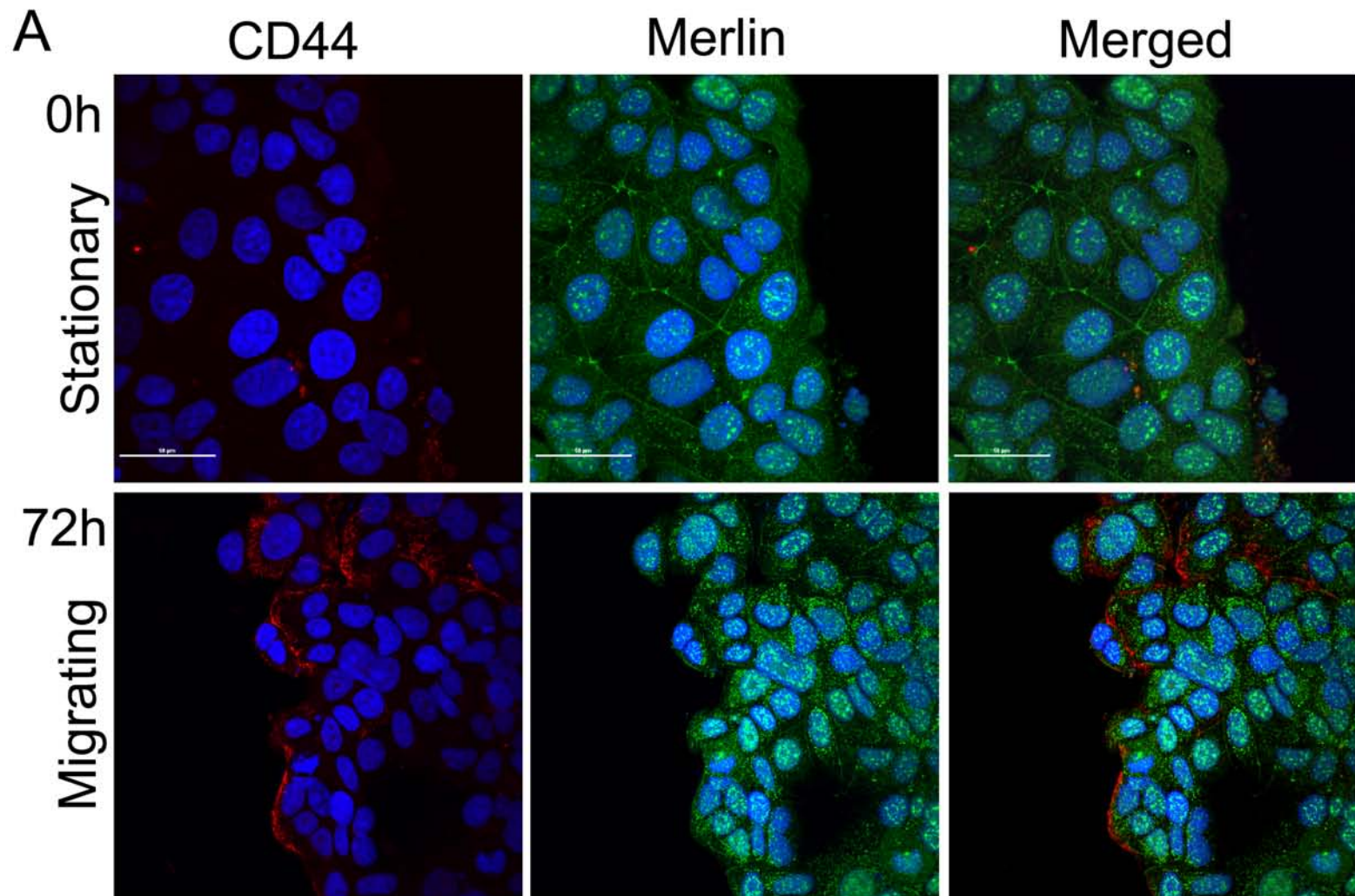




Table.S1 Correlation between CD44 expression and the clinico-pathological characteristics in the studied cohort

| Tissue marker        | n   | CD44 expression (IOD) |             | Significance   |
|----------------------|-----|-----------------------|-------------|----------------|
|                      |     | Median                | 95% CI      | <i>P</i> value |
| Age                  |     |                       |             |                |
| <50 y                | 97  | 4927                  | 3892-5962   | 0.2802         |
| ≥50 y                | 150 | 6599                  | 5558-7640   |                |
| Tumor size           |     |                       |             |                |
| <2 cm                | 108 | 4775                  | 3771-5779   | 0.4856         |
| ≥2 cm                | 149 | 5814                  | 4767-6861   |                |
| Histologic grade     |     |                       |             |                |
| 1                    | 53  | 2522                  | 1070- 10256 | 0.4792         |
| 2                    | 134 | 2755                  | 4126- 7176  |                |
| 3                    | 60  | 2537                  | 2984- 13207 |                |
| Lymphnode metastasis |     |                       |             |                |
| Yes                  | 54  | 2522                  | 2819- 7388  | 0.5405         |
| No                   | 193 | 2680                  | 4511- 7854  |                |
| ER                   |     |                       |             |                |
| Negative             | 112 | 3610                  | 6290-11240  | <0.0001*       |
| Positive             | 135 | 1930                  | 2206-3017   |                |
| PgR                  |     |                       |             |                |
| Negative             | 147 | 3588                  | 6032-10621  | <0.0001*       |
| Positive             | 100 | 1890                  | 2120-2986   |                |
| HER2                 |     |                       |             |                |
| Negative             | 126 | 3318                  | 5744-11282  | 0.0005*        |
| Positive             | 135 | 1999                  | 2918-4234   |                |
| Molecular subtype    |     |                       |             | < 0.0001*      |
| Luminal-like A       | 41  | 2060                  | 1754- 3154  |                |
| Luminal-like B       | 92  | 1963                  | 2249- 3186  |                |
| HER2 over-expressing | 44  | 2790                  | 3787- 7175  |                |
| Triple negative      | 70  | 4866                  | 7677- 17077 |                |

\*Indicated statistical significance (p <0.05)