

## **EZH2 promotes a bi-lineage identity in basal-like breast cancer cells**

Roy Z. Granit B.Sc.<sup>1</sup>, Yael Gabai B.Sc.<sup>1</sup>, Tal Hadar M.D.<sup>1,2</sup>, Yael Karamansha M.Sc.<sup>1</sup>, Leah Liberman B.Sc.<sup>1</sup>, Ithai Waldhorn<sup>1</sup>, Irit Gat-Viks Ph.D.<sup>5,6</sup>, Aviv Regev Ph.D.<sup>5</sup>, Bella Maly M.D.<sup>3</sup>, Tamar Peretz M.D.<sup>4</sup> and Ittai Ben-Porath Ph.D.<sup>1\*</sup>

<sup>1</sup> Department of Developmental Biology and Cancer Research, Institute for Medical Research – Israel-Canada, Hadassah School of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel, 91120

<sup>2</sup> Department of Surgery

<sup>3</sup> Department of Pathology

<sup>4</sup> Sharett Institute for Oncology

Hadassah Medical Center, Jerusalem, Israel, 91120

<sup>5</sup> Broad Institute of Harvard and MIT and Department of Biology, MIT, Cambridge Massachusetts, USA, 02142

<sup>6</sup> Current address: Department of Immunology, Tel-Aviv University, Tel-Aviv, Israel

\* **Correspondence:** Ittai Ben-Porath, E-mail: [ittaip@mail.huji.ac.il](mailto:ittaip@mail.huji.ac.il), Tel: 972-2-6757006, Fax: 972-2-6415848

**Running title:** EZH2 controls basal-like breast cancer differentiation state

## **Abstract**

The mechanisms regulating breast cancer differentiation state are poorly understood. Of particular interest are molecular regulators controlling the highly aggressive and poorly differentiated traits of basal-like breast carcinomas. Here we show that the Polycomb factor EZH2 maintains the differentiation state of basal-like breast cancer cells, and promotes the expression of progenitor-associated and basal-lineage genes. Specifically, EZH2 regulates the composition of basal-like breast cancer cell populations by promoting a “bi-lineage” differentiation state, in which cells co-express basal- and luminal-lineage markers. We show that human basal-like breast cancers contain a subpopulation of bi-lineage cells, and that EZH2-deficient cells give rise to tumors with a decreased proportion of such cells. Bi-lineage cells express genes that are active in normal luminal progenitors, and possess increased colony formation capacity, consistent with a primitive differentiation state. We found that GATA3, a driver of luminal differentiation, performs a function opposite to EZH2, acting to suppress bi-lineage identity and luminal-progenitor gene expression. GATA3 levels increase upon EZH2 silencing, mediating a decrease in bi-lineage cell numbers. Our findings reveal a novel role for EZH2 in controlling basal-like breast cancer differentiation state and intra-tumoral cell composition.

**Keywords:** Basal-like, breast cancer, differentiation, EZH2, Polycomb, GATA3

## **Introduction**

The mechanisms dictating the differentiation state of cancer cells are poorly understood, and their elucidation is essential for understanding cancer etiology as well as for therapy development. Breast cancers are grouped into several major subtypes that differ in their typical progression course, as well as in their differentiation traits (1, 2). Ongoing advances in the study of the lineages in the normal breast allow improved analyses of the links between normal cell types and tumor subtypes (1).

Breast carcinomas of the basal-like subtype are of particular interest. These tumors, which comprise approximately 10% of cases, are highly aggressive, and since no targeted therapy currently exists for their treatment they represent a major clinical challenge (2, 3). Unlike the majority of breast cancers, which express markers of the luminal lineage in the normal breast, basal-like tumors express markers of both the basal (or myoepithelial) and the luminal lineages (4, 5). This, as well as the poor histopathological differentiation typical of these tumors, suggests that they possess a stem or progenitor-like phenotype. Supporting this notion, these tumors express an embryonic stem (ES) cell-like signature: high expression of ES-enriched genes and repression of Polycomb targets (6). Gene expression analyses have established that in fact basal-like tumors are most similar to luminal-lineage progenitors in the normal mammary gland, from which they may originate (7-9). The molecular regulators controlling the differentiation state of basal-like tumors are largely unknown.

The Polycomb complex is a major regulator of stem cell identity, and an important link between stem cells and cancer (10, 11). EZH2, the catalytic subunit of the Polycomb Repressive Complex 2 (PRC2), methylates lysine 27 of histone H3 (H3K27me3) on target genes, leading to chromatin condensation and transcriptional silencing (10). Polycomb maintains stem cell states by silencing differentiation-inducing genes, and is often essential for proper differentiation (10, 11). EZH2 is frequently overexpressed in aggressive, metastatic tumors (12, 13) and mutations in this gene are found in

hematological malignancies (14). In breast cancers its expression is associated with high grade, ER-negative status, the basal-like subtype, and poor prognosis (13, 15, 16). EZH2 promotes cancer cell proliferation, anchorage-independent growth and invasiveness (13, 17-19). *In vivo*, EZH2 inhibition reduces tumor growth rates to various extents (18-21). However, the functions of EZH2 in controlling cancer cell differentiation remain poorly understood.

The known functions of Polycomb and the expression pattern of EZH2 in human breast cancers led us to examine the effects of EZH2 on breast cancer cell differentiation state. Our findings reveal that EZH2 promotes bi-lineage identity in basal-like breast cancer cells as well as progenitor-associated traits, thereby controlling subtype identity and intra-tumoral cell composition.

## Results

### **EZH2 maintains the gene expression signature associated with basal-like breast cancers**

Analysis of published breast cancer gene expression profiles (6), as well as immunohistochemical staining of an independent patient sample set, indicated that EZH2 overexpression is strongly associated with the basal-like breast cancer subtype, as previously reported (15) (**supplementary Figure 1**). Of other PRC2 components, EED expression was also associated with the basal-like subtype, to a less significant degree (**supplementary Figure 1**). Known targets for Polycomb repression (22) were negatively correlated with EZH2 levels, consistent with the potentially elevated Polycomb activity in these tumors; in contrast, genes highly expressed in normal luminal progenitor cells (7) showed a strong positive correlation with EZH2 expression (**supplementary Figure 1**). These associations suggested that EZH2 function contributes to the differentiation state of basal-like cancers. To test this, we stably silenced EZH2 in a panel of human breast cancer cell lines representing different tumor subtypes (**Figure 1a** and **supplementary Figure 2**). As faithful models of basal-like cancers we considered lines that maintain epithelial identity (have not undergone an epithelial to mesenchymal transition), express both basal and luminal cytokeratins, and display expression profiles and markers specific to this subtype (**supplementary Figure 3**). We found that in basal-like cell lines (HCC70, MDA-MB-468) silencing of EZH2 led to reduced expression of basal cytokeratins – CK5 and CK14 – accompanied by increased expression of the luminal cytokeratin CK18 (**Figure 1a**), suggesting a shift away from basal differentiation. Consistent with this, overexpression of EZH2 increased the expression of CK5 and CK14 (**Figure 1b**). These changes in cytokeratin expression were observed using two different hairpins targeting EZH2, as well as upon silencing of EED (**supplementary Figure 2**). Levels of the luminal CK18 were not, however, affected by EZH2 overexpression or by its silencing in luminal cells (**Figure 1b** and **supplementary Figure 2**). These

findings were supported by qRT-PCR analysis of additional lineage markers, which indicated that basal-lineage gene expression was reduced by silencing of EZH2 or EED, and promoted by EZH2 overexpression, with changes in luminal markers being less dramatic at the mRNA level (**Figure 1c**).

To obtain a more detailed picture of the effects of EZH2 on the differentiation state of the tumor cells, we performed expression profiling of HCC70 cells silenced for EZH2 or for EED, or overexpressing EZH2, as well as MDA-MB-468 cells silenced for EZH2. We analyzed changes in the expression of previously compiled gene sets associated with different breast cancer subtypes (23) (**supplementary Figure 4**). Genes highly expressed in basal-like tumors were preferentially suppressed upon EZH2 or EED silencing; EZH2 overexpression promoted the expression of a smaller number of these genes (**Figure 1d-f and supplementary File 1**). To gain further resolution, we compiled 12 sets of genes coordinately expressed across breast cancers, each associated with different cancer subcategories and gene functions (**supplementary Figure 4 and supplementary File 1**). Specific gene sets associated with high-grade and basal-like breast cancers were preferentially repressed in EZH2-silenced cells, and preferentially upregulated in EZH2 overexpressing cells (**supplementary Figure 4**). Upon EZH2 silencing, preferential upregulation of gene sets associated with luminal and claudin-low tumors were observed; however, these were less consistent across the different samples (**supplementary Figure 4**).

In light of the expression of luminal progenitor-associated genes in basal-like breast cancer cells (7), we tested the effect of EZH2 on these genes. Luminal progenitor-associated genes were preferentially downregulated in EZH2- and EED-silenced cells, and upregulated in EZH2-overexpressing cells (**Figure 1f-h**). Furthermore, FACS analyses revealed that CD133 (PROM1) and c-KIT, which are both expressed specifically in luminal progenitors in the normal breast (7) were dramatically downregulated upon EZH2 silencing (**Figure 1i,j**).

Together these analyses indicated that EZH2 maintains aspects of the gene expression program active in basal-like breast cells, specifically, basal-lineage and luminal progenitor-associated genes. However, we noted that EZH2 silencing did not result in a full phenotypic conversion or in the acquisition of an alternative differentiation identity, such as luminal differentiation.

### **EZH2 regulates the size of a bi-lineage cell fraction**

The expression of both luminal and basal cytokeratins is a central characteristic of basal-like breast cancers (4, 5) and of derived cell lines. We were intrigued whether these cancer cell populations are homogeneously comprised of cells that co-express both luminal and basal markers, or, instead, contain subpopulations that express distinct lineage markers.

We first examined this question in human basal-like breast cancer samples. We co-stained sections of ten basal-like tumors for CK14 and CK18. All tumors contained cells expressing only CK18 or CK14; in addition, nine out of ten tumors contained double-positive (CK18+CK14+) “bi-lineage” cells (**Figure 2a,b**). In all cases, the bi-lineage cells were a minority of cells (average 5%, range 0.5-18%). Often, relatively small tumor regions contained single- and double-positive cells in close proximity (**Figure 2a**), suggesting that differentiation state, as assessed by these markers, is dynamic even within close cell clones.

We next studied whether this composition is represented in basal-like breast cancer cell lines. Immunofluorescent staining and FACS analysis revealed that populations of HCC70 and MDA-MB-468 cells contained approximately 22% and 34%, respectively, of CK14+CK18+ bi-lineage cells, while the remainder of cells expressed the luminal CK18 only (**Figure 2c,d**). These lines therefore partly recapitulate the composition of human tumors.

In light of the observed heterogeneity within cancer cell populations, we examined the effects of EZH2 on population composition. We found that EZH2 silencing resulted in a substantial decrease in the fraction of bi-lineage cells to approximately half of its original size, as did silencing of EED (**Figure 3a,c, supplementary Figure 5**); conversely, EZH2 overexpression caused an increase in the bi-lineage fraction size (**Figure 3b,c**).

This was recapitulated in tumors formed by basal-like breast cancer cells upon transplantation in mouse mammary glands. EZH2-silenced MDA-MB-468 cells formed tumors that were slower in growth than those formed by control cells (**Figure 3d**), and these tumors contained a smaller fraction of bi-lineage cells (**Figure 3e,f**). These findings indicate that EZH2 controls the composition of the cancer cell population, promoting the presence bi-lineage cells at the expense of CK18-only cells.

### **EZH2 promotes the colony formation capacity of basal-like breast cancer cells**

We next asked whether the effect of EZH2 on the composition of the cell population was manifested in the behavior of the population. We tested the effects of EZH2 function on colony formation capacity. The ability to form colonies in liquid or on extracellular matrix (Matrigel) is a hallmark of normal mammary stem and progenitor cell populations, and of stem-like breast cancer cells (7, 24, 25). Normal luminal progenitors typically form hollow acini in 3D (7); in contrast, cancer cells form colonies lacking organized structure (26), and expression of a single oncogene is often sufficient to disrupt lumen formation (27, 28). We found that HCC70 and SUM149 cell populations seeded on Matrigel formed rounded colonies (filled spheres/acini) (**Figure 4a** and **supplementary Figure 6**). We calibrated conditions such that >96% of colonies were monoclonal. Similar colonies were formed when the cells were grown in semi-liquid conditions recently developed for the growth of primary stem and progenitor cells in 3D (29). In both conditions colony formation efficiency was partial: only



20-30% of seeded cells formed colonies. We noted that the majority (70%) of formed 3D colonies contained both CK18-only cells and bi-lineage cells (**Figure 4a** and **supplementary Figure 6**). The single cells forming these colonies can therefore give rise to both cell types, consistent with a change in differentiation state.

We found that EZH2-silenced HCC70 and SUM149 cells formed significantly fewer colonies than control cells (**Figure 4b-d**). EZH2-silenced cells showed mildly slower growth rates in conventional 2D culture, with no increase in apoptosis (**supplementary Figure 6**); however, an inherently slower proliferation rate would be expected to affect colony growth rates rather than the percentage of cells initiating colony formation. Indeed, colonies formed by EZH2-silenced cells were on average smaller than those formed by control cells, but to a degree less significant than the decrease in colony numbers (**Figure 4b-d**). Upon dissociation of colonies and replating, EZH2-deficient cells once again formed fewer colonies, indicating that this reduced capacity was maintained within the colonies (**Figure 4b**). Conversely, EZH2 overexpression increased the numbers of colonies formed by HCC70 cells, which were also larger in size (**Figure 4e**). Together these findings indicate that EZH2 promotes the colony formation capacity of the cancer cell population.

### **Bi-lineage cells possess increased colony formation capacity**

We next assessed whether colony formation capacity reflected the phenotypic diversity in the population. Recent studies of the developing mouse mammary gland have revealed that co-expression of CK14 and CK18 occurs in mammary stem cells during embryogenesis, but not subsequently in the adult gland (30, 31); such cells can be detected, however, in mouse mammary tumors (32). In the adult human breast, progenitor and stem cell-enriched subpopulations appear to contain cells that express bi-

lineage cytokeratins (7, 33). Bi-lineage marker expression in cancer cells could therefore represent a more primitive differentiation state.

We therefore assessed whether bi-lineage and CK18-only cells differ in their colony formation capacity. First, we followed the population dynamics during colony formation. 1 day after plating of HCC70 cells on Matrigel, a timepoint in which the culture largely contained single cells, 44% of cells showed bi-lineage CK14+CK18+ staining (**Figure 5a,b**). 4 days after plating, these cultures contained single cells that have not divided, as well as small clusters containing ~5 cells on average, indicating that approximately two cell divisions took place. Strikingly, the percentage of bi-lineage cells among the non-dividing single cells was decreased relative to the one-day time point, to 26% ( $P=0.007$ ) (**Figure 5b**); in contrast, 91% of the forming colonies contained bi-lineage cells (**Figure 5b**). 8 days after plating, this trend was further enhanced: only 21.8% of single cells were bi-lineage ( $P=0.004$ ), while 96% of colonies contained bi-lineage cells (**Figure 5b**). These data indicate that CK18-only cells preferentially remain undivided following seeding, while bi-lineage cells preferentially divide and form colonies.

To further support these findings, we generated a lentiviral reporter construct in which GFP expression is driven by the CK14 promoter (K14p-GFP) (**Figure 5c**). Since all cells in this population express CK18, the CK14 reporter was expected to label bi-lineage cells. We infected HCC70 cells with the K14p-GFP lentivirus and sorted GFP-high cells and GFP-negative cells. We found that GFP-high cells were indeed enriched for bi-lineage cells (**supplementary Figure 7**). Upon plating on Matrigel, GFP-high cells isolated from the K14p-GFP-infected population formed more colonies than GFP-negative cells, and these were also larger in size (**Figure 5c**). In contrast, GFP-high cells isolated from a population of cells infected with a virus constitutively expressing GFP, formed similar numbers of

colonies as GFP-negative cells isolated from the same population (**Figure 5d**). These findings provide direct evidence that bi-lineage cells possess increased colony initiation capacity.

We next compared the expression of lineage markers in bi-lineage and CK18-only cells. We sorted stained HCC70 cells to obtain these two subpopulations. Bi-lineage cells expressed, as expected, higher levels of basal-lineage genes; in addition, they expressed higher levels of luminal progenitor-associated genes (**Figure 5e**). This indicates that the bi-lineage cells are closer to luminal progenitors at the molecular level than the CK18-only cells.

Together, these findings support the hypothesis that bi-lineage cells within basal-like breast cancer cell populations represent a more primitive differentiation state than their counterparts, which is manifested in increased colony formation capacity and the expression of progenitor-associated genes. The promotion of colony formation by EZH2 can therefore be explained by its promotion of bi-lineage identity.

### **Bi-lineage cell numbers increase during early tumor formation**

In light of the increased ability of bi-lineage cells to form colonies in culture, we tested the dynamics of bi-lineage cell numbers within tumor cells implanted in mice. We injected MDA-MB-468 cells into mouse mammary glands, and then extracted the injected tissue at different time points, from 2 hours after injection to 3 weeks subsequently. Staining of tissue sections for CK14 and CK18 revealed that the numbers of bi-lineage cells progressively increased in these forming tumor nodules (**Figure 6a,b**). This phenomenon suggests that the bi-lineage identity provides an advantage during tumor growth.

## **GATA3 suppresses bi-lineage identity and luminal progenitor-associated genes**

We next considered which factors could mediate the decrease in bi-lineage cell numbers upon EZH2 silencing. We noted that EZH2-silenced cells showed increased expression of GATA3 and FOXA1, both known Polycomb targets and regulators of luminal identity in the normal and cancerous breast (34-37) (**Figure 7a**). In the normal gland GATA3 promotes the differentiation of luminal progenitors into mature luminal cells (34, 35). GATA3 expression is associated with luminal breast cancers, and it functions as a tumor suppressor gene (36-39). The expression levels of EZH2 and GATA3 in human breast cancers are strongly negatively correlated (**supplementary Figure 8**). However, while GATA3 levels are typically very high in luminal tumors, it is also expressed, albeit at lower levels, in basal-like tumors and cell lines and could therefore influence their differentiation state (**supplementary Figures 8**).

We therefore tested whether GATA3 affects the composition of basal-like breast cancer cell populations. Consistent with its described function, GATA3 overexpression repressed luminal progenitor-associated genes and basal-lineage genes, and induced luminal-lineage genes (**Figure 7b**). Furthermore, GATA3 overexpression dramatically reduced bi-lineage cell numbers, beyond the reduction observed upon EZH2 silencing (**Figure 7c**). Conversely, GATA3 silencing led to increased expression of luminal progenitor-associated and basal-lineage genes, and to an increase in the bi-lineage fraction (**Figure 7d,e**). These findings indicate that GATA3 suppresses bi-lineage identity within the basal-like breast cancer cell populations, and concomitantly represses luminal progenitor-associated genes.

To test whether GATA3 contributed to the decreased bi-lineage fraction observed upon EZH2 silencing, we co-infected HCC70 cells with shEZH2 and shGATA3 lentiviruses; GATA3 silencing

prevented to a large extent the decrease in bi-lineage cell numbers caused by loss of EZH2 (**Figure 7f**). We noted that GATA3 overexpression caused a reduction in EZH2 levels, while its silencing led to a reproducible increase in EZH2 levels (**Figure 7b,d**), suggesting a negative feedback loop between the two genes. Together these findings reveal that GATA3 performs a role opposite of that of EZH2, acting to decrease bi-lineage identity in favor of luminally directed identity. The interplay between the two proteins can therefore determine the composition of the cancer cell population (**Figure 7g**).

## Discussion

Little is known about the molecular regulators of poor differentiation and stem- or progenitor-like phenotypes in cancer cells. Basal-like breast cancers are unique in that they present a mixed lineage phenotype, expressing both basal and luminal genes (3, 4), as well as genes specifically expressed in luminal progenitors (7). Our results reveal a novel function for EZH2 in determining the differentiation state of these tumor cells. EZH2 promotes the expression of basal-lineage markers, as well as of luminal progenitor-associated genes. The absence of EZH2 is not, however, sufficient for cells to fully activate an alternative differentiation program, which would most likely require the concomitant activity of differentiation-promoting transcription factors.

We gained further insight by assessing the effects of EZH2 on the composition of the cancer cell population. Basal-like breast cancers contain a subpopulation of bi-lineage cells co-expressing CK14 and CK18, and EZH2 increases the relative fraction size of bi-lineage cells, at the expense of cells expressing only the luminal CK18. The gene expression changes observed in whole cell populations upon EZH2 silencing or overexpression therefore reflect, at least in part, these changes in cell composition. EZH2 thus controls tumor identity as a whole, as well as intra-tumor cellular heterogeneity.

We demonstrate that the bi-lineage identity reflects a differentiation state that is more primitive, and closer to that of normal progenitors, than that of the CK18-only cells. This is based on the expression of progenitor-associated genes and by the enhanced capacity of these cells to form heterogeneously-composed colonies. The changes in bi-lineage cell numbers upon EZH2 silencing or overexpression thus readily explain the corresponding changes in colony formation by these cells. Further characterization of the dynamics in which cells may exit and/or enter the bi-lineage state will shed additional light on its nature.

Prior descriptions of bi-lineage cells in the breast are consistent with our findings. Recent studies of the developing mouse mammary gland (30, 31) have revealed that bi-lineage cells exist among embryonic mammary stem cells that subsequently give rise to the two mature lineages. Bi-lineage cells appear to be absent from the adult mouse mammary gland (30), but have been detected in mouse mammary tumors (32), suggesting an adoption of this primitive differentiation state by the cancer cells. Bi-lineage cells do appear to exist in the adult human mammary gland (7, 33, 40): luminal progenitors contain a fraction of 41-50% cells co-expressing CK5 and CK18 (7), and mammary cells of BRCA1 mutation carriers, whose progenitor pools are expanded, generate colonies with increased numbers of CK14+CK18+ cells (33). Furthermore, stem cells isolated through label retention from normal and cancerous breasts (hNMSCs) express both luminal and basal markers (25). This marker profile therefore appears to be indicative of increase plasticity.

EZH2 has been shown to promote the size of the CD24-/CD44+ sphere forming population in breast cancers by driving Raf1 amplification through suppression of Rad51 and DNA damage repair (41). Our study is consistent with these findings in ascribing a role for EZH2 in promoting colony formation. However, we did not observe changes in Raf1 levels in basal-like cells upon EZH2 overexpression, and the kinetics of colony formation are not consistent with an amplification-driven mechanism, suggesting independent modes of action in the two systems.

EZH2-silenced cells give rise to slow growing tumors, as has been previously shown (18-21).

Strikingly, upon implantation of basal-like breast cancer cells in mouse mammary gland, the numbers of bi-lineage cells increase, indicating either that these cells have a growth or survival advantage *in vivo*, or that cells adopt this phenotype upon implantation. The manner by which bi-lineage identity and its control by EZH2 may affect tumor growth and progression requires further study.

Our finding that GATA3 represses bi-lineage identity in favor of a more luminally differentiated state is consistent with its described role as a master inducer of luminal differentiation (34, 35). However, our results shed new light on GATA3 function in breast cancer, demonstrating that it contributes also to the differentiation state of basal-like cancers, in which it is expressed at lower levels than in luminal tumors.

Together, our findings reveal a novel mechanism central to the determination of breast cancer subtype identity, and point to EZH2 as a promoter of the basal-like subtype. Furthermore, our work provides new insights into the complexity of intra-tumoral cellular heterogeneity, and reveals a function for EZH2 in controlling the equilibrium between different tumor cell populations.



## **Materials and Methods**

### **Cell culture and colony formation assays**

HCC70 cells were grown in RPMI media containing 10% FBS, MDA-MB-468 cells in L15- Leibovitz + 10% FBS and SUM149 cells in F-12 containing with 5% FBS, 1ug/ml hydrocortisone and 5 mg/ml insulin. For colony formation assays 500 viable cells were passed through a 40µm strainer, seeded in 96-well plates in wells coated with Matrigel (Growth Factor Reduced, BD), and covered in cell type matching media containing 2% Matrigel. Cultures were grown for 8 days or shorter time points as indicated. Conditions were calibrated to minimize cell aggregation by seeding of a 1:1 mixture of GFP-labeled and non-labeled cells and assessing the percentage of dual color colonies formed; in these conditions 1.6% of colonies were GFP/non-GFP mixed, and we extrapolated that bi-clonal colonies could account to ~3.2% in these conditions. For passaging, Matrigel was dissolved using Cell Recovery Solution (BD); colonies were then dissociated with trypsin, passed through a 40µm strainer and viable cells were reseeded. The semi-liquid assay conditions done as previously described (29) using Growth Factor Reduced Matrigel. Phase contrast images of colonies were obtained using a Nikon Eclipse Ti inverted microscope and a Nikon DS-Fi1 camera. Colony numbers and sizes were quantified using the NIS elements software (Nikon). Colonies with a >30µm diameter and round morphology were scored. Colonies were stained as described in (42) with antibodies against CK14 (Labvision RB-9020) and CK18 (MS-142) followed by labeled secondary antibodies (Jackson). Images were collected using a Zeiss LSM710 laser scanning confocal microscope. >30 colonies or >100 single cells were scored at each time point.

## Lentiviral constructs and infection

The pLKO.1-puro lentiviral vector was used for gene silencing, except for the experiments where EZH2 and GATA3 were co-silenced, in which pLKO.1-neo (Addgene plasmid 13425) was used for EZH2 silencing. Targeting sequences used: shEZH2 #1: ATTCTTGGTTTAAGATTTC CG, shEZH2 #2: AAGCTAAGGCAGCTGTTTCAG, shEED: AATTCCTATGTATGTCTCTGG, shGATA3 TTTCGGTTTCTGGTCTGGATG, shCont (scrambled): CCTAAGGTTAAGTCGCCCTCG. For EZH2 overexpression we used the Fip-EZH2 construct, provided by Marius Wernig. For GATA3 overexpression we cloned the GATA3 cDNA into the pLV-neo lentivirus. To construct the K14p-GFP reporter lentivirus, a SalI fragment containing the CK14 promoter provided by Sabine Werner was cloned upstream to GFP modifying the pLU-Jarid1Bp-hPGKp-Blast<sup>R</sup> vector provided by Meenhard Herlyn. As a control virus we used pRRL-GFP, which expresses GFP under the PGK promoter (43). We used standard virus generation and infection procedures, packaging with the pHRΔ8.2 and pCMV-VSV-G vectors.

## Expression profiling and analysis

RNA was extracted using the RNeasy Mini Plus kit (Qiagen). Gene expression profiling was done on Affymetrix Human Gene 1.0 ST Arrays. Each cell line was profiled in duplicate. Profiling data appear in GEO as GSE36939. Analysis of gene set enrichments was done using the hypergeometric distribution test as in Ref. (6) with a threshold of 0.3 in Log<sub>2</sub> change ( $Z$ -score >1.4) and FDR<0.05. Gene sets used in the study are listed in **supplementary File 1**. Subtype specific and luminal progenitor-associated gene sets were derived from Refs (7, 23). Gene Set Enrichment Analysis (GSEA) was done as described in Ref. (44).

## **Immunohistology, immunocytology and FACS**

Immunofluorescent staining of cultured cells for CK14 and CK18 was done using standard procedures. For assessment of cytokeratin expression in human tumors, sections from 10 basal-like cancers were co-stained for CK14 and CK18 using standard procedures. Cell percentages were calculated as averages of counts of ten confocal microscopic fields examined in each section. Bi-lineage cells in xenograft tumors were scored in the same method, assisted by the CellProfiler software. FACS analysis of live cells was done using standard procedures on an LSRII Analyzer (BD), using antibodies against CD133 (Miltenyi AC133) or c-KIT (Cell Signaling #3308). For intracellular cytokeratin FACS stains, cells were fixed in 2% paraformaldehyde for 10 minutes, permeabilized for 20 minutes in 100% methanol on ice, and stained with CK14/CK18 antibodies followed by conjugated secondary antibodies. Gating for bi-lineage cells was done relying on staining of mono-lineage cell lines (MCF7, MCF10A), the MDA-MB-231 line as a negative control, and single antibody stains. For isolation of bi-lineage and CK18-only cells, we sorted stained cells on FACS-ARIAIII (BD Biosciences). RNA was isolated using the RNeasy FFPE kit (Qiagen).

## **Tumor xenografts**

For tumor xenografts  $2.5 \times 10^6$  viable GFP-labeled MDA-MB-468 cells were injected in 20ul media containing 25% Matrigel into both #4 mammary glands of 6 wk old female NOD/SCID mice. Tumors were measured by palpation, and weighed upon excision, which was done either 125 days after implantation (Figure 4), or at the time points indicated in Figure 6. Tumors were formalin fixed and paraffin embedded for immunohistochemistry. The results shown are a representative experiment of three independent repeats. All experiments involving animals were performed under the approval of the Hebrew University Ethics Committee for Animal Use.

## **Human Patient Material**

Samples of 10 basal-like breast cancers were obtained from patients treated at Hadassah Medical Center. Basal-like tumors were defined on the basis of triple-negative status (HER2-, ER- and PR-negative) and CK5/6-positive stain. Samples were used under the approval of the Institutional Review Board.

## **Western blots and qRT-PCR**

Western blot and qRT-PCR analyses were performed according to standard procedures. We used the aforementioned antibodies as well as anti EZH2 (Upstate 07-689), CK5 (Abcam ab24647) and  $\beta$ -actin (Santa-Cruz sc-1615). Primer sequences appear in Supplementary Methods.

## **Acknowledgments**

We thank Eli Pikarsky, Yuval Dor and Yehudit Bergman for critical reviewing of the manuscript, Marius Wernig for the Fip-EZH2 construct, Alex Roesch and Meenhard Herlyn for the pLU-JARID1Bprom-EGFP-Blast construct, and Sabine Werner for the CK14 promoter plasmid. We thank Norma E. Kidess-Bassir for histological support. This study was supported by the Israel Science Foundation (Grant 1560/07), the Israel Cancer Association, the Israel Cancer Research Foundation, and the Joint Research Fund IMRIC-Hadassah.

The authors declare that they have no potential conflict of interest.

## References

1. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev* 2009; **23**:2563-2577.
2. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* 2011; **5**:5-23.
3. Rakha EA, El-Sayed ME, Reis-Filho J, Ellis IO. Patho-biological aspects of basal-like breast cancer. *Breast Cancer Res Treat* 2008; **113**:411-422.
4. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006; **19**:264-271.
5. Gusterson B. Do 'basal-like' breast cancers really exist? *Nat Rev Cancer* 2009; **9**:128-134.
6. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* 2008; **40**:499-507.
7. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* 2009; **15**:907-913.
8. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 2010; **7**:403-417.
9. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. Defining the cellular precursors to human breast cancer. *Proc Natl Acad Sci U S A* 2011; **109**:2772-2777.
10. Sparmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* 2006; **6**:846-856.

11. Sauvageau M, Sauvageau G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* 2010; **7**:299-313.
12. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002; **419**:624-629.
13. Klee CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003; **100**:11606-11611.
14. Hock H. A complex Polycomb issue: the two faces of EZH2 in cancer. *Genes Dev* 2012; **26**:751-755.
15. Pietersen AM, Horlings HM, Hauptmann M, Langerod A, Ajouaou A, Cornelissen-Steijger P, et al. EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer. *Breast Cancer Res* 2008; **10**:R109.
16. Alford SH, Toy K, Merajver SD, Klee CG. Increased risk for distant metastasis in patients with familial early-stage breast cancer and high EZH2 expression. *Breast Cancer Res Treat* 2011;
17. Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 2003; **22**:5323-5335.
18. Gonzalez ME, Li X, Toy K, DuPrie M, Ventura AC, Banerjee M, et al. Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. *Oncogene* 2009; **28**:843-853.
19. Richter GH, Plehm S, Fasan A, Rossler S, Unland R, Bennani-Baiti IM, et al. EZH2 is a mediator of EWS/FLI1 driven tumor growth and metastasis blocking endothelial and neuro-ectodermal differentiation. *Proc Natl Acad Sci U S A* 2009; **106**:5324-5329.

20. Yu J, Cao Q, Mehra R, Laxman B, Tomlins SA, Creighton CJ, et al. Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. *Cancer Cell* 2007; **12**:419-431.
21. Suva ML, Riggi N, Janiszewska M, Radovanovic I, Provero P, Stehle JC, et al. EZH2 is essential for glioblastoma cancer stem cell maintenance. *Cancer Res* 2009; **69**:9211-9218.
22. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 2006; **125**:301-313.
23. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 2010; **12**:R68.
24. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* 2003; **17**:1253-1270.
25. Pece S, Tosoni D, Confalonieri S, Mazzarol G, Vecchi M, Ronzoni S, et al. Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* 2010; **140**:62-73.
26. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol* 2007; **1**:84-96.
27. Debnath J, Brugge JS. Modelling glandular epithelial cancers in three-dimensional cultures. *Nat Rev Cancer* 2005; **5**:675-688.

28. Bouras T, Pal B, Vaillant F, Harburg G, Asselin-Labat ML, Oakes SR, et al. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. *Cell Stem Cell* 2008; **3**:429-441.
29. Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, et al. Slug and sox9 cooperatively determine the mammary stem cell state. *Cell* 2012; **148**:1015-1028.
30. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, et al. Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 2011; **479**:189-193.
31. Spike BT, Engle DD, Lin JC, Cheung SK, La J, Wahl GM. A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* 2012; **10**:183-197.
32. Zhang M, Behbod F, Atkinson RL, Landis MD, Kittrell F, Edwards D, et al. Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. *Cancer Res* 2008; **68**:4674-4682.
33. Proia TA, Keller PJ, Gupta PB, Klebba I, Jones AD, Sedic M, et al. Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* 2011; **8**:149-163.
34. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 2006; **127**:1041-1055.
35. Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* 2007; **9**:201-209.
36. Kouros-Mehr H, Bechis SK, Slorach EM, Littlepage LE, Egeblad M, Ewald AJ, et al. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell* 2008; **13**:141-152.



37. Asselin-Labat ML, Sutherland KD, Vaillant F, Gyorki DE, Wu D, Holroyd S, et al. Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor caspase-14. *Mol Cell Biol* 2011; **31**:4609-4622.
38. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004; **23**:7669-7678.
39. Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D, et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 2005; **65**:11259-11264.
40. Gusterson BA, Ross DT, Heath VJ, Stein T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res* 2005; **7**:143-148.
41. Chang CJ, Yang JY, Xia W, Chen CT, Xie X, Chao CH, et al. EZH2 Promotes Expansion of Breast Tumor Initiating Cells through Activation of RAF1-beta-Catenin Signaling. *Cancer Cell* 2011; **19**:86-100.
42. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* 2003; **30**:256-268.
43. Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, Naldini L, et al. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol* 1998; **72**:9873-9880.
44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**:15545-15550.

## Figure Legends

### Figure 1. EZH2 promotes the expression of basal-lineage and progenitor-associated genes. (a)

Western blot analysis of basal (CK5, 14) and luminal (CK18) cytokeratins in basal-like breast cancer cell lines expressing a shEZH2 or control (shCont) lentivirus. (b) Same analysis in basal-like HCC70 or luminal MCF7 cells overexpressing EZH2 (OE) or infected with control vector. (c) Fold change in expression of basal- and luminal-lineage markers in EZH2 or EED silenced HCC70 cells, and in EZH2 over-expressing cells, relative to control cells, assessed by qRT-PCR. Values indicate average of triplicate reactions  $\pm$ SEM. (d) Preferential repression (green) or activation (red) of genes highly expressed in basal-like cancers in the indicated cells relative to controls. Values indicate statistical significance of preferential change in gene set expression, shown as  $-\log_{10}$  of *P* value. (e) Gene set enrichment analysis (GSEA) of genes highly expressed in basal-like cancers in shEZH2 versus shCont infected HCC70 cells. ES – enrichment score, NES – normalized ES. (f) Fold change in expression of individual genes from the basal-like gene set and the luminal progenitor gene set in indicated cells relative to control cells, as in (c). VGLL1 and CD133 are included in both sets. (g) Preferential repression (green) or activation (red) of a set of luminal progenitor-associated genes in the indicated cells relative to controls. (h) GSEA of luminal progenitor-associated genes in shEZH2 versus shCont infected HCC70 cells. (i,j) FACS analysis of control and EZH2-silenced HCC70 cells stained for CD133 or c-KIT. Percentages of high CD133 or c-KIT expressing cells are indicated.

**Figure 2. A bi-lineage cell subpopulation is found in basal-like breast tumors. (a)** Co-staining of CK14 (green) and CK18 (red) on tumor sections of human basal-like breast cancers. Double-positive (bi-lineage) cells, appear as yellow/orange. (b) Percentage of CK14+ only, CK18+ only, and bi-lineage CK14+CK18+ cells in tumor sections of ten basal-like breast tumors obtained from different patients.

(c) Immunofluorescent co-staining of CK14 (red) and CK18 (green) in luminal (MCF7), non-tumorigenic myoepithelial (basal) (MCF10A) and basal-like cell lines (MDA-MB-468, HCC70). (d) FACS analysis of CK14 and CK18 expression in indicated cell lines. Gate indicates bi-lineage subpopulation.

**Figure 3. EZH2 controls the bi-lineage cell fraction in basal-like cell lines and xenograft tumors.**

(a) FACS analysis of CK14 and CK18 expression in EZH2-silenced and control HCC70 cells. (b) Same analysis in EZH2-overexpressing and control cells. (c) Fold change of the bi-lineage cell fraction size relative to matching control cells (Cont) in the indicated cells, as determined by FACS analysis. Values indicate average of three or more independent experiments  $\pm$ SEM. \*  $P < 0.03$ , \*\*  $P < 0.003$ , *t*-test. (d) Growth curves (left) and final masses (right) of tumors formed by control or EZH2-silenced MDA-MB-468 cells injected into the mammary glands of NOD-SCID mice. Values indicate average across tumors  $\pm$ SEM. (e) The percentage of bi-lineage cells in xenograft tumors formed by control and EZH2-silenced MDA-MB-468 cells, as assessed by immunofluorescent staining, as in (f). Circles represent individual tumors. Arrows indicate tumors in which no EZH2 silencing was detected. Black bars indicate average across tumors. (f) Representative images of xenograft tumor sections stained for EZH2 (white), CK14 (green) and CK18 (red); bi-lineage cells appear as yellow or orange in the merged images.

**Figure 4. EZH2 promotes colony formation capacity.** (a) Confocal image of a colony formed on Matrigel by HCC70 cells and stained for CK14 and CK18 as indicated. (b) Number of colonies formed on Matrigel by control and EZH2-silenced HCC70 upon initial seeding (passage 0) and upon dissociation of colonies and replating (passage 1). Average colony sizes are shown on right (sizes were

the same after passage). Representative phase-contrast images of colonies are shown on right. **(c)** Colony numbers and sizes of same cells in semi-liquid stem cell culture conditions. **(d)** Colony numbers and sizes of control or EZH2-silenced SUM149 cells; colonies had morphology and composition similar to those formed by HCC70 cells. **(e)** Colony numbers and sizes of control and EZH2 over-expressing HCC70 cells. Values indicate average of 3-5 replicates  $\pm$ SEM.

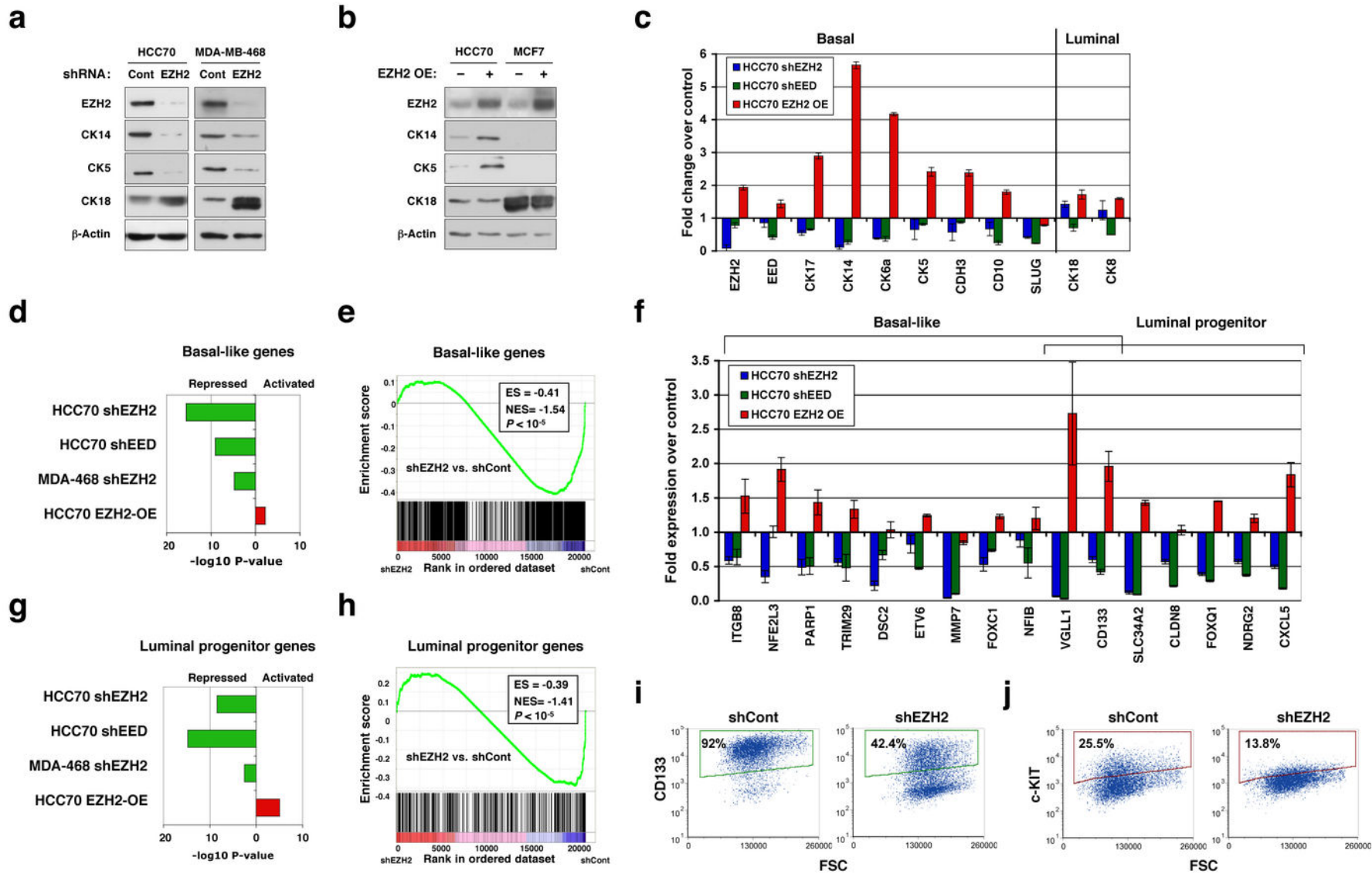
**Figure 5. Bi-lineage cells possess increased colony formation capacity and express progenitor-associated genes.** **(a)** Confocal images of single HCC70 cells or colonies stained for CK14 (green) and CK18 (red), 1,4 or 8 days after seeding on Matrigel. **(b)** Percentage of bi-lineage cells among single HCC70 cells 1,4 or 8 days following seeding (yellow bars), and percentage of colonies containing bi-lineage cells 4 or 8 days after seeding (blue bars). No colonies were present 1 day after seeding. \*  $P < 0.007$ , Pearson's chi-squared test. **(c)** Numbers and sizes of colonies formed by GFP-high or GFP-negative HCC70 cells infected with the K14p-GFP reporter vector (shown in diagram). Values indicate average of 5 replicates  $\pm$ SEM. Representative images of colonies are shown on right, inset shows GFP expression in colonies formed by GFP-high cells. **(d)** Numbers and sizes of colonies formed by GFP-positive and GFP-negative HCC70 cells infected with the control PGKp-GFP vector. N.S. – non-significant. **(e)** qRT-PCR analysis of basal, luminal and luminal progenitor-associated genes performed on RNA extracted from FACS-isolated bi-lineage and CK18-only HCC70 cells. Values indicate fold expression in bi-lineage cells over CK18-only cells, as average of triplicate reactions  $\pm$ SEM.

**Figure 6. Bi-lineage cell numbers increase during tumor formation.** **(a)** Percentage of bi-lineage cells in sections of mouse mammary glands implanted with MDA-MB-468 cells and excised 2 hours (2hr), 1 day, 1 week (1wk) or 3 weeks (3wk) after implantation. Each circle represents a single tumor.

Black bars indicate average of 4 tumors. \*  $P < 0.01$ , \*\*  $P < 0.001$ ,  $t$ -test. **(b)** Representative confocal images of sections of tumors excised at each time point and stained for CK18 (red) and CK14 (green). Note that CK18 expression was detected in all cells, albeit at low levels in cells that appear green in the images.

**Figure 7. GATA3 represses bi-lineage identity.** **(a)** Fold change in GATA3 and FOXA1 expression in EZH2-silenced HCC70 cells relative to control cells (shCont), assessed by qRT-PCR. Values indicate average of triplicate reactions  $\pm$ SEM. **(b)** Fold change in expression of basal-lineage, luminal-lineage, and luminal progenitor-associated (lum prog) genes upon GATA3 overexpression in HCC70 cells, assessed by qRT-PCR. **(c)** Images of control (pLV) HCC70 cells or cells overexpressing GATA3 (pLV-GATA3), either without (shCont) or with (shEZH2) concomitant EZH2 silencing, stained for CK18 (green) or CK14 (red). Histogram shows bi-lineage cell numbers relative to control cells, as quantified from images. Values indicate average of 7 fields each containing  $>50$  cells,  $\pm$ SEM. \*  $P < 0.01$ , \*\*  $P < 0.002$ ,  $t$ -test **(d)** Fold change in expression of basal-lineage, luminal-lineage, and luminal progenitor-associated genes upon GATA3 silencing in HCC70 cells, assessed by qRT-PCR. **(e)** CK14 and CK18 FACS analysis of HCC70 cells silenced for GATA3. **(f)** CK14 and CK18 FACS analysis of HCC70 cells silenced for EZH2, or for both EZH2 and GATA3. **(g)** Schematic model: EZH2 and GATA3 play opposing roles in determining bi-lineage content in basal-like breast cancers.

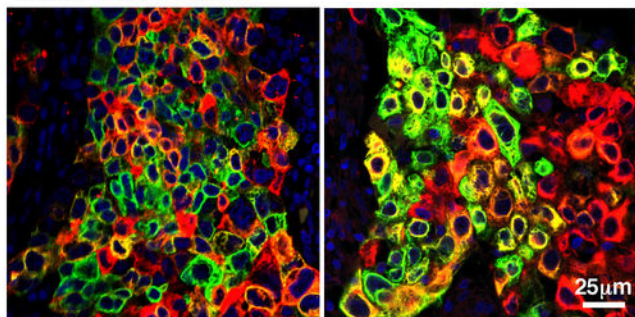
# Figure 1



# Figure 2

**a**

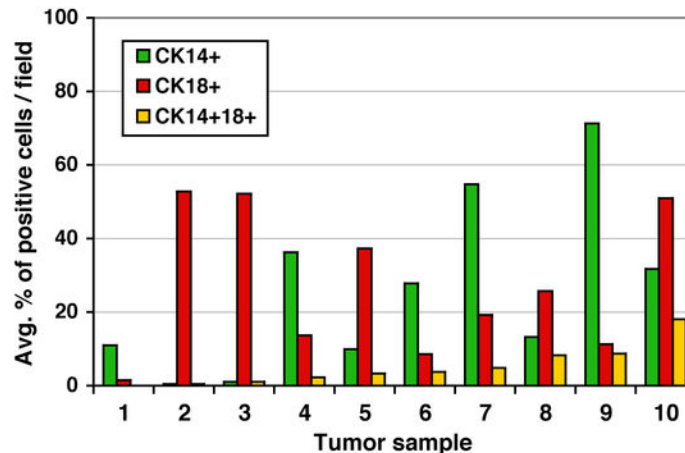
CK18 / CK14 / DNA



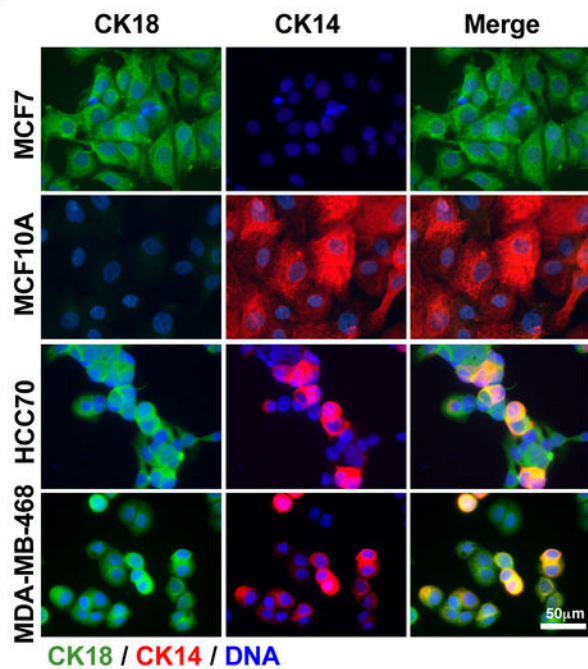
Sample #9

Sample #5

**b**

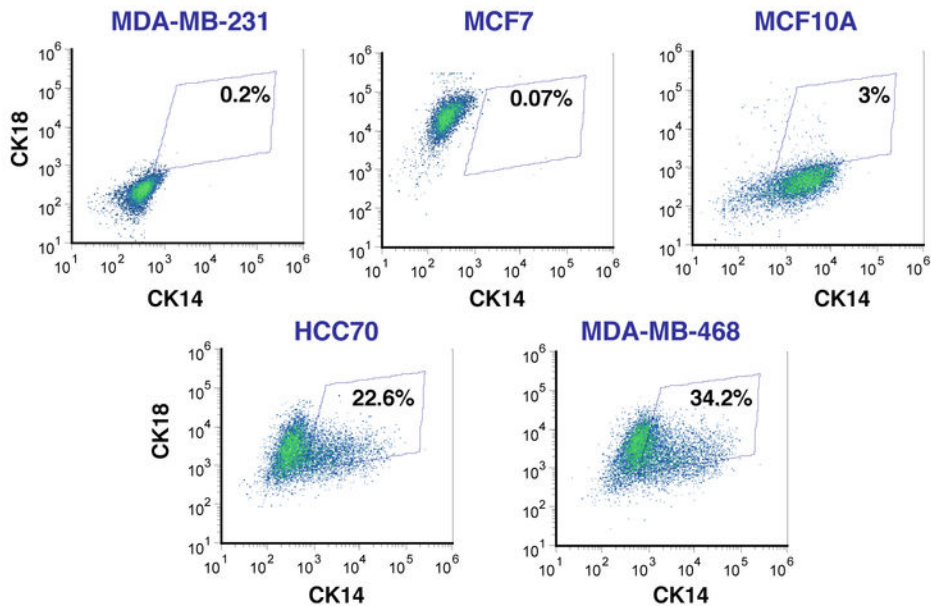


**c**

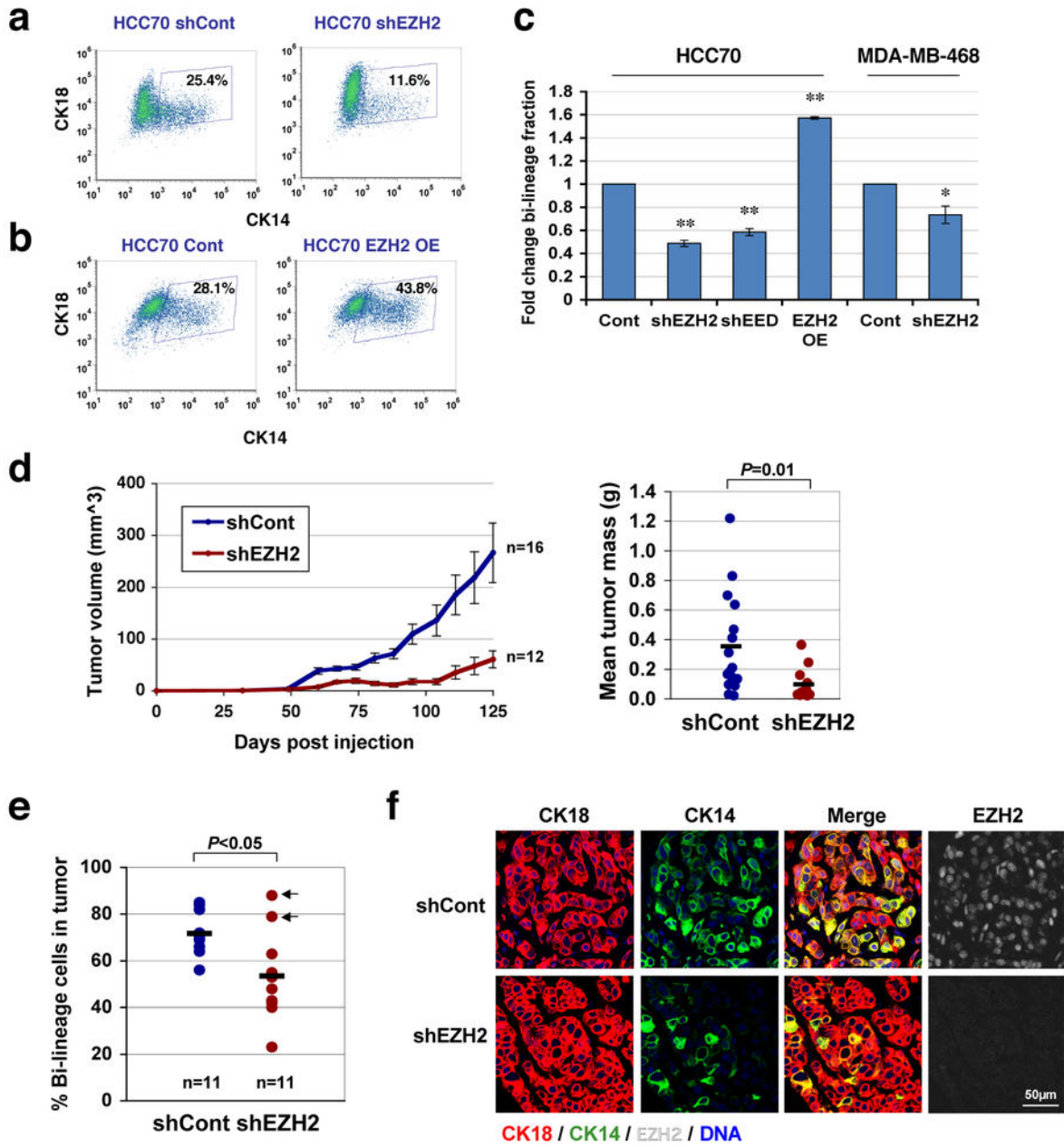


CK18 / CK14 / DNA

**d**



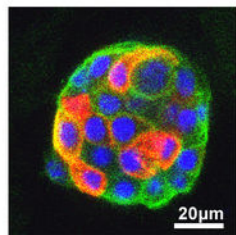
**Figure 3**



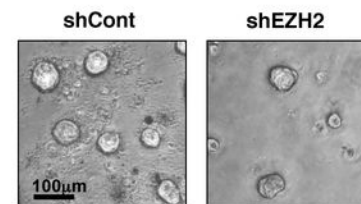
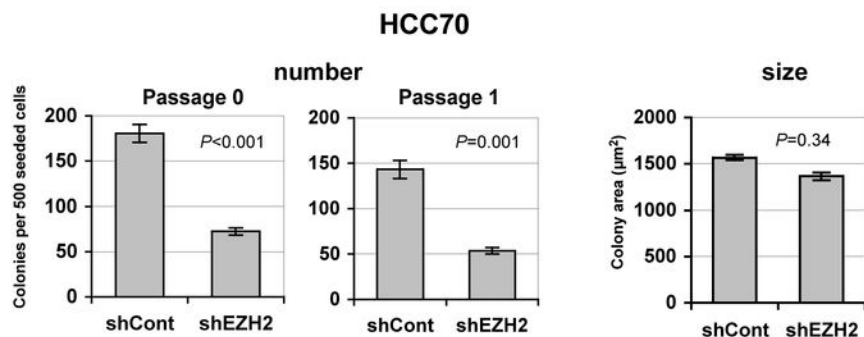


# Figure 4

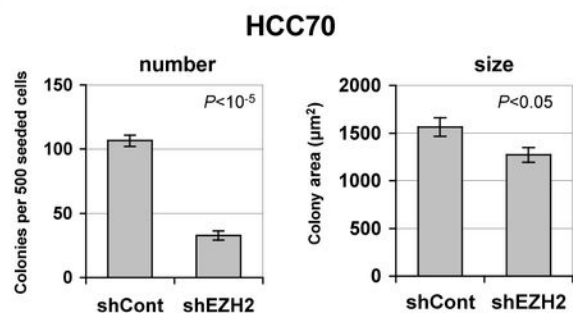
**a**



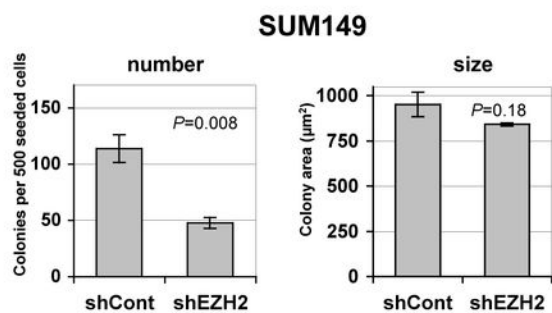
**b**



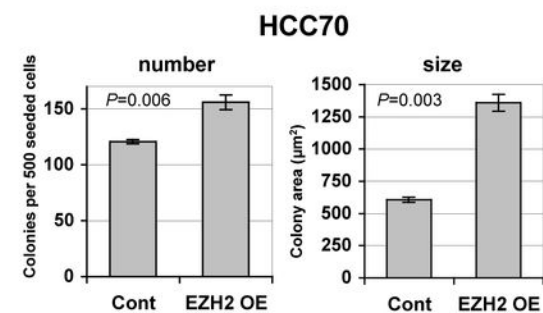
**c**



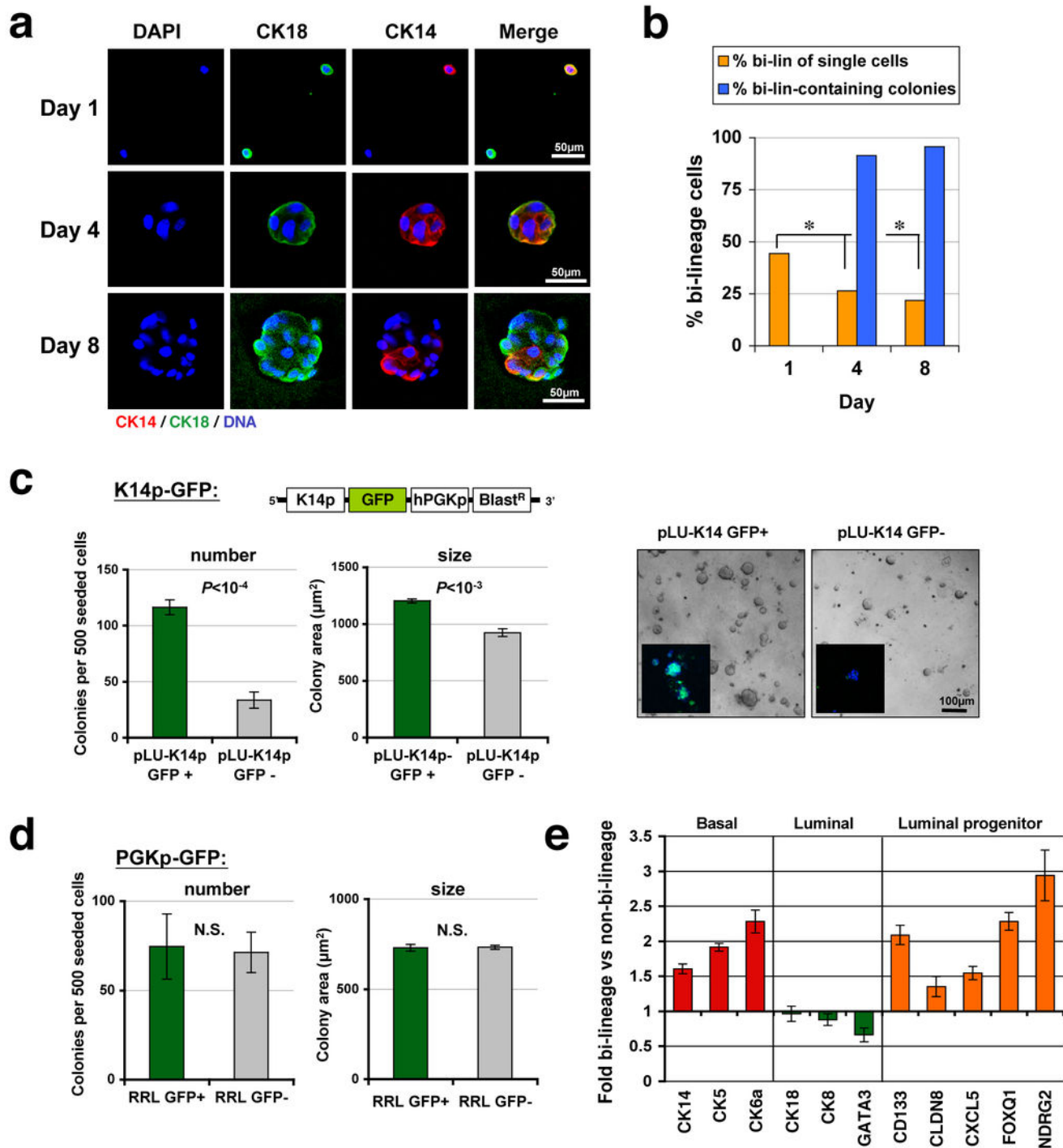
**d**



**e**

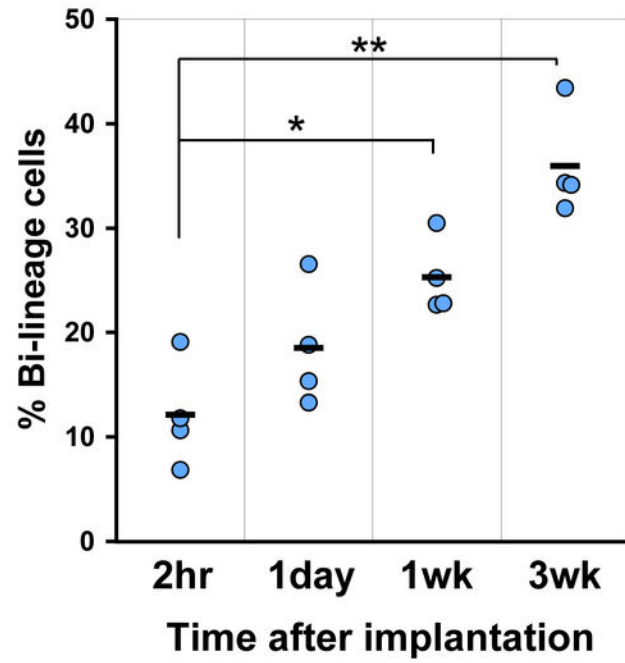


# Figure 5

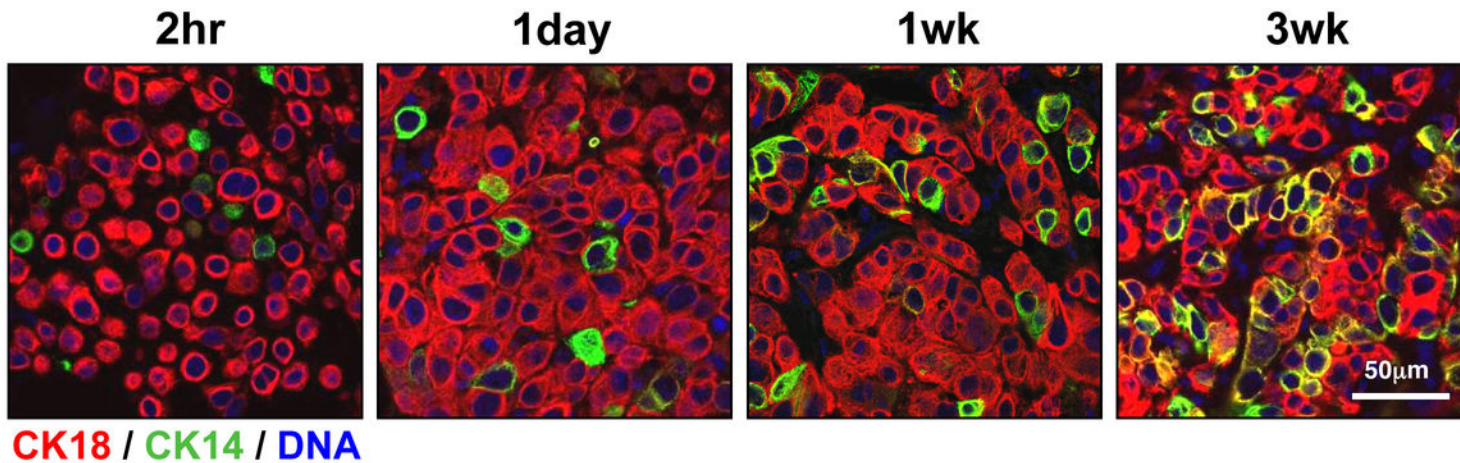


# Figure 6

**a**

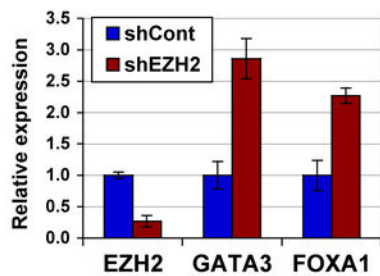


**b**

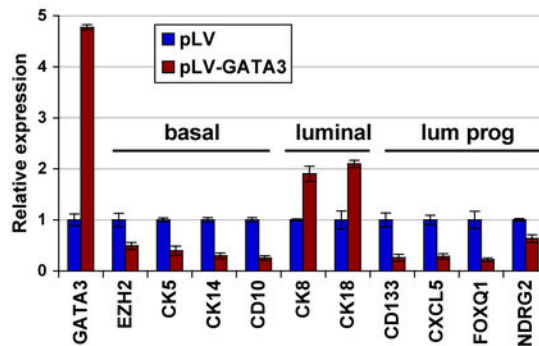


# Figure 7

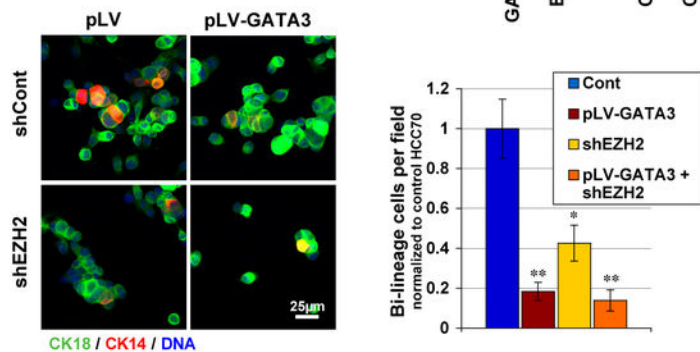
**a**



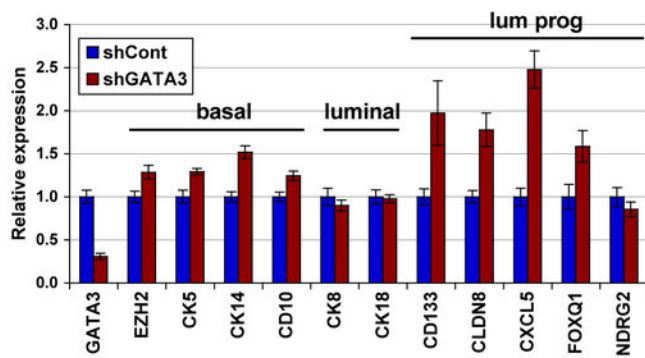
**b**



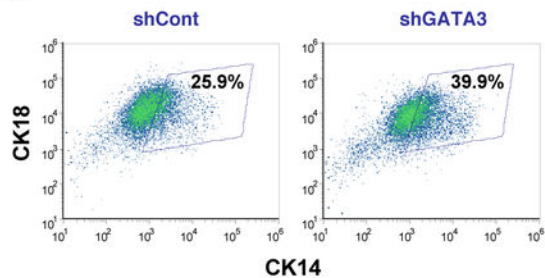
**c**



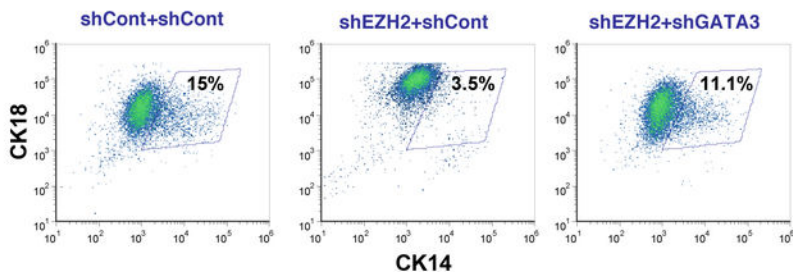
**d**



**e**



**f**



**g**

