# TRANSCRIPTIONAL REGULATION OF BREAST CANCER DIFFERENTIATION STATES

Thesis submitted for the degree of

"Doctor of Philosophy"

By

Roy Zvi Granit

Submitted to the Senate of the Hebrew University of Jerusalem

December 2016

# TRANSCRIPTIONAL REGULATION OF BREAST CANCER DIFFERENTIATION STATES

Thesis submitted for the degree of

"Doctor of Philosophy"

By

Roy Zvi Granit

Submitted to the Senate of the Hebrew University of Jerusalem

December 2016

# בקרה שעתוקית של מצבי התמיינות בסרטני שד

חיבור לשם קבלת תואר דוקטור לפילוסופיה

מאת

רועי צבי גרניט

הוגש לסנט האוניברסיטה העברית בירושלים

דצמבר 2016

# בקרה שעתוקית של מצבי התמיינות בסרטני שד

חיבור לשם קבלת תואר דוקטור לפילוסופיה

מאת

רועי צבי גרניט

הוגש לסנט האוניברסיטה העברית בירושלים

דצמבר 2016

עבודה זו נעשתה בהדרכתו של דר' איתי בן-פורת

# תקציר


מחלת סרטן השד מתאפיינת בהטרוגניות רבה, הנצפית בין גידולים שונים ובין התאים המרכיבים את הגידול. להטרוגניות זו השלכות מעשיות על אופן הטיפול בחולה והיא פוגעת ביכולת למגר את כלל התאים הסרטניים. ההבנה של ההרכב התאי של גידולים, התכונות הביולוגיות של תתי-אוכלוסיות בתוך הגידול והמנגנוניים המבקרים יצירת הטרוגניות לוקה בחסר.

במחקריי כיוונתי לגלות את הגנים המבקרים מצבי התמיינות בסרטני שד, ולהבין את היחסים בין מצבים אלו ובין סוגי התאים הקיימים ברקמת השד הבריאה והשפעתם על תכונות סרטניות. בעבודתי התמקדתי בסרטני שד מתת הסוג ה"בזאלי", המאופיינים במצב בלתי ממוין, התנהגות קלינית אגרסיבית ומידה רבה של הטרוגניות. באמצעות אנליזה של דגמי ביטוי גנים בגידולי שד, זיהיתי קבוצה של פקטורי שעתוק וכרומטין המבוטאים באופן ייחודי בסרטני שד בזאליים, המהווים מועמדים לבקרת התמיינות בסרטני שד. מבין אלו חקרתי תחילה את הגן EZH2 ממשפחת Polycomb ומצאתי כי הוא חיוני לשמירת זהותם של תאים מגידולים בזאליים. בנוסף, מצאתי כי EZH2 מבקר את הרכבם של אותם גידולים, תוך שהוא מקדם את הימצאותה של אוכלוסיית תאים ייחודית בעלת תכונות 'פרוגניטוריות' ויכולות סרטניות מוגברות. כדי לזהות גנים נוספים המבקרים את הרכב הגידול ביצעתי סריקה פונקציונאלית מקיפה ובחנתי את תפקידם של כלל הגנים המועמדים שזוהו. סריקה זו הובילה לזיהויים של קבוצה של גנים בעלי השפעה על מצב ההתמיינות: חלקם תרמו להגדלה של האוכלוסייה הפרוגניטורית, בעוד אחרים את ביצעו את התפקיד ההפוך. בין גנים אלו נכללו כאלו המבקרים גם את זהות התאים ברקמת השד הבריאה.

עוד מצאתי כי מנגנון מרכזי השולט על המעבר בין זהות פרוגניטורית למצב ממוין יותר הוא חלוקות אסימטריות המאפשרות יצירתם של שני תאי בת בעלי זהות שונה. זיהיתי כי פקטורי השעתוק EZH2 ו- NFIB פועלים באמצעות מסלול Notch ומגדילים את חלקה של האוכלוסייה הפרוגניטורית על ידי דיכוי חלוקות אסימטריות, בעוד הפקטורים GATA3 ו-FOXA1 פועלים בכיוון ההפוך, ומעודדים חלוקות אסיטמריות המייצרות תאים ממוינים. בנוסף, כאשר חקרתי את השפעתם של שינויים אפיגנטיים על זהותם של גידולי שד מצאתי כי יוביקוויטינציה של היסטון H2 מתרחשת ברמה שונה בתתי סוגים של סרטני שד, ומשפיעה בהם בצורה מנוגדת על ביטוי גנים ותכונות סרטניות.

עבודתי חושפת מנגנונים חדשים המבקרים את זהותם ומידת ההטרוגניות של סרטני שד, מצביעה על התפקוד השונה של תתי אוכלוסיות תאים בגידול, ומלמדת כי הבקרה על חלוקות אסימטריות הינה מנגנון ליצירת הטרוגניות בגידול. עבודתי מסייעת להבנה הבסיסית של הבקרה על הרכב הגידול ומצביעה על מסלולים הניתנים לעיכוב תרופתי כאמצעי להשפעה על מצב התמיינותם של תאי הגידול.

This work was carried out under the supervision of:

Dr. Ittai Ben-Porath

# Acknowledgments

I am thankful to have been given the opportunity to work under the supervision of Dr. Ittai Ben-Porath. Thank you for believing in me, being open to ideas, allowing me to conduct innovative research and pushing me to the edge when needed. Few are the people that pursue science with such passion and expertise. I would also like to thank all of the lab members, past and present, that have accompanied me on this journey and provided professional and personal support. I have met a truly amazing group of people, highly dedicated and genuinely committed to mutual help. Thanks to the many collaborators, particularly in the labs of Prof. Moshe Oren, Prof. Aviv Regev and Dr. Dave Root who have made these studies possible.

Finally, I owe many thanks to my wife Neta and my family, for their endurance and the support they have provided me along this path.

Dedicated to my grandfather Prof. Dov Nir which has seeded in me the love of knowledge and curiosity.

# Abstract

Breast cancer is a disease displaying a high level of heterogeneity, which is observed between tumors as well as between the individual cells that comprise them. This heterogeneity affects treatment choices for patients, and hinders the ability to effectively target the entire malignant cell population. A detailed understanding of the composition of tumors, the biological traits of distinct cell subtypes present in them, and the mechanisms that govern the generation of heterogeneity, is lacking.

In my research I aimed to uncover regulators that determine the differentiation states of breast cancer cells, and understand the relation of these states to the cell lineages in the normal mammary epithelium, as well as their influence on cellular tumorigenic traits. I focused on breast tumors of the "basal-like" subtype, which exhibit poor differentiation, aggressive clinical behavior, and a high level of heterogeneity. I identified, through analysis of tumor expression profiles, a group of transcriptional and chromatin regulators specifically expressed in basal-like tumors, representing candidate regulators of differentiation in breast cancer. I studied among these the Polycomb factor EZH2, and found that it is necessary for maintenance of basal-like subtype identity. Furthermore, I found that EZH2 regulates the composition of the tumor cell population, by promoting the expansion of subset of cells that display a progenitor-like differentiation state and enhanced tumorigenic traits. To identify additional regulators of tumor cell composition, I conducted a functional screen testing the effects of many of these candidates. This led me to identify a group of regulators, some of which increased the numbers of progenitor-like breast cancer cells, and others acting to reduce it. Importantly, these included known regulators of normal mammary progenitor identity.

I found that a central mechanism controlling the transitions of breast cancer cells between a progenitor-like and a more differentiated state is asymmetric divisions, which allow the production of two distinct daughter cells. EZH2, and the NFIB transcription factor, acting through the Notch pathway, increase the numbers of progenitor-like tumor cells by repressing asymmetric divisions, while GATA3 and FOXA1 increase the rates of asymmetric divisions, thereby producing more differentiated cells. In studying the effects of chromatin modifications on breast cancer identity and function, I found that regulation of histone H2 ubiquitinylation differs between breast cancer subtypes and influences tumor

gene expression and behavior in a subtype-dependent manner.

Altogether, my work uncovers novel pathways that govern subtype identity and intratumoral heterogeneity in breast cancer, highlights the functional significance of distinct tumor cell subpopulations, and reveals the role played by regulation of asymmetric divisions in generating heterogeneity. This work thus provides basic insights into how tumor composition is controlled, and suggests that tumor cell differentiation state could be modified by drugs that target these regulatory pathways.

# Letter of Contribution

I have conducted all the experiments and analyzed the resulting experimental data described in this Ph.D. thesis. I was also responsible for obtaining, or generating the necessary materials to conduct my research. I initiated and planned the research and the experiments that were done. I prepared the figures and participated in writing the papers. These studies were conducted under the supervision and the advice of Dr. Ittai Ben-Porath. Additional authors that have assisted with specific parts of the research are indicated below:

## Role of the additional authors:

## Chapter 1

Y. Gabai: assisted in the conduction of various experiments

T. Hadar: stained and characterized breast cancer tissue samples

Y. Karamansha: assisted with cloning and western blots

L. Liberman: assisted with cloning of the K14p reporter

I. Waldhorn: helped with EZH2 over-expression experiments

I. Gat-Viks: conducted bioinformatics analysis

A. Regev: advised on bioinformatics analysis

B. Maly: analyzed tissue micro-arrays and provided pathologic advice

M. Darash-Yahana: obtained tissue sections

T. Peretz: consulted and provided tissue samples

I. Ben-Porath: supervised the entire study and wrote the manuscript

## Chapter 2

E. Carmon: stained and quantified breast cancer section panel

Y. Fixler: assisted in the conduction of various experiments

S. Dalin: helped in the conduction of BrdU pulse-chase experiments

R. Moran: assisted in the conduction of various experiments

T. Sella: consulted and provided tissue samples

A. Sonnenblick: obtained PDX tumor samples

D. Katz: assisted in the conduction of *in-vivo* experiments

U. Lehmann: provided pathologic advice

K. Paz: obtained PDX tumor samples

F. Piccioni: assisted in the conduction of shRNA screen

A Regev: advised on bioinformatics analysis

D.E. Root: assisted in the conduction of shRNA screen

I. Ben-Porath: supervised the entire study and wrote the manuscript

## Chapter 3

This chapter was conducted in an equal collaboration with Ohad Tracic, from the lab of Moshe Oren (Weizmann institute), each contributing a substantial and unique contribution. I have contributed from my experience and knowhow in breast cancer research and bioinformatics, conducting both experiments and analysis, while Ohad contributed his experience with the H2Bub1 modification. Both principle investigators supervised the study.

O. Tarcic: co-authored the manuscript and conducted experiments

I.S. Pateras: stained and quantified breast cancer section panel

H. Masury: assisted with tissue culture and cell migration assay

Bella Maly: provided pathologic advice and analysis

V.G. Gorgoulis: provided pathologic advice and analysis

E. Pikarsky: provided pathologic advice and reviewed the MS

I. Ben-Porath: supervised the study and wrote the manuscript

M. Oren: supervised the study and wrote the manuscript

## Chapter 4

M. Slyper: assisted in writing the review and preparing the figures

I. Ben-Porath: supervised the entire study and wrote the manuscript

**Approved by:**

_____

Dr. Ittai Ben-Porath

# Table of Contents

# 1. Introduction

## 1.1 Breast cancer heterogeneity

### 1.1.1 Tumor heterogeneity

Tumors are heterogeneous in numerous aspects. Individual tumors differ in their gene mutation profiles, epigenetic modifications, gene-expression patterns and cell traits and composition. These differences and others give rise to heterogeneity in tumor growth, progression and responses to therapy (Marusyk et al., 2012). Heterogeneity can be seen in two levels: intertumoral heterogeneity, observed between tumors in individual patients with distinct characteristics, and intratumoral heterogeneity, observed amongst individual cells or cell populations within a single tumor (Figure 1).

Intertumoral heterogeneity can be generated due to initiation of tumors from distinct cells of origin, the accumulation of different mutations along tumor formation, environmental cues such as hypoxia, changes in vasculature, as well as many additional factors (Marusyk et al., 2012). Similar factors can also produce intra-tumor heterogeneity: clonal mutational evolution can give rise to genetic diversity within the tumor (Eirew et al., 2015), while environmental factors can act locally, for instance, some cells can reside in a more hypoxic region of the tumor and differently active genes relative to their counterparts, found in oxygen rich regions (Allen et al., 2016; Wagenblast et al., 2015). Importantly, differentiation events within tumors can generate diverse cellular subtypes within the tumors, and these may execute distinct roles within them (Almendro et al., 2014; Marusyk et al., 2012).

Intratumoral heterogeneity is recognized to have important clinical implications in the course of tumor development and treatment. Cell-specific variation due to genetic factors (e.g. treatment resistance mutations) or epigenetic changes (e.g. expression of resistance genes) (Lackner et al., 2012; Meacham and Morrison, 2013) can hinder the ability to effectively target the entire malignant population. Interestingly, experiments in culture suggest that tumor composition can reach an equilibrium that is kept even after therapy and other bottlenecks, suggesting a guided process (Eirew et al., 2015; Gupta et al., 2011); yet

the factors that govern the composition of tumors and the cellular mechanisms controlling individual cell identity are poorly characterized.

Despite its high clinical significance, the role of epigenetic regulation of this diversity has not been well characterized. My work is focused on investigating tumor heterogeneity and aims to elucidate its functional implications and the molecular mechanisms that govern it. I studied the contribution of differentiation events to generating heterogeneity in breast cancer, and the links between the phenotypes of breast cancer cells and those of cells lineages in the normal breast. I focused on tumors of the basal-like subtype, a highly aggressive form of the disease which demonstrates a high degree of heterogeneity.

I have conducted a broad screen of candidate transcriptional regulators to identify those that influence heterogeneity, which was measured by inspecting the differentiation state of tumor cells. This screen pointed to 15 regulators, amongst them the Polycomb regulator EZH2, members of the Notch pathway, and additional genes that were studied in depth. Along these efforts I have also studied the functional implications of intratumoral heterogeneity on tumor cell traits and the mechanisms that allow shifts in population balance. These efforts led to the discovery of novel role for EZH2 and Notch in the regulation of tumor heterogeneity. I have also studied heterogeneity in the ubiquitylation of histone H2B (H2Bub1), another major epigenetic modification implicated in cancer. In this study I have measured the regulation and function of H2Bub1 in luminal versus basal-like tumors and uncovered subtype-specific roles for this modification.

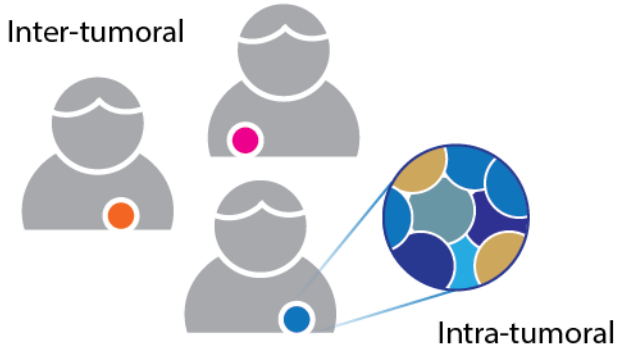



**Figure 1: Types of tumor heterogeneity.** Tumor heterogeneity, represented by different colors, is found at two levels: intertumoral heterogeneity that occurs among tumors and

intratumoral heterogeneity that is found among individual cells that compose the tumor.

### 1.1.2 Breast cancer classification

The realization of the importance of intertumoral heterogeneity has led to considerable efforts to classify the diversity between different breast tumors. Breast cancer is the most frequently diagnosed cancer in women in developed countries, affecting 12.6% of women (one in eight) (Feuer et al., 1993). The most common type of breast cancer is invasive ductal carcinoma, which could be classified into subgroups. Immunohistochemical staining allows clinical classification of invasive ductal carcinomas into three major subtypes based on the expression of the Estrogen Receptor (ER) and Progesterone Receptor (PR), which are typically co-expressed, and the receptor tyrosine kinase HER2 (Blows et al., 2010). The majority of tumors (70-80%) are $ER^+$ and often respond to endocrine therapy. The second subtype includes tumors that overexpress HER2 positive (20-15%), depend on this signaling pathway for their growth, and often respond to anti-HER2 therapies such as Trastuzumab (Herceptin). The remainder (10-15%) are termed triple-negative (TN), since they do not express ER, PR and HER2, and are typically highly aggressive and poorly differentiated, and currently lack targeted treatment options (Cancer Genome Atlas Network, 2012; Polyak and Metzger Filho, 2012; Stingl and Caldas, 2007). Other cases of less common breast cancers, such as lobular carcinoma and others, are normally considered to have relative good prognosis and are mostly responsive to endocrine therapy (Reed et al., 2015).

The emergence of advanced molecular technologies, allowing the profiling of the entire gene expression pattern and mutation spectrum of a given sample, led to the development of more refined classifications of breast ductal carcinomas. Clustering of gene expression profiles of hundreds of breast tumors revealed the existence of at least four major so called 'intrinsic subtypes' (Perou et al., 2000). These include the Luminal A and Luminal B subtypes, which largely encompass the $ER^+$ tumors, with Luminal B being more proliferative; the HER2 over-expressing subtype; and the basal-like subtype, which is largely composed of TN tumors (Perou et al., 2000; Prat and Perou, 2011; Prat et al., 2012). Tumors of the basal-like subtype are clinically aggressive, poorly differentiated and are of particular interest due to their unique biological features (discussed below) and because they currently lack targeted treatment options (Banerjee et al., 2006; Rakha et al., 2009). Additional, more minor, subtypes have also been suggested, such as the claudin-low subtype that is TN and

enriched for a mesenchymal gene expression signature (Prat et al., 2010).

Altogether, it is now accepted that breast carcinoma patients with local disease can be classified in the clinic into four major subtypes (Luminal A, Luminal B, HER2 and TN) that are distinguished by a combination of immunohistochemical and molecular tools. Intrinsic subtype diagnosis has become critical for prognosis and determination of treatment course, including the need for and course of neoadjuvant and adjuvant chemotherapy (before and following surgery) for localized disease or metastatic, and treatment with targeted therapies (anti-estrogenic, anti HER2 and others) (Prat et al., 2015). For example, of ER+ patients, only those that will be classified as Luminal B will receive adjuvant chemotherapy to combat their high proliferative and aggressive phenotype (Goldhirsch et al., 2011).

It is also thought that these tumors represent distinct biological entities, potentially reflecting different cells of origins and/or courses of mutation accumulation (Prat et al., 2012). Importantly, while these subtypes are clearly distinguished at the levels of gene and protein expression, as well as methylome analyses, and common mutations, there is a very high level of heterogeneity between tumors of the same subtype (Cancer Genome Atlas Network, 2012). For example, while PI3K mutations are very common in luminal tumors (45%), there are no additional mutations that are found in more than ~15% of patients of this subtypes. In basal like tumors, 80% of patients carry p53 mutations, yet mutation profiles in other patients are highly diverse and there are no genes that are common, and the combination of driver mutations observed in tumors are highly diverse (Cancer Genome Atlas Network, 2012; Nik-Zainal et al., 2016). Individual tumor diversity is thus high despite these groupings, emphasizing the importance of molecular classification of each tumor and the need for understanding the mechanisms that dictate this heterogeneity and its functional implications, which are some the aims of my study.

### 1.1.3 The normal mammary gland

In order to determine whether particular tumors reflect distinct biological identities or differentiation states, it is necessary to understand the structure and cellular hierarchy of the normal tissue from which the tumors arise. The healthy breast tissue is composed mostly of fat embedded with milk-producing alveoli, which are active during lactation, and are connected by a network of ducts to the nipple (Watson and Khaled, 2008). This network is present in the non-lactating breast and organized in terminal ductal-lobular units, in which

inactive alveoli are connect by small ducts that unite to form larger inter-lobular ducts (Watson and Khaled, 2008). All of these ducts are comprised of two unilayers of epithelial cells, each exhibiting distinct biological properties and distinguished by specific proteins that can be used as markers. Cells in the outer basal layer present a spindle-like myoepithelial morphology and are marked by the expression of alpha-smooth muscle actin (αSMA), that allows cells to contract to assist milk flow. In addition they express basal-epithelium intermediate-filaments - cytokeratins 5,14 (K5, K14), the transcription factor p63, and other structural and regulatory proteins (Gudjonsson et al., 2005; Stingl and Caldas, 2007; Visvader, 2009). Cells in the inner, luminal, layer express luminal-epithelium intermediate-filaments cytokeratins 8,18 (K8, K18) and some specific regulators such as GATA3 and FOXA1 (Stingl and Caldas, 2007; Visvader, 2009). Some of these cells express the hormone receptors ER and PR, and the androgen receptor (AR).

The cell hierarchy controlling to the formation of the normal breast tissue has been extensively studied yet much is still unknown. It is thought that at the top of the hierarchy are mammary stem cells (MaSCs) which are thought to reside within the basal compartment. These cells can give rise to all cells in the tissue, including basal cells, and luminal progenitor cells that in turn can give rise to terminally differentiated luminal cells (Lim et al., 2009; Prater et al., 2014; Rios et al., 2014). Fluorescence-activated cell sorting (FACS) allowed the isolation and profiling of these populations, revealing gene expression signature that are characteristic to each (Lim et al., 2009).

The basal/MaSC cell population could be enriched for by means of FACS sorting of both mouse and human origin mammary tissue, using several markers. Most widely used is the combination of low expression epithelial cell adhesion molecule (EpCAM) and high levels of alpha 6 integrin (CD49f) (Lim et al., 2009). Other approaches employ other sets of markers such as Aldehyde Dehydrogenase 1 (ALDH1), or a CD10 (Ginestier et al., 2007; Raouf et al., 2008).

MaSCs are defined functionally by their ability to grow in anchorage independent environment and give rise to cells of both basal and luminal lineages. The ultimate demonstration of their function is that basal/MaSCs cells isolated both from mouse and human tissue were found to generate a full reconstituted epithelial tree when transplanted *in vivo* (Shackleton et al., 2006). To date MaSCs could not be distinguished from the rest of

basal cells based on marker expression, and so it is argued whether all basal/myoepithelial cells are MaSCs or just a minority (Granit et al., 2013). Recently, a study that inspected the reconstitution capacity of isolated basal cells found that ~50% of cells demonstrate reconstitution capacity (Prater et al., 2014). During embryonic development the entire mammary epithelium is generated from bipotent MaSCs expressing basal markers, yet their subsequent fate remains disputed: some lineage-tracing studies in mice have argued that the basal and luminal lineages are actively maintained by bipotent MaSCs (Rios et al., 2014), while others claim that each compartment is maintained by a separate, unipotent stem cell (Giraddi et al., 2015; Van Keymeulen et al., 2011). The situation in the human mammary gland is currently unknown in this respect.

The luminal layer presents its own complexity: luminal progenitors that are marked by an $EpCAM^{+}CD49^{+}$ profile can grow under anchorage independent conditions, similarly to MaSCs, yet are only able to give rise to luminal cells (Lim et al., 2009). These progenitors are largely $ER^{-}$, yet recent studies have further classified luminal progenitors to several subtypes, including $ER^{+}$ progenitors that give rise to ductal cells, $ER^{-}$ progenitors which give rise to alveolar cells, and a rare population of ERBB3-low expressing population whose function remains unknown (Shehata et al., 2012). Activity of ALDH1 has also been used to identify undifferentiated luminal cells with clonogenic capacity, largely overlapping with the $ER^{-}$ luminal progenitor (Shehata et al., 2012).

In summary, current understanding allows the identification and isolation of three major cell types present in the mouse and human glands: basal cells, which overlap with MaSCs, luminal progenitors that rare largely $ER^{-}$ and mature luminal cells that are largely $ER^{+}$. These profiles are also reflected in the cancerous tissue, as detailed below.

### 1.1.4 Links between breast cancer subtype and normal lineages

Following the characterization of tumor subtypes and identification of cell populations in the normal tissue researchers have begun to explore the possible links between the two. Tumors may retain features of their cell of origin, and therefore tumors of different subtypes could reflect different cells of origin. In addition, molecular pathways that determine cell differentiation state, may act in tumors to maintain or promote these states. The activation

or repression of particular pathways and regulators could drive changes in differentiation of the tumor overall state or of cell subsets in it, leading to heterogeneity (Granit et al., 2013; Skibinski and Kuperwasser, 2015). Understanding these links could therefore prove valuable to the understanding of tumor heterogeneity and biology and could potentially be used to alter tumor characteristics.

Relying on these assumptions, researchers have compared the expression signature of FACS-isolated healthy mammary tissue cell subpopulations, labeled by a combination of surface markers as described above, to that of different breast cancer subtypes (Lim et al., 2009; Raouf et al., 2008). This comparison revealed important links: tumors of the luminal A, B and HER2-like subtypes showed highest levels of expression of the signatures of mature luminal cells, with HER2 showing somewhat lower levels of this signature. Importantly, basal-like tumors showed the highest level of expression of the luminal progenitor cells signature (Figure 2) (Lim et al., 2009). This finding was important because it established a link between this tumor types and the luminal layer, as well as to progenitor identity, as opposed to the basal/MaSC layer. It was tumors of the rare claudin-low subtype, which is enriched for mesenchymal-like cells, which showed high expression of the basal/MaSC cell signature (Figure 2) (Lim et al., 2009).

Several studies have provided experimental support to these observations. Isolation of human $CD10^{+}$ basal/myoepithelial cells and their subsequent transformation generated metaplastic tumors that are rarely found in the clinic, yet these present some features that resemble claudin-low tumors and thus support a possible connection between the two (Gudjonsson et al., 2005). In contrast, it is was noted that transformation of cells originating in the luminal layer generated much of the diversity found among breast tumors, including basal-like cancers (Proia et al., 2011).

Further ties between the luminal layer and cancer are emphasized by the notable connection between normal luminal progenitors and basal-like tumors which has been extensively studied (Blanpain, 2013; Molyneux et al., 2010; Proia et al., 2011). Basal-like tumors are highly prevalent in patients carrying germline mutations in BRCA1, and as noted above, often carry p53 mutations, which are detected in 80% of these malignancies (Cancer Genome Atlas Network, 2012; Molyneux et al., 2010). The link between luminal progenitors and basal-like tumors has been further supported by the fact that BRCA1 carriers were found

to have an expanded pool of luminal progenitor cells (Lim et al., 2009; Proia et al., 2011), and by functional experiments in mice demonstrating that the deletion of BRCA1 or p53 in luminal progenitors led to the formation of tumors with basal-like features (Liu et al., 2007; Molyneux et al., 2010). Additional studies illustrated that mouse luminal cells transformed *in vivo* with the oncogenic Etv6-NTRK3 fusion or mutant PI3K, formed tumors of mixed basal and luminal differentiation (Meyer et al., 2011; Tao et al., 2014). Thus, it appears that normal differentiation programs, inherited from the cell of origin, significantly contribute to intertumoral heterogeneity, yet the study of specific pathways that mediate this heterogeneity and how the luminal layer can give rise to such diverse subtypes lacking.

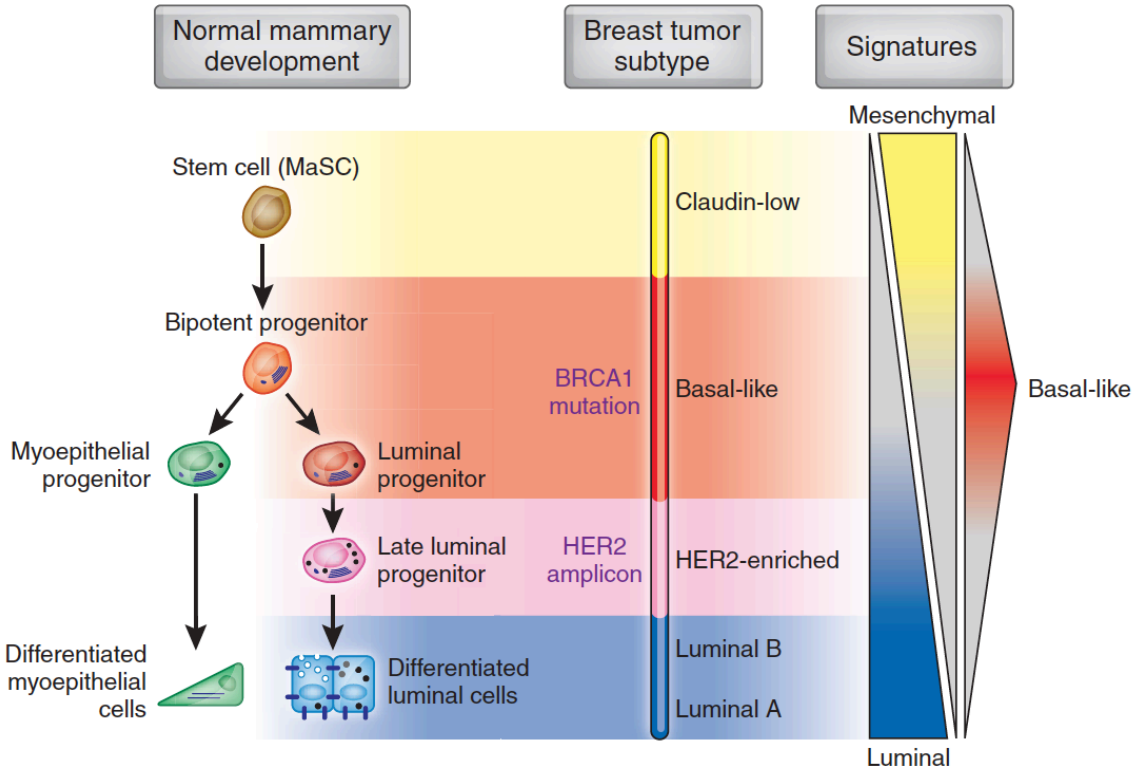



**Figure 2: Ties between the normal breast lineage and tumor subtypes.** Illustration of the normal breast lineages (left) and proposed similarities to different tumor subtypes (right) that is based on resemblance in their gene-expression signature (Prat and Perou, 2009).

### 1.1.5 Breast cancer stem cells

Differentiation-like events are one of the factors that can contribute to intratumoral heterogeneity: tumor cells may undergo differentiation changes that endow them with a different phenotype. The cancer stem cell (CSC) model proposed the existence of subpopulations of cells in tumors that have stem-like traits. CSCs are hypothesized to possess self-renewal capacity and to give rise to more differentiated cell progeny (Reya et al., 2001; Visvader and Lindeman, 2008). CSCs are also believed to harbor greater resistance to chemotherapy (Li et al., 2008; Magee et al., 2012) and this, together with their increased self-renewal capacity, could allow CSCs to support tumor growth and initiate metastases (Visvader and Lindeman, 2008). Therefore, the study of CSCs and the cellular hierarchy within the tumor offers potential therapeutic value: therapeutic elimination of CSCs could root out the driving force of the tumor and open a door for curative treatment. The CSC model thus provided a model describing the potential roles of differentiation events in cancer.

Breast cancers were the first solid tumor type in which the existence of CSCs was reported (Al-Hajj et al., 2003). To date several markers of breast CSCs have been identified, allowing to enrich for these cells and study their properties. To determine the tumorigenic potential of epithelial tumor cells, these are injected into mice to estimate their tumor initiating capacity, or seeded *in vitro* to test their anchorage-independent growth as spheroid structures. Formed tumors or spheroids are assessed to determine whether they reconstitute the initial heterogeneity, and further passaged to determine the self-renewal of the CSCs. Tumorigenic cells presenting the $CD44^{high}$ $CD24^{low}$ marker profile, or high activity of ALDH1, were found to be enriched with CSC traits (Al-Hajj et al., 2003; Ginestier et al., 2007). Interestingly, these two approaches for isolating CSC appear to select for cells with distinct characteristics: $CD44^{high}$ $CD24^{low}$ cells were found to be enriched with mesenchymal traits (discussed below), while $ALDH1^{+}$ cells are of epithelial nature (Visvader and Lindeman, 2008). Additional biological features characterize breast CSCs: it was shown that in certain cases these cells are slow-cycling and can be labeled using dye retention assay, which marks slow-diving cells (Moore and Lyle, 2011; Pece et al., 2010).

Until recently, it was thought that the transition of CSCs into a more differentiated state is a unidirectional process, similarly to the hierarchy attributed to normal stem cells. Yet recent experimental data suggested that more plasticity exists, and have provided a

revised model of CSCs according to which cancer cells can readily transition in and out of the stem-like state, and CSCs can be generated *de novo*, as opposed to being derived from transformed tissue stem cells (Gupta et al., 2009; Marjanovic et al., 2013; Meacham and Morrison, 2013). Moreover, it has been noted that intermediate differentiation states endow tumor cells with increased flexibility and support malignancy (Brooks et al., 2015; Condiotti et al., 2014; Marjanovic et al., 2013; Yu et al., 2013). Together, it appears that CSCs can differ from normal tissue SCs in the expression of markers as well as their characteristics. In some cases, markers such as ALDH1 are conserved between cancerous and normal tissue (Ginestier et al., 2007), while the "main" markers, $CD44^{high}$ $CD24^{low}$, do not appear to label a normal stem/progenitor population. It is therefore likely that different cancer subtypes harbor distinct types of CSCs (Ricardo et al., 2011).

Altogether while it appears that CSCs may contribute to the generation of intratumoral heterogeneity and the complexity found in some breast cancers. A broad characterization of the existence of different types of CSCs across multiple breast cancers of different subtypes, of their respective functions, and of their mode of generation, is still lacking. The investigation of the mechanisms that control their generation, plasticity and function are clearly of importance.

### 1.1.6 Epithelial to mesenchymal transition and breast cancer stem cells

One of the most recognized transcriptional programs affecting breast cancer differentiation and contributing to its intratumoral heterogeneity is the epithelial to mesenchymal transition (EMT) program. The origins of this process are normal developmental processes, in which epithelial cells undergo trans-differentiation to acquire mesenchymal traits (Thiery et al., 2009). In this process, the cells remodel their cytoskeleton structure, downregulating cytokeratins and upregulating the mesenchymal filament vimentin. Cell membrane proteins are also altered, replacing integrins and E-cadherin, that are required for adhering to the basement membrane and neighboring cells, by N-cadherin and additional proteins that support a less adherent state (Morel et al., 2008; Thiery et al., 2009). EMT features such as a spindle-like morphology and an associated typical gene expression signature were found to be associated with normal tissue MaSCs and the basal compartment (Guo et al., 2012; Visvader, 2009), suggesting that the two state are associated (Granit et al.,

2013).

To execute EMT, cells activate a transcriptional plan that is regulated by several master regulator TFs, including Snail1, Snail2 (Slug), Twist1 and ZEB1/2 (Lamouille et al., 2014). The specific activated transcription factors appear to operate under different settings, such as the TGF-β pathway that can stimulate EMT, and additional pathways can also do so in different conditions (Thiery et al., 2009; Polyak and Weinberg, 2009; Heldin et al., 2009).

In light of this, it was demonstrated that breast cancer cells can active these factors and pathways to undergo EMT, which in turn endows them with migratory and invasive abilities that allow them to migrate and invade neighboring tissues and increase their metastatic capacity (Chaffer and Weinberg, 2011). Once cells have migrated and reached the blood stream, it is thought that they must undergo the reverse process of mesenchymal to epithelial transition (MET) in order to colonialize the target organ and form metastases (Gunasinghe et al., 2012; Tsai and Yang, 2013; Tsai et al., 2012). Furthermore, it was found that, circulating metastatic breast cancer cells in the blood display mixed expression of epithelial mesenchymal markers, fitting with the enhanced migratory abilities associated with this phenotype and with their ability to undergo MET (Yu et al., 2013).

Expression of EMT TFs in breast tumors was found to contribute to the expansion of the $CD44^{high}$ $CD24^{low}$ CSC population (Mani et al., 2008; Polyak and Weinberg, 2009). Claudin-low tumors represent an extreme case in which these factors are highly expressed and majority of the population appear to undergo EMT, and are therefore enriched with $CD44^{high}$ $CD24^{low}$ cells (Prat et al., 2010). The key regulator Slug was found to be active also in normal basal cells, and plays a critical role in basal/MaSC differentiation (Guo et al., 2012), further supporting the association between the basal lineage and EMT. However, it has recently been shown that in cancer EMT activation relies more significantly on Snail than on Slug relative to MaSC, illustrating that there is some distinction in the activation of EMT among these states (Ye et al., 2015). EMT thus promotes not only invasive properties, but also stem-like traits, manifested in some cases by the acquisition of the $CD44^{high}$ $CD24^{low}$ profile. This adds an additional complexity to breast cancer and its heterogeneity that is of great functional significance.

### 1.1.7 Axes of differentiation in breast cancer

From the analyses described above, it is possible to consider differentiation states in breast cancer along several axes: a **basal <-> luminal** differentiation axis on which cancer cells may transition between these normal lineage states, the EMT axis, on which they transition between the **mesenchymal <-> epithelial** states, and a **stem-like <-> differentiated** axis (Granit et al., 2013). These axes apply to tumors and tumor subtypes, as well as to individual cells within tumors. This type of view allows to evaluate differentiation states on a continuous (rather than discrete) scale and to track the effect of different treatments. Moreover, this approach places increased emphasis on the consideration of intermediate differentiation states and the plasticity of changes along each axis.

These axes of differentiation could be co-aligned, yielding an even more complex description of the tumors' state, and allowing to inspect how these axes correlate with each other (Figure 3). Placing these three axes parallel to each other illustrates the links discussed above between the basal identity, mesenchymal traits, and stem cell function, and between luminal identity to an epithelial, more differentiated state. Intersecting the axes on a three-dimensional grid provides a potentially more detailed means to dissect shifts in differentiation, since it allows to transpose along certain axis independently of the others (Figure 3).

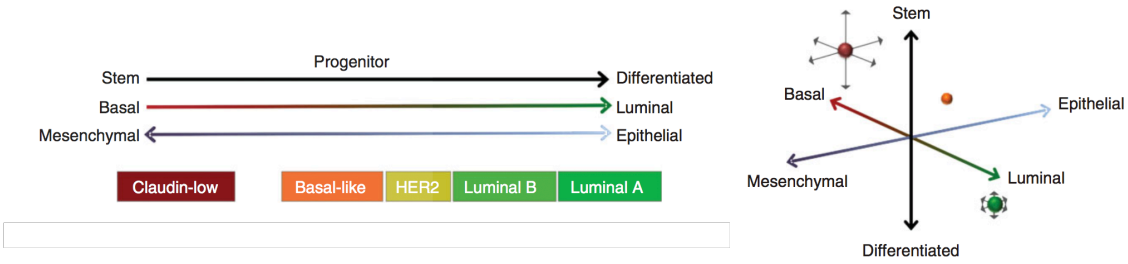



**Figure 3: Axes of differentiation in the normal and cancerous breast.** Left: alignment of breast cancer subtypes on a combined linear axis. Right: the same differentiation axes, now positioned in a three-dimensional scale, with spheres illustrating potential positions of breast cancer subtypes (Granit et al., 2013).

The directionality of shifts along these axes have gained increased research attention in recent years. As mentioned above, the changes alongside the stem <-> differentiated axis, have previously been considered to be unidirectional, yet this view has recently changed, supporting a more plastic behavior in the cancerous tissue (Marjanovic et al., 2013). It is

generally accepted that basal cells are the source of all cells in the mammary gland, thus suggesting that the basal <-> luminal axis is unidirectional, yet it appears that progenitor cells possess an intermediate basal-luminal identity in some respects, and transitions between luminal versus differentiated cells are poorly characterized. Finally, across the EMT <-> epithelial axis movement appears to be bidirectional, as it was shown that metastatic cells that have undergone EMT go through a reversal to an epithelial state during seeding in the target organ (Yu et al., 2013).

In my research I explore the traits of cells found in an intermediate state, the relationship between these axes and attempt to uncover the regulatory mechanisms that regulate each of them (chapters 2,4).

## 1.2 Regulators of differentiation in the breast

Key regulators that control normal breast differentiation might perform a similar role in cancer, therefore their study can potentially uncover mechanisms that control tumor heterogeneity. This could be done by evaluating how these regulators and their perturbation translates to changes across the axes of differentiation, using the direction and magnitude of change as means to evaluate their regulatory significance.

Various transcription factors have been implicated, before and during my studies, in regulating differentiation events in the breast. Among these are GATA3, FOXA1 and TBX3 that govern luminal differentiation, SLUG and TP63 which control basal differentiation, and SOX9, FOXM1, EZH2 and ELF5 that maintain luminal progenitor state (Granit et al., 2013; Visvader, 2009).

As I began my studies, several regulatory factors have been noted to be highly expressed in basal-like tumors. However, a broad view of the regulation of differentiation states in breast cancer cells was lacking and functional contribution of these factors remained unknown. Factors that have been suggested to regulate normal luminal progenitors or those that drive luminal identity, such as GATA3 and FOXA1, could also potentially affect/inhibit basal-like differentiation, yet a functional dissection of this hypothesis was lacking.

To investigate how differentiation is functionally regulated in basal-like tumors I first compiled a list of potential regulators using a broad computational analysis of breast cancer gene expression profiles. In this bioinformatics analysis I have extracted genes with potential regulatory function such as TFs, chromatin regulators and kinases that I found to be statistically associated with basal-like tumors using published breast cancer expression profiles (Figure 4). Together this candidate list contained a wide-array of genes with diverse regulatory potential, some are sequence specific TFs, such as FOXA1, while others are broad chromatin modifiers such as EZH2 and RNF20 (of which I discuss later). This analysis provided a starting point which guided me to test specific regulators, starting with the Polycomb factor EZH2, as well as to conduct a functional screen aimed at identifying proteins that control the composition of breast cancers, and included many of these candidates (Figure 4).

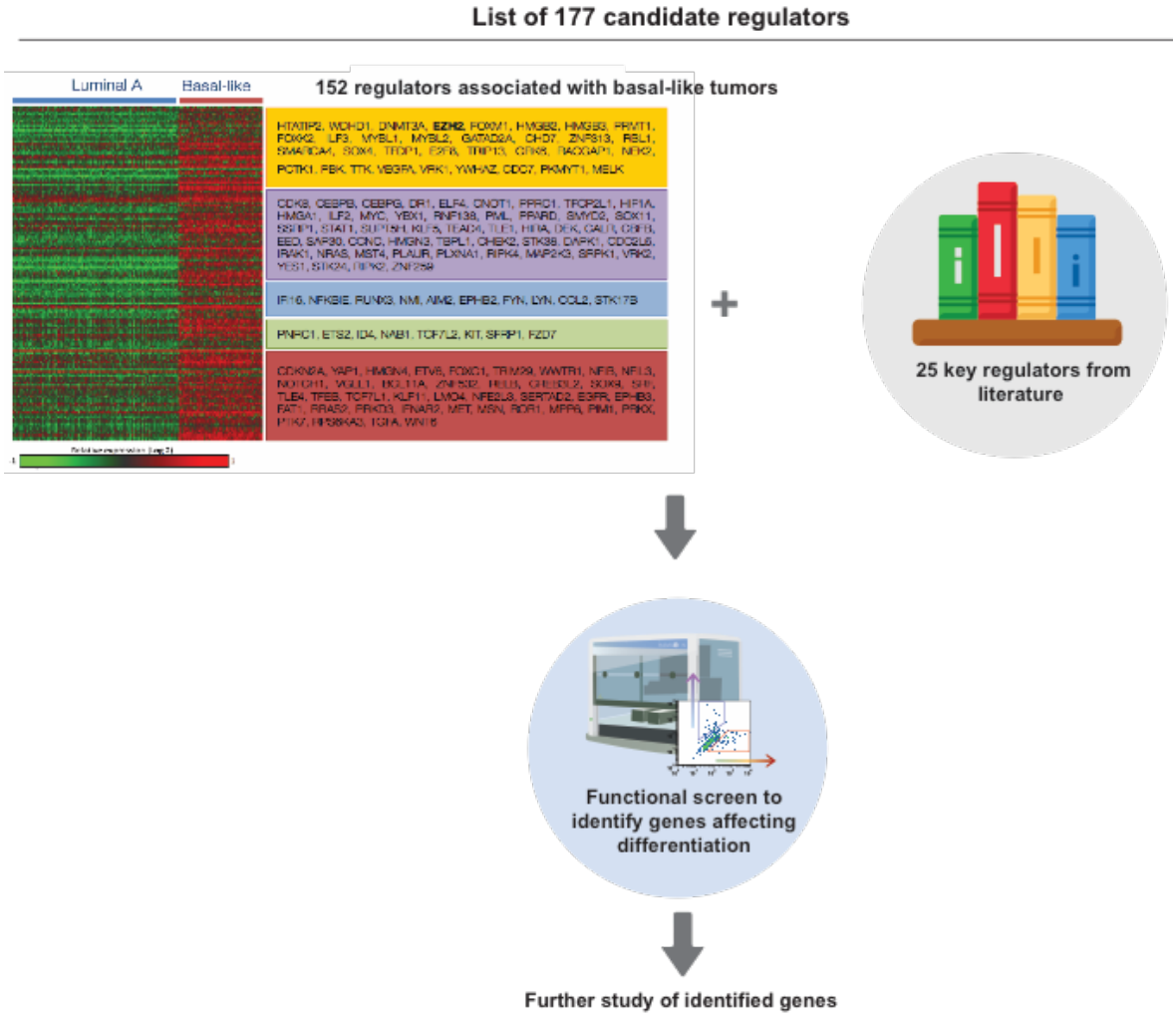


**Figure 4: Outline of the screening process.** 152 genes with potential regulatory function (TFs, chromatin regulators, kinases) that are specifically associated with basal-like tumors were identified in a bioinformatic analysis. These were supplemented with 25 known differentiation regulators derived from the literature. Each of these genes was than silenced using shRNA's and screened using FACS to identify its effect on differentiation. Next, top hits were selected for in-depth investigation.

In my work I have conducted a functional screen methodology to systematically identify which of these genes functionally affect basal-like differentiation. These genes were then further studied and the relationship between these was characterized (chapter 2). In parallel to this screen I have also taken a biased approach in which I have individually studied genes that appeared to have promising regulatory potential (chapter 1,3). Below I describe some of the main pathways and genes that I identified and studied during my work.

### 1.2.1 The Polycomb group

The bioinformatics analysis pointed to several known regulators of differentiation and stemness that were highly active specifically in basal-like breast cancer. Among them was the Polycomb factor EZH2, which I decided to focus on in the initial stages of the work. The Polycomb group (PcG) is a group of transcriptional repressors, known to regulate lineage choices during development and differentiation (Sparmann and van Lohuizen, 2006). PcG complexes maintain the non-differentiated state of SCs and progenitors by silencing developmental TFs, and are also critical for proper activation of differentiation programs upon lineage commitment (Boyer et al., 2006; Lee et al., 2006; Sauvageau and Sauvageau, 2010). Gene silencing is initiated by the recruitment of the Polycomb Repression Complex 2 (PRC2) to the target loci where EZH2, its catalytic subunit, trimethylates Histone 3 at Lysine 27 ($H3K27Me^3$) leading to chromatin condensation and transcriptional repression (Sauvageau and Sauvageau, 2010). Another member of this complex is EED, which can bind $H3K27Me^3$ to perform a positive feedback loop, and is also necessary for the methylation process (Margueron and Reinberg, 2011). Subsequently, the PRC1 complex is recruited to the $H3K27me^3$ labeled loci to facilitate long-term repression by monoubiquitylation of lysine 119 of histone H2A (H2AK119ub1), further inhibiting transcriptional elongation and facilitating chromatin condensation (Sauvageau and Sauvageau, 2010). More recent studies have suggested that under certain conditions EZH2 activity may promote gene activation

(Asangani et al., 2013; Gonzalez et al., 2014; Xu et al., 2012a), yet the significance of these mechanism remains unclear.

PRC2 and its catalytic subunit EZH2 have been noted to play a major role in the development and maintenance of the breast tissue. EZH2 is important for the maintenance of the MaSC and luminal progenitor pool and its absence leads to impaired growth and morphogenesis of milk ducts (Michalak et al., 2013; Pal et al., 2013). The pattern of $H3K27me^3$ differs among the different cell compartments of the breast, and it appears to participate in the regulation of expression programs in each cell type, by influencing the levels of the main lineage-restricted regulators (Pal et al., 2013). During pregnancy hormonal cues elevate $H3K27me^3$ levels in the breast to its peak, to support the process of alveologenesis and lactation (Pal et al., 2013). Together these finding suggest a critical role for PRC2 in influencing breast differentiation and function.

EZH2 is an established oncogene, highly expressed in several solid tumor types, among them breast and prostate cancers where it is associated with proliferation and poor survival (Collett, 2006; Kleer et al., 2003; Varambally et al., 2002). More recent evidence revealed that EZH2 is also critically involved in hematological malignancies, exhibiting frequent activating mutations in lymphomas and recurrent inhibitory mutations in myeloid cancers (Chase and Cross, 2011; McCabe et al., 2012; Shih et al., 2012).

To counter these functions several pharmacological inhibitors have been developed to target EZH2 specifically, and are currently under clinical evaluation (Kim and Roberts, 2016). In my work I have used the specific EZH2 inhibitor, GSK-126, which has been found to be efficient in targeting EZH2 in mutant lymphomas and also additional tissues types (McCabe et al., 2012). In breast cancer EZH2 is thought to play an tumor-promoting role, and its expression is highly associated with $ER^-$, basal-like tumors, increased metastasis and poor patient survival (Kleer et al., 2003; Alford et al., 2012; Pietersen et al., 2008).

Despite its association with basal-like tumors, little is known about the role of EZH2 in determining breast cancer heterogeneity and differentiation. Expression of EZH2 was suggested to mark early-stage hyperplasia in the breast, and its overexpression in normal mammary cells was found to induce hyperplasia (Ding and Kleer, 2006; Kleer et al., 2003). Furthermore EZH2 was found to impair DNA repair in breast cells, and demonstrates a

negative relationship with the DNA repair protein BRCA1 (Puppe et al., 2009; Wang et al., 2013b). Mutations in EZH2 in breast cancer are rare and it appears to be regulated at the transcriptional level (Nik-Zainal et al., 2016; Yamaguchi and Hung, 2014): it is activated transitionally during cell cycle, and also by factors such as HIF1 and the MEK-ERK pathway and is inhibited by BRCA1 and by several microRNAs (Bracken et al., 2003; Fujii et al., 2011; Sander et al., 2008; Wang et al., 2013b). We were specifically intrigued to study EZH2 due to it known role as a master regulator of differentiation, and speculated that it might preform the same role in cancer, and this analysis is detailed in chapters 1,2. We found that EZH2 is a key regulator important for the maintenance of basal-like subtype identity and progenitor-like traits, indicating a role in intertumoral heterogeneity. I next found that EZH2 in fact also controls the relative proportions of subpopulations within the tumor, and does so mechanistically by promoting symmetric versus asymmetric cell divisions, thereby executing an important role in controlling intratumoral cell diversity and tumor composition.

### 1.2.2 Notch signaling pathway

Compelling results originating from the functional screen I have conducted, demonstrated a role for the Notch pathway in regulating heterogeneity and tumor composition. This, along with my finding that Notch acts downstream to EZH2, and is a known regulator of cell fate choices (Gonzalez et al., 2014; Kopan and Ilagan, 2009), have driven me to study this pathway in-depth. Notch is a central, highly conserved and studied pathway, which plays a critical role in communicating multi-cellular processes during stages of development and tissue maintenance (Chiba, 2006; Mumm and Kopan, 2000). Notch performs these functions by directing cell fate choices of stem cells and progenitors to determine their abundance in various tissues, including the mammary gland (Artavanis-Tsakonas et al., 1999). The pathway is activated when one of the four Notch receptors binds an appropriate ligand, of the Delta or Jagged protein family, expressed on the membrane of a neighboring cell. Ligand binding triggers proteolysis of the receptor by the γ-secretase complex and release of the Notch intracellular domain (NICD), which in turn enters the nucleus, binds the TF RBPJ and activates Notch target genes (Kopan and Ilagan, 2009). Activated Notch can maintain the signal-receiving cell in an uncommitted state, supports self-renewal and induces expression of the Notch receptors, while cells in a more

differentiated state may express the Notch ligands (Chiba, 2006; Kopan and Ilagan, 2009). Together, this structure provides a mechanism for effective self-maintenance of tissue composition by the cells that constitute it. Consequently, aberrations in the pathway are often observed in several cancer types including hematological malignancies and solid tumors (Ranganathan et al., 2011). In most cases, these aberrations lead to an oncogenic function, yet in several of solid tumors activation of Notch performs the opposite and acts as a tumor suppressor, suggesting a context dependent effect (Lobry et al., 2011; Ranganathan et al., 2011).

Notch plays an important regulatory role in the development of the mammary gland, where it was found to restrict MaSC function and drive cells to commit towards the luminal lineage, while its aberrant activation has led to the expansion of the luminal progenitor pool and hyperplasia (Bouras et al., 2008; Farnie and Clarke, 2007). Additional support for its role in promoting self-renewal comes from tissue culture experiments demonstrating that Notch supports the formation of anchorage-independent spheroids (mammospheres) (Dontu et al., 2004).

Deregulation of the Notch pathway has been implicated to play a critical part in the cancerous breast tissue. Overexpression of the Notch receptors in transgenic mice has been demonstrated to promote hyperplasia and induce tumors (Farnie and Clarke, 2007; Hu et al., 2006; Raafat et al., 2004). Moreover, in culture, Notch was found to be highly active in breast CSCs, and its inhibition using gamma secretase inhibitors (GSI) or inhibitory antibodies led to a decrease in mammosphere formation and also hampered tumor formation *in vivo* (Farnie and Clarke, 2007; Harrison et al., 2010). Additionally, in a study that I have contributed to, it was shown that Notch activity is important for proliferation and 'luminal filling' of milk ducts in breast carcinomas over-expressing HER2 (Pradeep et al., 2012). In the clinic, Notch activation was found to mark early-stage lesions and that tumors over-expressing Notch1 or the ligand Jag1 correlate with poor clinical survival (Farnie and Clarke, 2007; Reedijk et al., 2005). An additional mechanism of Notch hyper-activation in tumors is the loss of its negative regulator Numb, which ubiquitinates the Notch1 receptor, leading to proteasomal degradation of NICD (McGill and McGlade, 2003). Furthermore, Numb has been demonstrated to inhibit ubiquitylation of p53, and its loss may also reduce p53 tumor suppressive activity (Colaluca et al., 2008).

Several studies have indicated that activation of Notch is associated with basal-like tumors and suggested that it might contribute the identity of these cancers (Lee et al., 2008; Xu et al., 2012b; Yamamoto et al., 2013). Notch appears to interact with additional pathways active in these malignancies: inflammatory cytokines secreted by these tumors activate the nuclear factor κB (NF-κB) pathway, which in turn induces Jag1 expression in stromal and non-CSCs that induces the expansion of the CSC population (Yamamoto et al., 2013). Interestingly a recent study from the Kleer group has demonstrated that EZH2 acts as a positive regulator of Notch1 in basal-like tumors and that they together act to increase the fraction of breast CSCs (Gonzalez et al., 2014), findings that support my own. These central regulatory roles played by Notch, the association with EZH2 and basal-like subtype and our initial experimental findings have led us to explore Notch's function as a regulator of tumor differentiation. I found that Notch regulates the intratumoral heterogeneity and tumor composition, acting down-stream to EZH2 and NFIB to drive symmetric cell divisions. The full findings are entailed in chapter 2.

### 1.2.3 Histone H2Bub modification

My study of EZH2 as well as additional studies have demonstrated that chromatin regulators play an important role in the regulation of tumor differentiation, yet the role of broad chromatin modifiers in regulating subtype-associated gene expression programs is largely unknown. One such histone post-translational modification that plays a regulatory role in differentiation and contributes to heterogeneity is H2B mono-ubiquitylation, which in mammalian cells occurs on Lysine 120 of the histone tail (Fuchs and Oren, 2014). H2Bub1 orchestrates various processes by changing the biochemical properties of chromatin, including nucleosome assembly, transcriptional elongation, recruitment of DNA repair factors and additional functions (Cao and Yan, 2012; Fuchs and Oren, 2014; Fuchs et al., 2014). The ubiquitylation of H2B is catalyzed by an E3 ligase heterodimer complex, composed of the proteins RNF20 and RNF40, and inhibited by SMURF2, which acts as an inhibitor of RNF20 (Blank et al., 2012).

At the cellular level, H2Bub1 plays a critical role in allowing proper differentiation of embryonic stem cells and of muscle cells (Chen et al., 2012; Vethantham et al., 2012) and in the context of the breast was found to be required for the expression of ER target genes

(Bedi et al., 2015; Prenzel et al., 2011). During inflammation, levels of H2Bub1 decrease, driving the expression of NF-κB target genes and inflammatory cytokines and its down-regulation was observed in several inflammation driven malignancies (Tarcic et al., 2016). Moreover loss of H2Bub1 has been linked to high grade and poor prognosis in colorectal cancers (Melling et al., 2016).

H2Bub1 has been suggested to play a tumor suppressive role in mammary cells, where depletion of RNF20 promoted cell migration and tumorigenesis in non-transformed MCF10A cells (Shema et al., 2008). Furthermore, RNF20 was found to be important for the expression of p53 and the attenuation of several oncogenes and its promoter was found to be hyper-methylated in tumors, consistent with a tumor suppressive function and its downregulation that was observed in several cancers (Melling et al., 2016; Shema et al., 2008; Wang et al., 2013c). In contrast, recently it has been suggested that H2Bub1 levels are up-regulated in breast tumors and in MLL-rearranged leukemia, and that its negative regulator SMURF2 is down-regulated in cancers, suggesting an oncogenic role for H2Bub1 in these settings (Blank et al., 2012; Wang et al., 2013a). Thus the role of H2Bub1 in breast cancer remains to be elucidated, and the differential findings suggest specific contexts may dictate the role of this modification. I investigate these questions in chapter 3.

# 2. Research aims

Tumor heterogeneity has pronounced implications on tumor biology and clinical characteristics that can impact patient survival. Despite this, the biological mechanisms that control of such diversity are poorly understood.

Breast cancers offer a powerful system for the study of heterogeneity, as these display diversities at several levels, ranging from the occurrence of different tumor subtypes to the identification of subpopulations of cells with distinct biological traits. As discussed above, several axes of differentiation appear to act within breast tumors, including the basal <-> luminal, mesenchymal <-> epithelial and stem <-> differentiated. These provide the means to measure phenotypic diversity, and evaluate heterogeneity and changes in tumor composition. Moreover, the activity of different pathways such as Notch and Polycomb has been characterized in these tumors, and these, alongside other pathways and factors, represent potential regulators of heterogeneity.

My research aims to uncover molecular mechanisms regulating heterogeneity in the differentiation state of breast cancer cells, focusing on the roles of transcription factors and epigenetic regulators. Specifically, I addressed this in basal-like breast cancers, which appear to have a high level of heterogeneity, and represent a major clinical challenge. The study intends to address both subtype identity control, and also to dissect aspects of regulation of intratumoral heterogeneity, assuming that the two are intertwined. Additionally, the research aims to unravel the functional tumorigenic implications of changes in heterogeneity and differentiation.

**Specific research aims**

1. To study the regulation of the differentiation properties of basal-like breast cancer by the Polycomb factor EZH2 and its mechanism of action.
2. Define the nature of differentiation states of cells in basal-like breast cancers, their links to normal tissue and their functional tumorigenic traits.

3. Investigate the regulatory network governing the heterogeneity and differentiation state of basal-like breast cancer cells and the mechanistic manner by which transitions between states occur.
4. Investigate the roles of the epigenetic modification H2B ubiquitylation and its regulators in different breast cancer subtypes.

The following chapters explore in depth the questions that were raised along this introduction. The findings add important new insights into the biological significance of tumor heterogeneity and the mechanisms that control it.

**Chapter 1-** Provides an investigation of the role of EZH2 in basal-like breast tumors, and establish it as an important factor in maintaining progenitor/subtype identity. Moreover, it points to the existence of a subpopulation of cells with mixed basal-luminal differentiation that is unique to basal-like tumors, possess progenitor functions, and links EZH2 to its maintenance. Lastly it reveals the repression of GATA3 by EZH2 and the opposing effects on the mixed-lineage phenotype. Together, this chapter addresses Aims 1-3.

**Chapter 2-** Delivers direct continuation to chapter 1, adding further investigation of the mixed-lineage population, revealing the transcriptional programs that are active in it and its functional tumorigenic traits. This study deepens the analysis of the regulatory networks that control tumor heterogeneity by dissecting the contribution of 177 additional candidate regulators that were functionally screened. The screen exposes additional factors that influence basal-like tumor composition, including Notch, and sheds light about the relationship amongst these factors. It also identifies symmetric/asymmetric divisions as mechanism that could allow differentiation transition to take place in breast cancer and exposes the roles of EZH2, Notch, as well as the luminal regulators GATA3 and FOXA1 in regulating division type. Together, this chapter addresses Aims 1-3.

**Chapter 3-** In this chapter I explore H2Bub1, an epigenetic modification that has been implicated in several cancer types, yet its role in breast cancer has been contradictory. We study the distribution of H2Bub1 across breast cancer subtypes and find that this modification and its drivers are more prevalent in luminal versus basal-like tumors. We also find that in luminal tumors this modification performs an oncogenic function by promoting

the expression of ER target genes, while in basal-like tumors it acts as a tumor suppressor by inhibiting the expression of inflammatory cytokines that act as drivers in this setting. This chapter addresses Aims 3,4.

**Chapter 4 (supplement)-** A review article that explores the different axes of differentiation active in breast cancer, their origin, significance and the overlap among these axes.

# 3. Results

## Chapter 1 : EZH2 promotes a bi-lineage identity in basal-like breast cancer cells

Status: published in Oncogene (2013)

R.Z. Granit, Y Gabai, T Hadar, Y Karamansha, L Liberman, I Waldhorn, I Gat-Viks, A Regev, B Maly, M Darash-Yahana, T Peretz and I Ben-Porath

Oncogene (2012), 1–10

www.nature.com/onc

ORIGINAL ARTICLE

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

RZ Granit[1], Y Gabai[1], T Hadar[1,2], Y Karamansha[1], L Liberman[1], I Waldhorn[1], I Gat-Viks[3,6], A Regev[3], B Maly[4], M Darash-Yahana[5], T Peretz[5] and I Ben-Porath[1]

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.

*Oncogene* advance online publication, 17 September 2012; doi:10.1038/onc.2012.390

**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 cell 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 ~10% of cases, are highly aggressive, and as 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 normal mammary gland luminal lineage, 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 levels 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, cells 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.[10,11] Polycomb maintains stem cell states by silencing differentiation-associated 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 a bi-lineage identity and progenitor-associated traits in basal-like breast cancer cells, 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

[1]Department of Developmental Biology and Cancer Research, Institute for Medical Research – Israel-Canada, Hadassah School of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; [2]Department of Surgery, Hadassah Medical Center, Jerusalem, Israel; [3]Broad Institute of Harvard and MIT and Department of Biology, MIT, Cambridge, MA, USA; [4]Department of Pathology, Hadassah Medical Center, Jerusalem, Israel and [5]Sharett Institute for Oncology, Hadassah Medical Center, Jerusalem, Israel. Correspondence: Dr I Ben-Porath, Department of Developmental Biology and Cancer Research, Institute for Medical Research – Israel-Canada, Hadassah School of Medicine, The Hebrew University of Jerusalem, Jerusalem 91120, Israel.
E-mail: ittaibp@mail.huji.ac.il
[6]Current address: Department of Immunology, Tel-Aviv University, Tel-Aviv, Israel
Received 13 December 2011; revised 16 July 2012; accepted 16 July 2012

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, express both basal and luminal cytokeratins, and express protein markers and mRNA profiles typical of 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 of 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 (Figures 1d–f and Supplementary File 1); EZH2 overexpression promoted the expression of a smaller number of these genes (Figures 1d, f). To gain further resolution, we compiled 12 sets of genes coordinately expressed across breast cancers, each associated with different cancer subcategories and gene

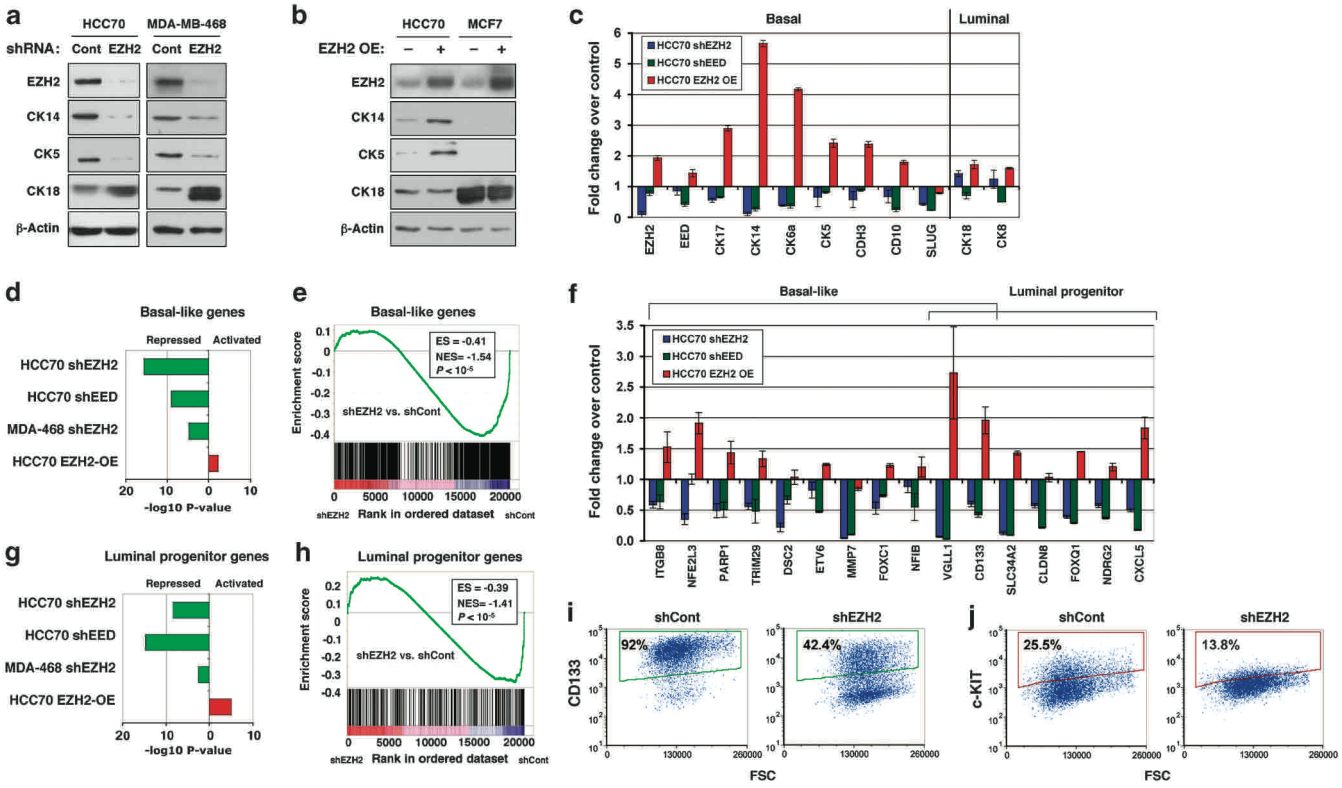



**Figure 1.** EZH2 promotes the expression of basal-lineage and progenitor-associated genes. (**a**) Western blot analysis of basal (CK5,14) and luminal (CK18) cytokeratins in basal-like breast cancer cell lines expressing a shEZH2 or control (shCont) lentivirus. (**b**) Same analysis in basal-like HCC70 and 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 overexpressing cells, relative to control cells, assessed by qRT–PCR. Values indicate average of triplicate reactions ± s.e.m. (**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. VGLL1 and CD133 are included in both sets. (**g**) Preferential repression (green) or activation (red) 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.

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 (Figures 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 (Figures 1i, j).

Together these analyses indicate 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 involving 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 homogenously composed of cells that coexpress 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 10 basal-like tumors for CK14 and CK18. All tumors contained cells expressing only CK18 or CK14; in addition, 9 out of 10 tumors contained double-positive ($CK18^+CK14^+$) 'bi-lineage' cells (Figures 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 closely related 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 ~22% and 34%, respectively, of $CK14^+$ $CK18^+$ bi-lineage cells, whereas the remainder of cells expressed the luminal CK18 only (Figures 2c, d). These lines therefore partly recapitulate the composition of human tumors.

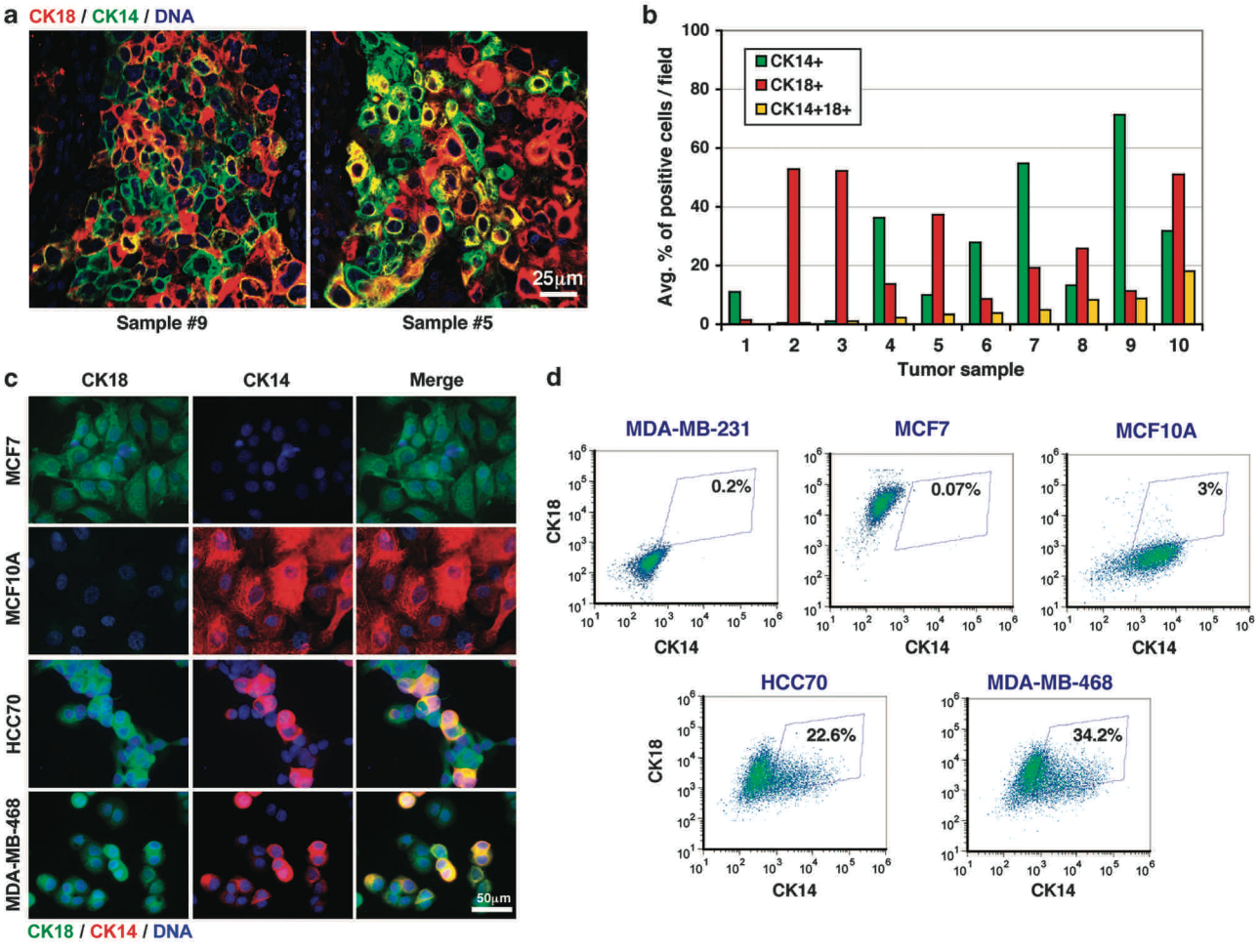



**Figure 2.** A bi-lineage cell subpopulation is found in basal-like breast tumors. (**a**) Co-staining of CK14 (green) and CK18 (red) on tumor sections of human basal-like breast cancers. Double-positive (bi-lineage) cells, appear as yellow/orange. (**b**) Percentage of $CK14^+$ only, $CK18^+$ only and bi-lineage $CK14^+CK18^+$ cells in tumor sections of 10 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 the bi-lineage subpopulation.

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 (Figures 3a and c, Supplementary Figure 5); conversely, EZH2 overexpression caused an increase in the bi-lineage fraction size (Figures 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 (Figures 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 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 their undergoing a change in differentiation state.

We found that EZH2-silenced HCC70 and SUM149 cells formed significantly fewer colonies than control cells (Figures 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 (Figures 4b–d). Upon dissociation of colonies and re-plating, 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 number of colonies formed by HCC70 cells, and these were also larger in size (Figure 4e). Together these findings indicate that EZH2 promotes the colony-formation capacity of the cancer cells.

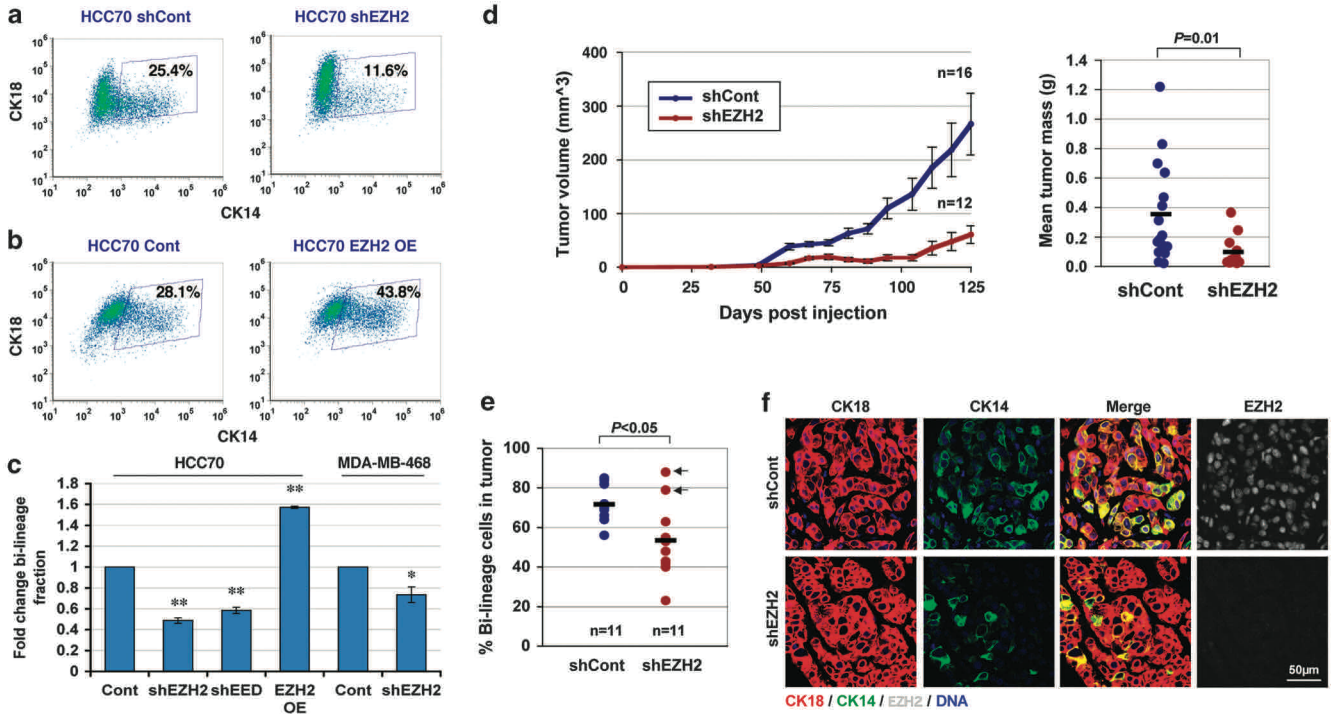



**Figure 3.** EZH2 controls the bi-lineage cell fraction in basal-like cell lines and xenograft tumors. (**a**) FACS analysis of CK14 and CK18 expression in EZH2-silenced and control HCC70 cells. (**b**) Same analysis in EZH2-overexpressing and control cells. (**c**) Fold change of the bi-lineage cell fraction size in the indicated cells relative to matching control cells (Cont), as determined by FACS analysis. Values indicate average of three or more independent experiments $\pm$ s.e.m. $^{*}P<0.03$, $^{**}P<0.003$, *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 $\pm$ s.e.m. (**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. 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.

### 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 coexpression of CK14 and CK18 occurs in mammary stem cells during embryogenesis, but not subsequently in the adult gland;[30,31] however, such cells can be detected 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. One day after plating of HCC70 cells on Matrigel, a time point at which the culture largely contained single cells, 44% of cells showed bi-lineage $CK14^+CK18^+$ staining (Figures 5a, b). Four days after plating, these cultures contained single cells that have not divided, as well as small clusters containing $\sim$5 cells on average, indicating that approximately two cell divisions took place. Strikingly, the percentage of bi-lineage cells among the nondividing single cells was decreased relative to the one-day time point, to 26% ($P = 0.007$) (Figure 5b); in contrast, 91% of the forming colonies contained bi-lineage cells (Figure 5b). Eight days after plating, this trend was further enhanced: only 21.8% of single cells were bi-lineage ($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, whereas 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). As all the 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 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-formation 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 within these forming tumor nodules (Figures 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

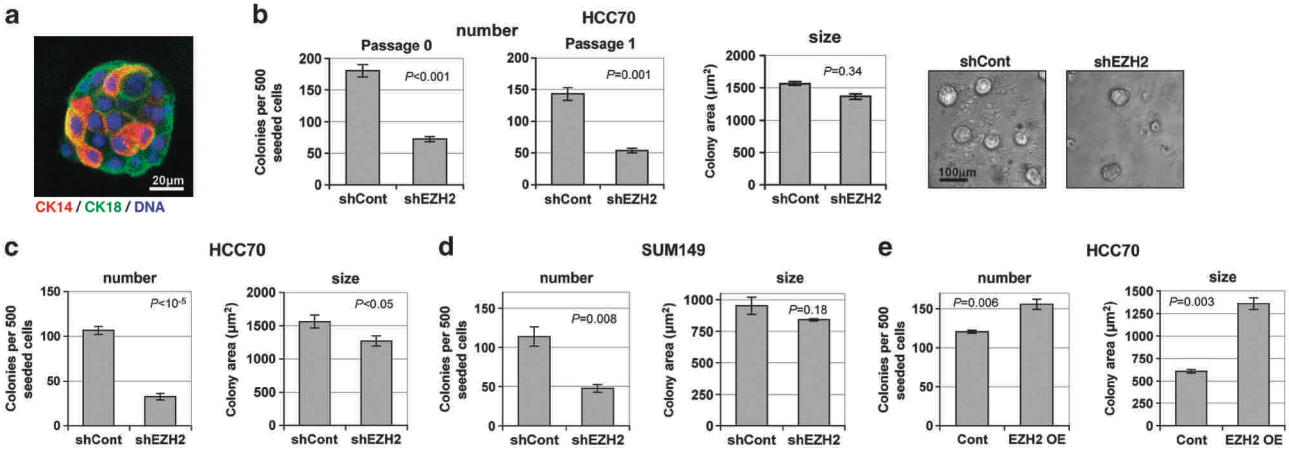



**Figure 4.** EZH2 promotes colony-formation capacity. (**a**) Confocal image of a colony formed on Matrigel by HCC70 cells and stained for CK14 and CK18 as indicated. (**b**) Number of colonies formed on Matrigel by control and EZH2-silenced HCC70 cells upon initial seeding (passage 0) and upon dissociation of colonies and replating (passage 1). Graph on right shows average colony sizes (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 and EZH2-silenced SUM149 cells; colonies displayed morphology and composition similar to those formed by HCC70 cells. (**e**) Colony numbers and sizes of control and EZH2-overexpressing HCC70 cells. Values indicate average of 3–5 replicates $\pm$ s.e.m.

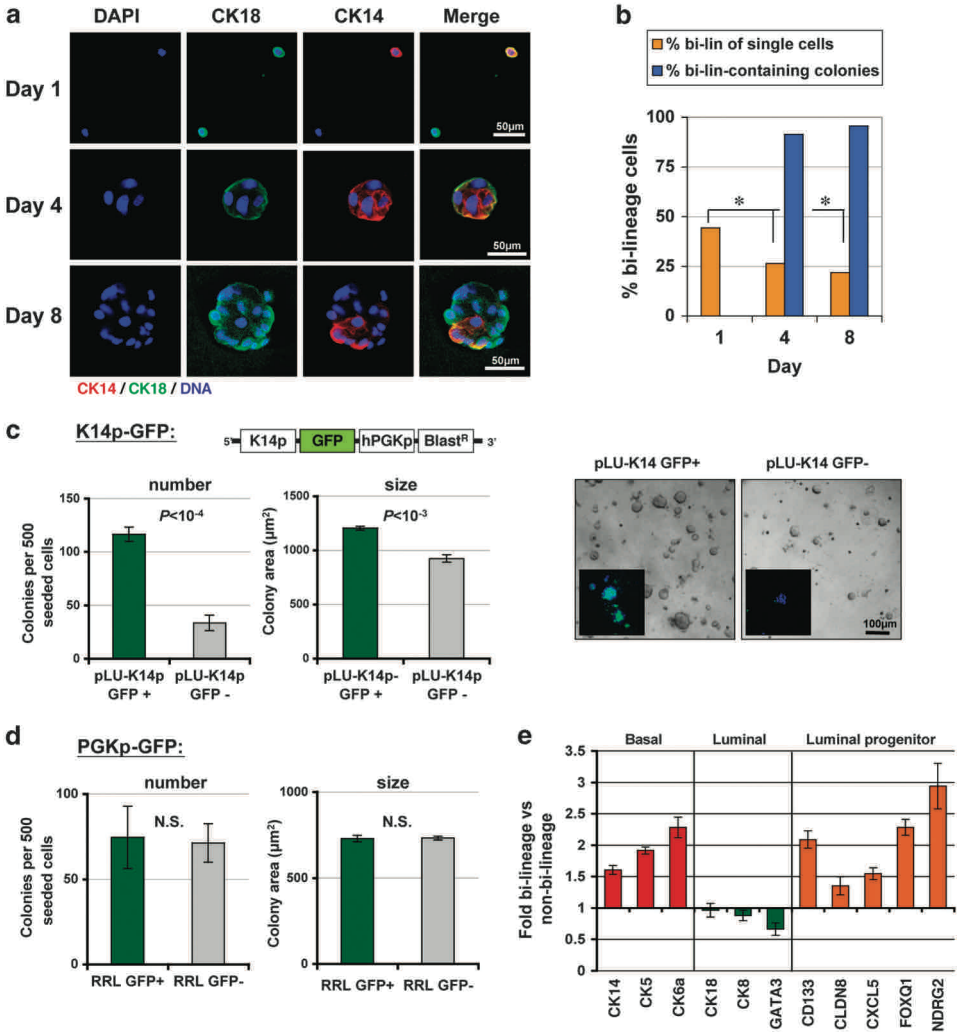



**Figure 5.** Bi-lineage cells possess increased colony-formation capacity and express progenitor-associated genes. (**a**) Confocal images of single HCC70 cells or colonies stained for CK14 (red) and CK18 (green), 1,4 or 8 days after seeding on Matrigel. (**b**) Percentage of bi-lineage cells among 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 $\chi^2$-test. (**c**) Numbers and sizes of colonies formed by GFP-high and GFP-negative HCC70 cells infected with the K14p-GFP reporter vector (shown in diagram). Values indicate average of five replicates ± s.e.m. 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 ± s.e.m.

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, although 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 Figure 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 (Figures 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 contributes 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, whereas its silencing led to a reproducible increase in EZH2 levels (Figures 7b, d), suggesting a negative-feedback loop between the two genes. Together, these findings reveal that GATA3 performs a role opposite to that of EZH2, acting to decrease bi-lineage identity in favor of a luminally-directed identity. The interplay between the two proteins can therefore determine the composition of the cancer cell population (Figure 7g).

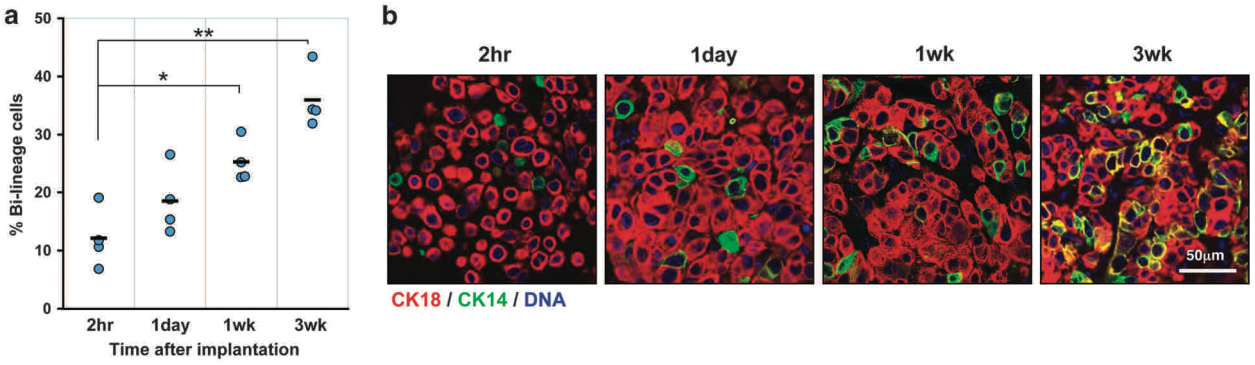



**Figure 6.** Bi-lineage cell numbers increase during tumor formation. (**a**) Percentage of bi-lineage cells in sections of mouse mammary glands implanted with MDA-MB-468 cells and excised 2 hours (2hr), 1 day, 1 week or 3 weeks (1wk, 3wk) after implantation. Each circle represents a single tumor. Black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

## 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, as well as genes specifically expressed in luminal progenitors.[3,4,7] 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 subtype identity 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 on 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 have revealed that bi-lineage cells exist among embryonic mammary stem cells that subsequently give rise to the two mature lineages.[30,31] 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% of cells that co-express 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 increased phenotypic 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 glands, the numbers of bi-lineage cells increase, indicating either that these cells have a growth or survival advantage *in vivo*, or that the 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 tumors, 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 containing 10% FBS and SUM149 cells in F-12 containing 5% FBS, 1μg/ml hydrocortisone and 0.5 μg/ml insulin. For colony-formation assays, cells were passed through a 40-μm strainer, and seeded in 96-well plates coated with growth factor reduced Matrigel (BD Biosciences, Franklin Lakes, NJ, USA), 500 cells per well, 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 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,

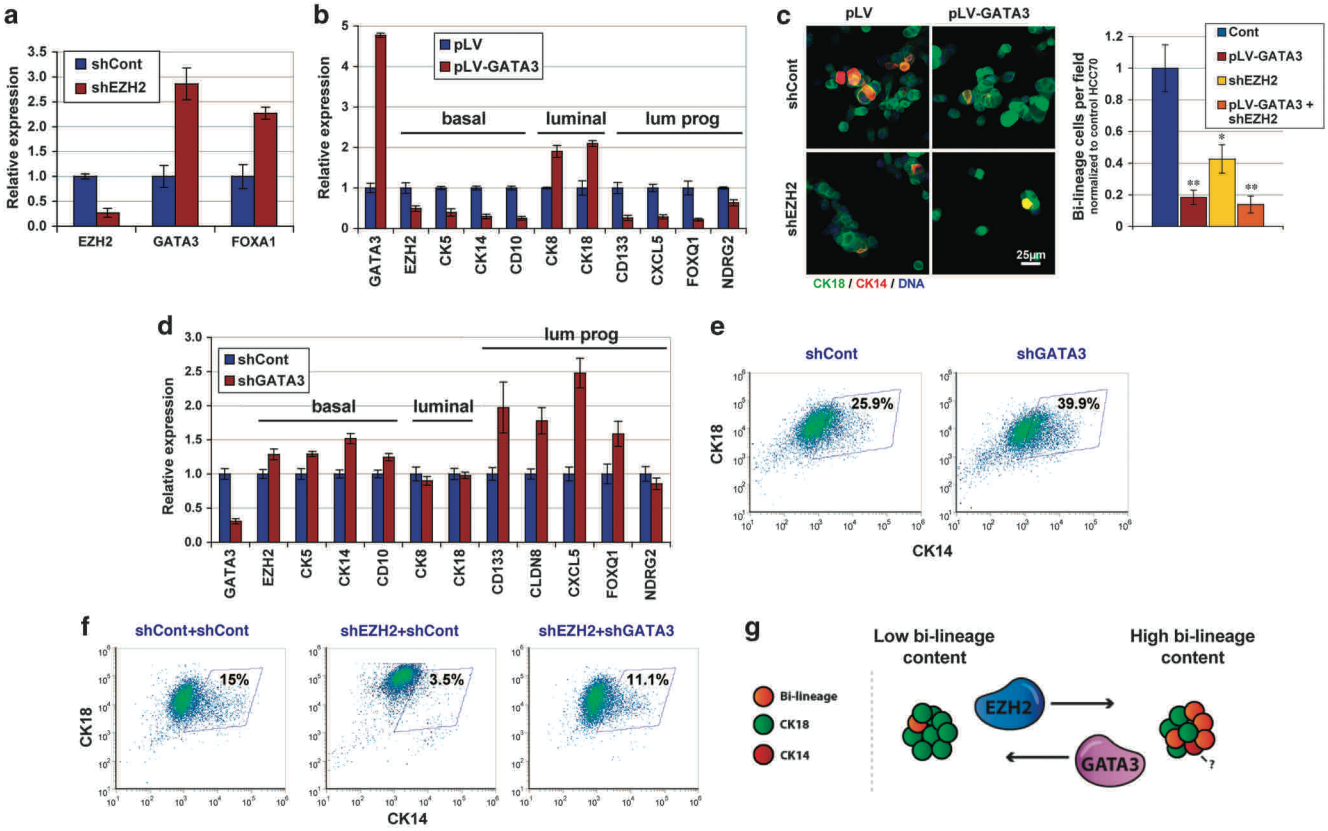



**Figure 7.** GATA3 represses bi-lineage identity. (**a**) Fold change in GATA3 and FOXA1 expression in EZH2-silenced HCC70 cells relative to control cells (shCont), assessed by qRT–PCR. Values indicate average of triplicate reactions ± s.e.m. (**b**) Fold change in expression of basal-lineage, luminal-lineage, 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 seven fields each containing >50 cells, ± s.e.m. $*P<0.01$, $**P<0.002$, *t*-test (**d**) Fold change in expression of basal-lineage, luminal-lineage and luminal progenitor-associated genes upon GATA3 silencing in HCC70 cells. (**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.

and we extrapolated that bi-clonal colonies account to ~3.2%. For passaging, Matrigel was dissolved using cell-recovery solution (BD); colonies were then dissociated with trypsin, passed through a 40-μm strainer and reseeded. The semi-liquid assays were performed as described in Guo *et al.*[29] Phase-contrast images of colonies were obtained using an Eclipse Ti inverted microscope (Nikon, Tokyo, Japan) and a DS-Fi1 camera (Nikon). Colony numbers and sizes were quantified using the NIS Elements software (Nikon). Colonies with a diameter >30 μm and a round morphology were scored. Colonies were stained as described in Debnath *et al.*[42] with antibodies against CK14 (RB-9020, Thermo, Waltham, MA, USA) and CK18 (MS-142, Thermo), followed by labeled secondary antibodies (Jackson, West Grove, PA, USA). Images were collected using a LSM710 laser scanning confocal microscope (Zeiss, Oberkochen, Germany). >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 was used for EZH2 silencing. Targeting sequences used: shEZH2 #1: 5′-ATTCTTGGTTTAAGATTTCCG-3′, shEZH2 #2: 5′-AAGCTAAGG CAGCTGTTTCAG-3′, shEED: 5′-AATTCCTATGTATGTCTCTGG-3′, shGATA3 5′-TTTCGGTTTCTGGTCTGGATG-3′, shCont (scrambled): 5′-CCTAAGGTTAA GTCGCCCTCG-3′. 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-GFP-Blast$^R$ 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, Venlo, Netherlands). Gene expression profiling was done on Human Gene 1.0 ST Arrays (Affymetrix, Santa Clara, CA, USA). 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 Ben-Porath *et al.*[6] with a threshold of 0.3 in $\log_2$ 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 obtained from Lim *et al.*[7] and Prat *et al.*[23] Gene Set Enrichment Analysis (GSEA) was performed as described in Subramanian *et al.*[44]

## Immunohistology, immunocytology and FACS

Immunofluorescent staining of cultured cells and tumor sections for CK14 and CK18 was done using standard procedures. Bi-lineage cell percentages in tumors were calculated as averages of counts of 10 confocal microscopic fields examined in each section, assisted by the CellProfiler software (http://www.cellprofiler.org/). FACS analysis of live cells was done using standard procedures on an LSRII Analyzer (BD), using antibodies against CD133 (AC133, Miltenyi, Gladbach, Germany) or c-KIT (#3308, Cell Signaling, Danvers, MA, USA). For intracellular cytokeratin FACS stains,

cells were fixed in 2% paraformaldehyde for 10 min, permeabilized for 20 min 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. For isolation of bi-lineage and CK18-only cells, we sorted stained cells on FACS-ARIAII (BD). RNA was extracted 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 20μl media containing 25% Matrigel into both #4 mammary glands of 6-week-old female NOD/SCID mice. Tumors were measured by palpation, and weighed upon excision, which was done either 125 days after implantation (Figure 3), or at the time points indicated in Figure 6. Tumors were formalin-fixed and paraffin-embedded for immunohistochemistry. The experiment shown is representative 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. Additional antibodies used: EZH2 (07-689, Millipore, Billerica, MA, USA), CK5 (ab24647, Abcam, Cambridge, UK) and β-actin (sc-1615, Santa Cruz, Santa Cruz, CA, USA). Primer sequences appear in Supplementary Methods.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

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-JARID1Bp-GFP-Blast$^R$ 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.

## 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 USA* 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 Kleer 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 USA* 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, Kleer CG. Increased risk for distant metastasis in patients with familial early-stage breast cancer and high EZH2 expression. *Breast Cancer Res Treat* 2012; **132**: 429–37.
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 USA* 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 USA* 2005; **102**: 15545–15550.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)

**Supplementary Information**

**Supplementary Figure Legends**

**Supplementary Figure 1. EZH2 is overexpressed in basal-like breast cancers. (a)** Expression levels of EZH2 and other indicated genes (rows) across 1,211 breast cancer samples (columns) within a compendium of six published studies (6). Tumors are grouped by intrinsic subtype (indicated above). Red and green indicate over- or under-expression, respectively. *P*-values (right) indicate the significance of difference in expression levels of each gene in basal-like tumors versus tumors of other subtypes (*t*-test). Note that for BMI1 the *P*-value denotes lower expression in basal-like tumors. Histogram (right) shows the average expression levels of EZH2 within each subtype (in $\log_2$, ±s.e.m) relative to its average expression across samples (=0). The number of tumors included in each subtype is listed below. **(b)** Preferential over- (red) or under-expression (green) of indicated gene sets (rows) in the same tumor samples. On right: Pearson correlation (R) between average expression levels of genes in each set and EZH2 levels in the tumor samples. H3K27me3 bound – genes bound by this modified histone in human ES cells (22), ES exp1 – genes enriched in hES cells (6), Luminal, Lum Prog, Myo/MaSC – genes enriched in the mature luminal, luminal progenitors, or the mammary stem cell-containing myoepithelial cell fraction in the normal human breast, respectively (7). **(c)** Percentages of basal (CK5/6-positive) and non-basal (CK5/6-negative) breast cancers showing high (2+3) or low (0+1) EZH2 levels, assessed by immunohistochemical staining of a panel of 284 breast cancer patients treated at Hadassah Medical Center. Right: images of EZH2 and CK5/6 immunostaining in two representative samples. Patient material was collected and used under the approval of Hadassah-Hebrew University Medical Center's Institutional Review Board.

**Supplementary Figure 2. Additional data supporting the role of EZH2 in maintaining basal-like marker expression. (a)** Western blot analysis of indicated proteins in cell lines representing different tumor subtypes, indicated above: Luminal, HER2 (HER2 overexpressing), EMT (possess

mesenchymal traits); Myo (immortalized normal myoepithelial). **(b)** Western blot analysis of HCC70 cells infected with two different shRNA sequences targeting EZH2, an shRNA targeting EED, or a scrambled sequence control shRNA. T47D – luminal cell line as reference. **(c)** Levels of EZH2 silencing achieved with the two shRNA hairpins as assessed by qRT-PCR.

**Supplementary Figure 3. Lineage marker expression in breast cancer cell lines.** Expression patterns of genes (rows) associated with luminal, basal or luminal progenitor cells, as well as with the epithelial to mesenchymal transition (EMT) in a set of previously profiled breast cancer cell lines (columns) (Neve et al., 2006). We grouped the lines (as indicated above) according to their marker expression profiles; red and green indicate over- or under-expression, respectively, of each gene relative to its average expression (black) across all samples. Bottom: in brown, subtypes as defined in Neve et al. Bottom in colors: intrinsic tumor subtype determined using the SSP tool (Hu et al., 2006).

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 2006; 10:515-527.
Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006; **7**:96.

**Supplementary Figure 4. Gene sets associated with tumor subtypes. (a)** Gene sets highly expressed in each of the indicated breast cancer subtypes were obtained from Ref. (23). The heat map presents preferential over- or under-expression (red and green, respectively) of genes within each gene set in each of 1,211 breast tumors (6) (columns). Tumors are grouped by intrinsic subtype, indicated above. The number of genes included in each set is indicated on right. **(b)** Heat map representing preferential over- or under-expression of 12 sets of coordinately expressed genes across the same samples. Each set contains genes associated with different subtypes and functions. The compilation of these gene sets is

described in Supplementary Methods. **(c)** Expression pattern of selected individual genes (assayed in Figure 1f) from the basal-like-associated gene set and the luminal progenitor-associated gene set, across the same breast cancers. CLDN8 and FOXQ1 were omitted due to low representation on the arrays used in the dataset. Red and green indicate over- or under-expression, respectively, relative to average expression (black) of each gene across samples. **(d)** Preferential repression (green) or activation (red) of the indicated gene sets following silencing of EZH2 in HCC70 cells. Values indicate statistical significance of preferential change in gene set expression, shown as $-\log_{10}$ of *P* value. **(e)** Preferential repression (green) or activation (red) of the 12 subtype-associated gene sets following silencing (left) or overexpression (right) of EZH2 in HCC70 cells. The basal-like breast cancer-associated set 9 is enriched for developmental related Gene Ontology functions, while set 1 is enriched for proliferation-related functions **(**see **supplementary File 1)**.

**Supplementary Figure 5. Decrease in the bi-lineage fraction following the silencing of EZH2 or EED. (a)** FACS analysis of CK14 and CK18 in HCC70 cells infected with two different shRNAs targeting EZH2, with shEED or with control shRNA (shCont). **(b)** FACS analysis of CK14 and CK18 in MDA-MB-468 cells infected with shEZH2 or shCont.

**Supplementary Figure 6. 3D Colony morphologies of tested cells, and proliferation and apoptosis rates upon EZH2 manipulation (a)** A representative HCC70 cell colony formed in semi-liquid stem cell growth conditions (left) and a colony formed by SUM149 cells on Matrigel (right). Colonies were stained for CK14 and CK18 as indicated. **(b)** Growth curve in 2D culture of HCC70 cells infected with two different shEZH2 targeting vectors or with shCont. **(c)** Growth curve in 2D culture of SUM149 cells infected with shEZH2 or shCont. **(d)** Growth curve in 2D culture of HCC70 cells infected with EZH2 over-expression or control vector. In all growth curves values indicate average of 10 replicate wells ±s.e.m. **(e)** Percentage of apoptotic cells in HCC70 cells infected with shEZH2 or shCont vectors

as assessed by Annexin V FACS stain. Values indicate average of triplicate reads ±s.e.m. N.S. – non-significant, *t*-test.

**Supplementary Figure 7. K14p-GFP reporter enables isolation of bi-lineage enriched and deprived cell fractions (a)** FACS sorting of HCC70 cells either uninfected (black line), infected with the K14p-GFP virus (green line), or infected with a virus carrying GFP with no upstream promoter (grey line). Grey boxes indicate gates used to isolate GFP-high and GFP-negative fractions of CK14p-GFP infected cells. The no-promoter line served as a negative control indicating levels of spurious GFP expression. **(b)** FACS sorting of HCC70 cells infected with the pRRL-GFP vector. Data is presented as GFP vs. FSC (forward scatter) due to high levels of GFP expression observed with this vector. Gray boxes indicate isolated populations. **(c)** Immunofluorescent co-staining of CK14 (red), CK18 (white) and GFP (green) in CK14p-GFP-infected HCC70 cells sorted for GFP-high (GFP+) or GFP-negative (GFP-). Images show correlation between CK14 expression and GFP expression, and enrichment for CK14+CK18+ cells in the GFP+ population. **(d)** Expression levels of indicated genes in GFP-high (GFP+) or GFP-negative (GFP-) sorted populations of K14p-GFP-infected HCC70 cells, assessed by qRT-PCR. Changes show higher expression levels of basal markers in the GFP+ fraction, indicative of enrichment for bi-lineage cells in this fraction. Values indicate average of triplicate reactions ±s.e.m.

**Supplementary Figure 8. EZH2 and GATA3 display inverse expression patterns in breast cancers. (a)** Average expression levels of EZH2 and GATA3 in 1211 breast cancers (6) grouped by subtype. Values are in $\log_2$ relative to mean expression of each gene (=0). Data for EZH2 is same as in supplementary Figure 1. Values indicate average expression in tumors within subtype ±s.e.m. **(b)** Scatter plot showing the expression of EZH2 versus that of GATA3 in the same 1,211 breast cancers. Each dot represents an individual tumor; tumors of the basal-like subtype are shown in red, tumors of

other subtypes are in blue. Line indicates linear regression of data. Pearson correlation R-value and corresponding *P*-value are displayed. **(c)** Relative expression levels of EZH2 and GATA3 in breast cancer cell lines of indicated subtypes assessed by qRT-PCR. Values indicate average of triplicate reactions ±s.e.m. **(d)** Images of human breast cancer sections stained for GATA3. GATA3 expression is observed in basal-like tumor (right), at lower levels than in luminal tumor (left).

**Supplementary File 1. Gene sets used in study. Sheet 1:** Breast cancer subtype-associated gene sets obtained from Ref. 23. **Sheet 2:** Luminal progenitor-associated gene set obtained from Ref. 7. **Sheet 3:** Breast cancer subtype-associated #1-12 gene sets. **Sheet 4:** Gene Ontology category enrichments of subtype associated 12 gene sets. **Sheet 5:** Changes in expression (in $\log_2$) of gene sets of interest in different samples. Shown are genes from within subset that are changed in at least one of the four samples.

**Supplementary Methods**

**Compilation of subtype-associated gene sets**

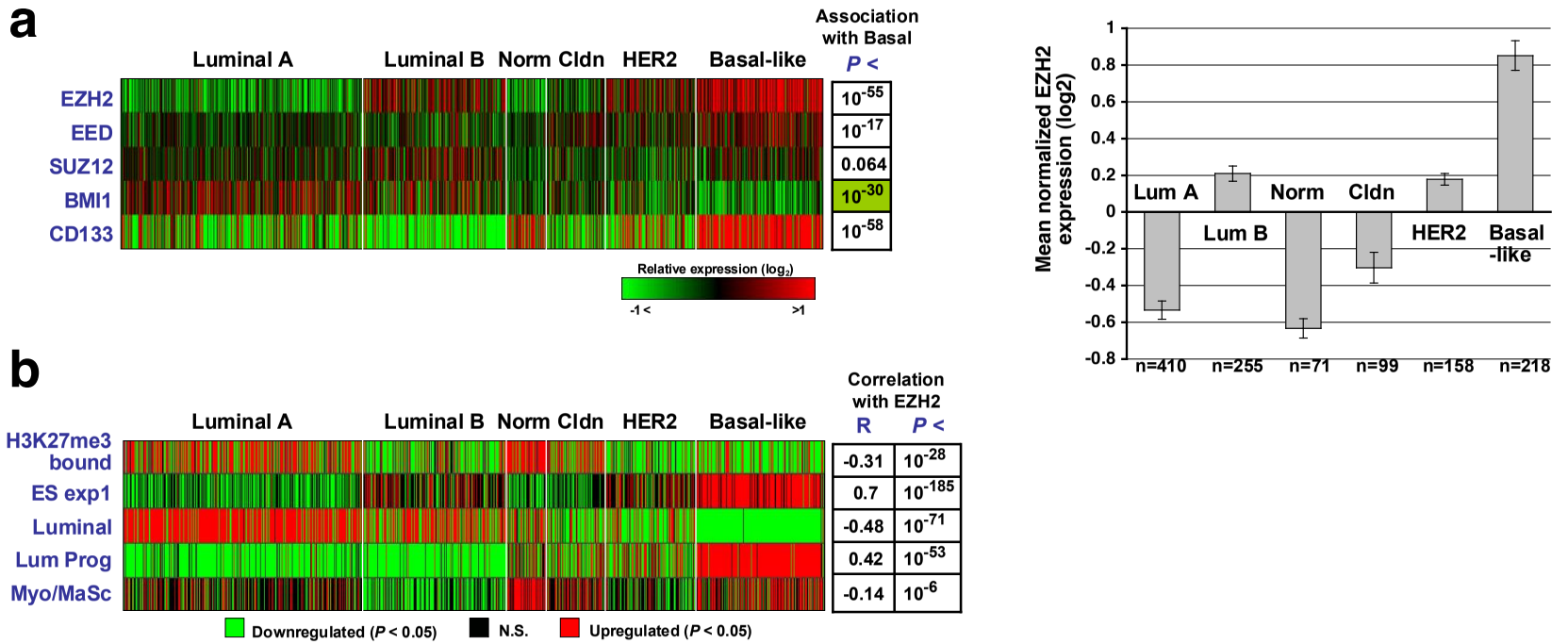
We compiled 12 sets of genes coordinately expressed in human breast cancers and associated with different subtypes. First, we identified genes discriminating between any two tumor categories within several published expression datasets (6), using the Wilcoxon signed-rank statistical test as a measure of association. All genes showing at least one high association score ($P<10^{-4}$) were clustered using the K-means method into gene sets based on their pattern of association across tumor categories. The 12 resulting gene sets were tested for their functional enrichments in Gene Ontology.

**Primers used**

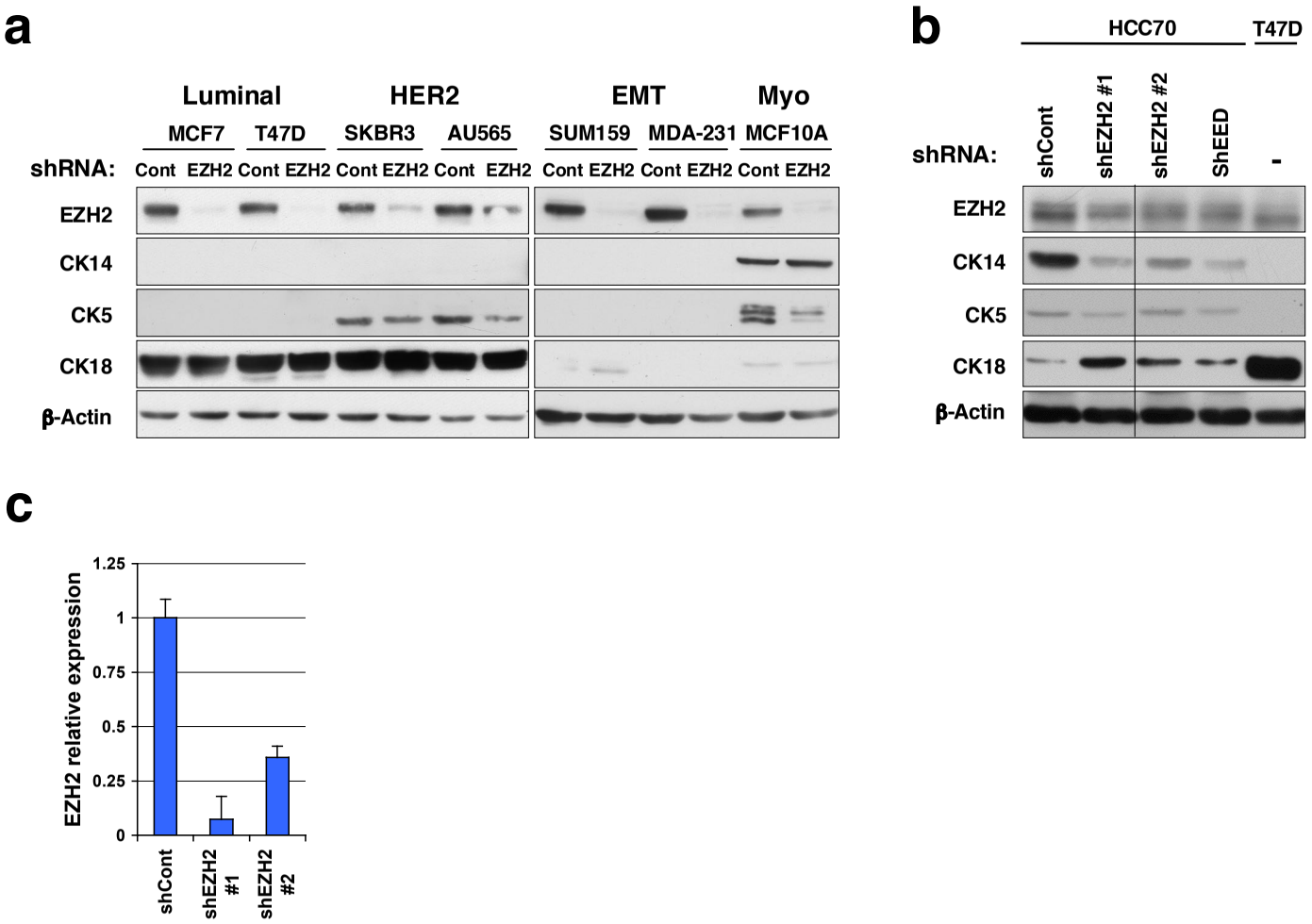
| *Gene* | *Forward (5'-3')* | *Reverse (5'-3')* |
|---|---|---|
| **EZH2** | **GCGCGGGACGAAGAATAATCAT** | **TACACGCTTCCGCCAACAAACT** |
| **EED** | **GAGGAATATGTCCGAGAGGG** | **GACGAGAATGATGACGCTGT** |
| **HPRT1** | **TGACACTGGCAAAACAATGCA** | **GGTCCTTTTCACCAGCAAGCT** |
| **GATA3** | **CTCATTAAGCCCAAGCGAAG** | **TCCTCCAGAGTGTGGTTGTG** |
| **FOXA1** | **AGACACGCAGGAGGCCTACTC** | **CATGTTGCCGCTCGTAGTCA** |
| **HOXA9** | **TGCAGCTTCCAGTCCAAGG** | **GTAGGGGTGGTGGTGATGGT** |
| **ALDH1** | **CGCAAGACAGGCTTTTCAG** | **TGTATAATAGTCGCCCCCTCTC** |
| **CK19** | **CTACACGACCATCCAGGACC** | **GAGCCTGTTCCGTCTCAAAC** |
| **CK18** | **GAGGCTGAGATCGCCACCTA** | **CCAAGGCATCACCAAGATTAAAG** |
| **CK17** | **AATTGAGGAGCTGCAGAACAA** | **AAACTTGGTGCGGAAGTCAT** |
| **CK14** | **GACCATTGAGGACCTGAGGA** | **CATACTTGGTGCGGAAGTCA** |
| **CK8** | **CCGTGGTTGTGAAGAAGATC** | **TGTTCACTTGGGCAGGAC** |
| **CK6a** | **CTGAGATCGACCACGTCAAG** | **CAGCTTGTTCTTGGCATCCT** |
| **CK5** | **CTGCGTGAGTACCAGGAGC** | **CTGGTCCAACTCCTTCTCCA** |
| **CDH3** | **TTCAGGGAGGCTGAAGTGAC** | **CAGTGCTAAACAGAGCTGGC** |
| **MME** | **TGGGTTCTTGAAGGACATCTTTC** | **CGTTACGGCAACTTTGACATTTT** |
| **SLUG** | **GCATTTCTTCACTCCGAAGC** | **TGAATTCCATGCTCTTGCAG** |
| **ARNTL2** | **CAAGTGGCTCCTGCGATG** | **AGCTGTTGGTCTTGTCCCTG** |
| **ETV6** | **GCCCAGTGCCGAGTTACGCT** | **AAATTGGCTGCAAGCGCAGGTG** |
| **FOXC1** | **TAAGCCCATGAATCAGCCG** | **GCCGCACAGTCCCATCTCT** |
| **MMP7** | **GTGAGCTACAGTGGGAACAGG** | **CATCTCCTTGAGTTTGGCTTCT** |
| **PIM1** | **AACTGGTCTTCCTTTTTGGTT** | **TACCATGCCAACTGTACACAC** |
| **SLC34A2** | **CAACATCTCCGGCATCTTGCTGTG** | **GGAGTCGGAGGCACAGTACCAGGA** |
| **VGLL1** | **GGCTCAGTTCACTATAAGAA** | **TATTTCCCAGGTGTCTCTAA** |
| **CLDN8** | **CACTCTCCACTGAGGCATGA** | **TCCCAGAGGATAATGGCAAC** |
| **FOXQ1** | | |

| | | |
|---|---|---|
| **NDRG2** | **GAGATATGCTCTTAACCACCC** | **GCTGCCCAATCCATCCAA** |
| **CXCL5** | **AGGTGGAAGTGGTAGCCTCC** | **TTGGACGGTGGAAACAAG** |
| **CD133** | **ACCGACTGAGACCCAACATC** | **TGAACAGCACCTTGAAGAGCT** |
| **DSC2** | **CTGCTCCTGCTGACCCTC** | **CCAACAAGTTTCTCGGCATC** |
| **NFE2L3** | **GGGAAGAGGAGAAGGCACC** | **ATCCACAGCTTCGTGCTTTT** |
| **PARP1** | **CCCAGGGTCTTCGGATAG** | **AGCGTGCTTCAGTTCATACA** |
| **ITGB8** | **TGGTCGAGGAGTTTGTGTTTG** | **AGCCACTGAAGCATTGGCA** |
| **TRIM29** | **CATAGCACCGTGACAGTGGA** | **GCCACTTCTCAGCTTCATCC** |
| **NFIB** | **TTTGTGTCCAGCCACATCAT** | **GTGGCTTGGACTTCCTGATT** |

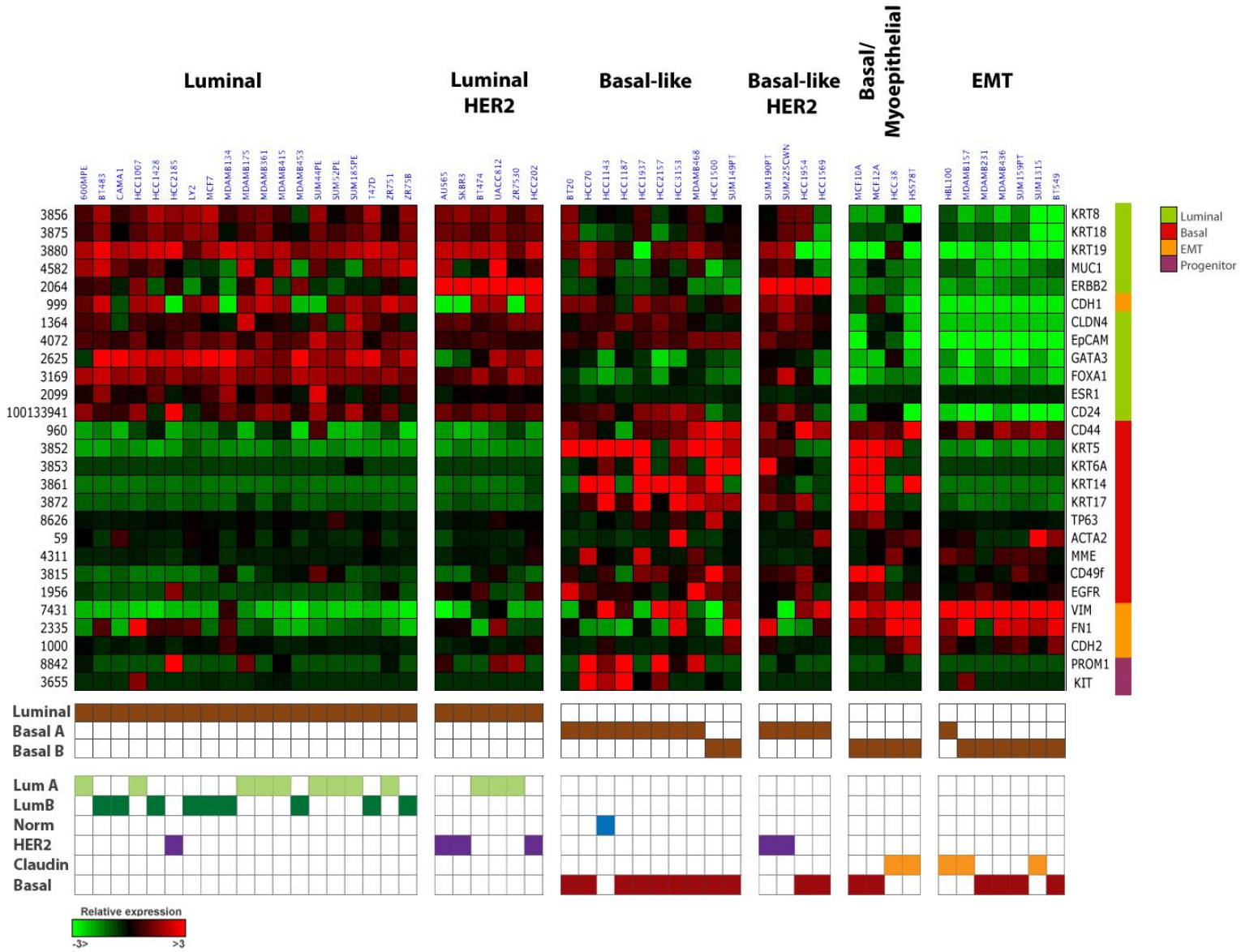
# Figure S1


a
Luminal A
Luminal B
Norm Cldn
HER2
Basal-like
Association with Basal
P <
EZH2
EED
SUZ12
BMI1
CD133
10-55
10-17
0.064
10-30
10-58
Relative expression (log2)
-1 <
>1
Mean normalized EZH2 expression (log2)
Lum A
Lum B
Norm
Cldn
HER2
Basal -like
n=410
n=255
n=71
n=99
n=158
n=218
b
Correlation with EZH2
R
P <
H3K27me3 bound
ES exp1
Luminal
Lum Prog
Myo/MaSc
-0.31
10-28
0.7
10-185
-0.48
10-71
0.42
10-53
-0.14
10-6
Downregulated (P < 0.05)
N.S.
Upregulated (P < 0.05)
c
EZH2 0+1
EZH2 2+3
P < 10-5
Percentage of tumors
CK5/6 positive
CK5/6 negative
n=20
n=264
EZH2
CK5/6
Sample # 1
Basal-like
Sample # 2
Non Basal-like

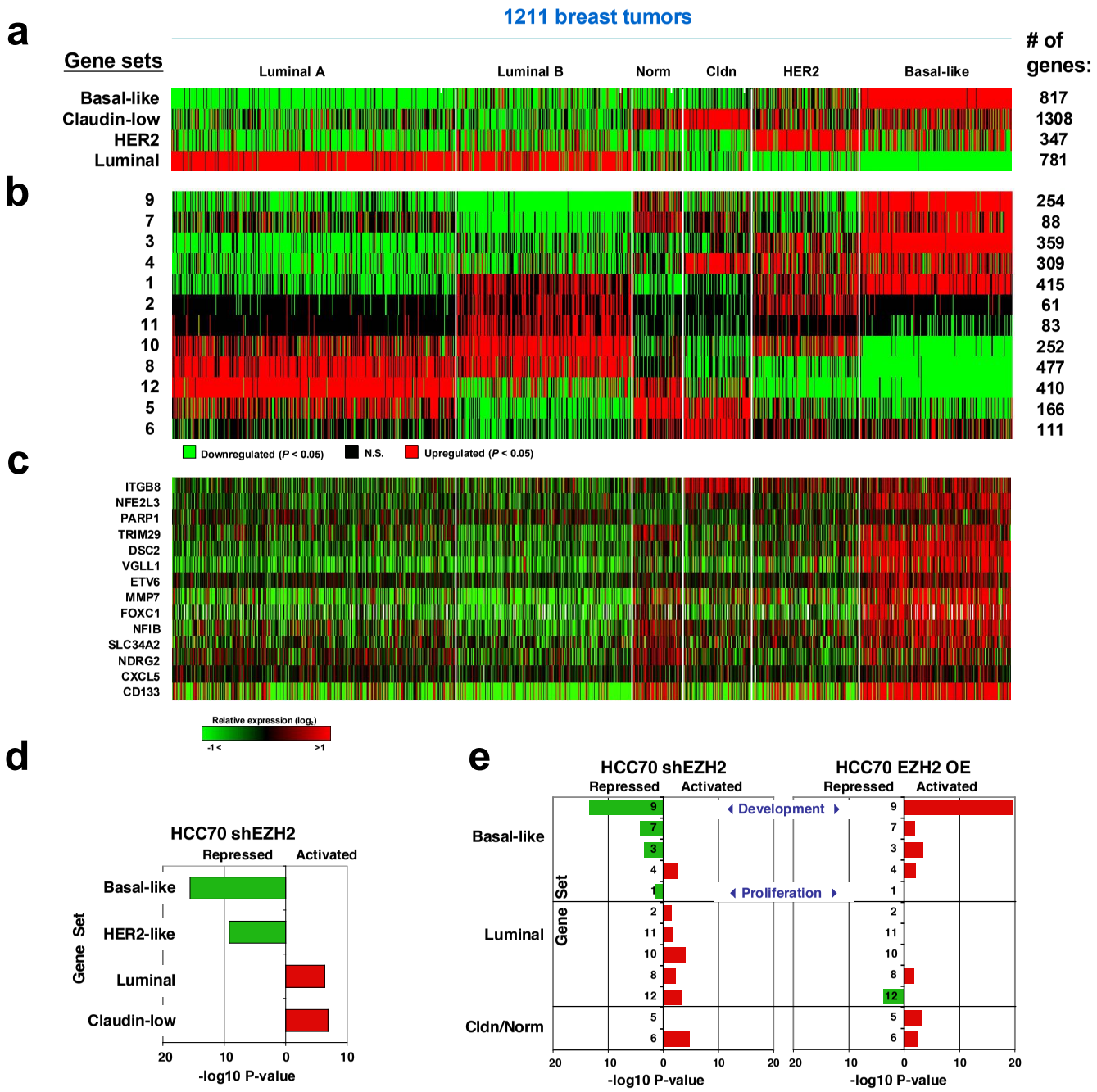
# Figure S2


a
Luminal
HER2
EMT
Myo
MCF7
T47D
SKBR3
AU565
SUM159
MDA-231
MCF10A
shRNA:
Cont EZH2 Cont EZH2 Cont EZH2 Cont EZH2 Cont EZH2 Cont EZH2 Cont EZH2
EZH2
CK14
CK5
CK18
β-Actin
b
HCC70
T47D
shRNA:
shCont
shEZH2 #1
shEZH2 #2
ShEED
-
EZH2
CK14
CK5
CK18
β-Actin
c
EZH2 relative expression
1.25
1
0.75
0.5
0.25
0
shCont
shEZH2 #1
shEZH2 #2

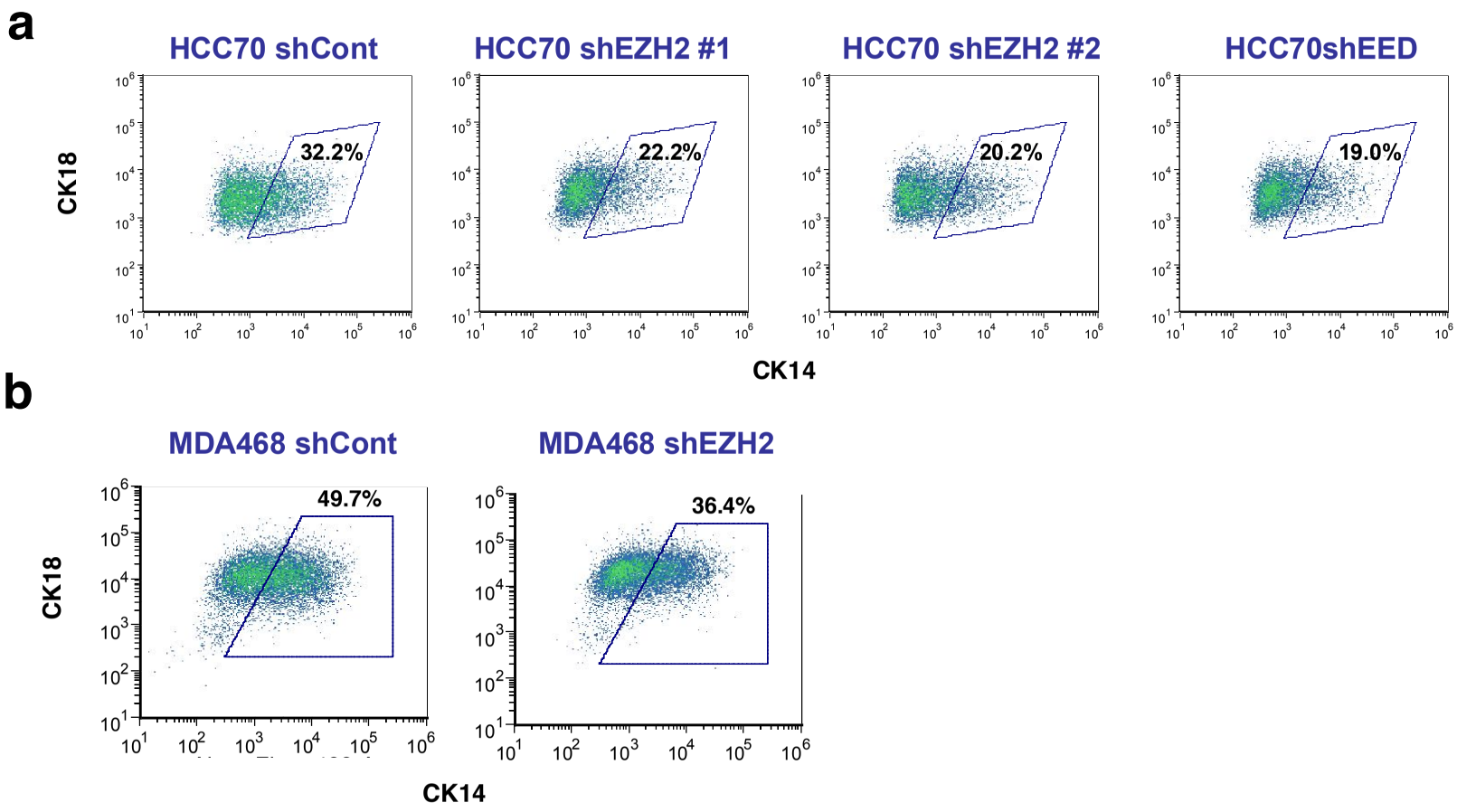
# Figure S3


Luminal
Luminal HER2
Basal-like
Basal-like HER2
Basal/ Myoepithelial
EMT
600MPE
BT483
CAMA1
HCC1007
HCC1428
HCC2185
LY2
MCF7
MDAMB134
MDAMB175
MDAMB361
MDAMB415
MDAMB453
SUM44PE
SUM52PE
SUM185PE
T47D
ZR751
ZR75B
AU565
SKBR3
BT474
UACC812
ZR7530
HCC202
BT20
HCC70
HCC1143
HCC1187
HCC1937
HCC2157
HCC3153
MDAMB468
HCC1500
SUM149PT
SUM190PT
SUM225CWN
HCC1954
HCC1569
MCF10A
MCF12A
HCC38
HS578T
HBL100
MDAMB157
MDAMB231
MDAMB436
SUM159PT
SUM1315
BT549
3856
3875
3880
4582
2064
999
1364
4072
2625
3169
2099
100133941
960
3852
3853
3861
3872
8626
59
4311
3815
1956
7431
2335
1000
8842
3655
KRT8
KRT18
KRT19
MUC1
ERBB2
CDH1
CLDN4
EpCAM
GATA3
FOXA1
ESR1
CD24
CD44
KRT5
KRT6A
KRT14
KRT17
TP63
ACTA2
MME
CD49f
EGFR
VIM
FN1
CDH2
PROM1
KIT
Luminal
Basal
EMT
Progenitor
Luminal
Basal A
Basal B
Lum A
LumB
Norm
HER2
Claudin
Basal
Relative expression
-3>
>3

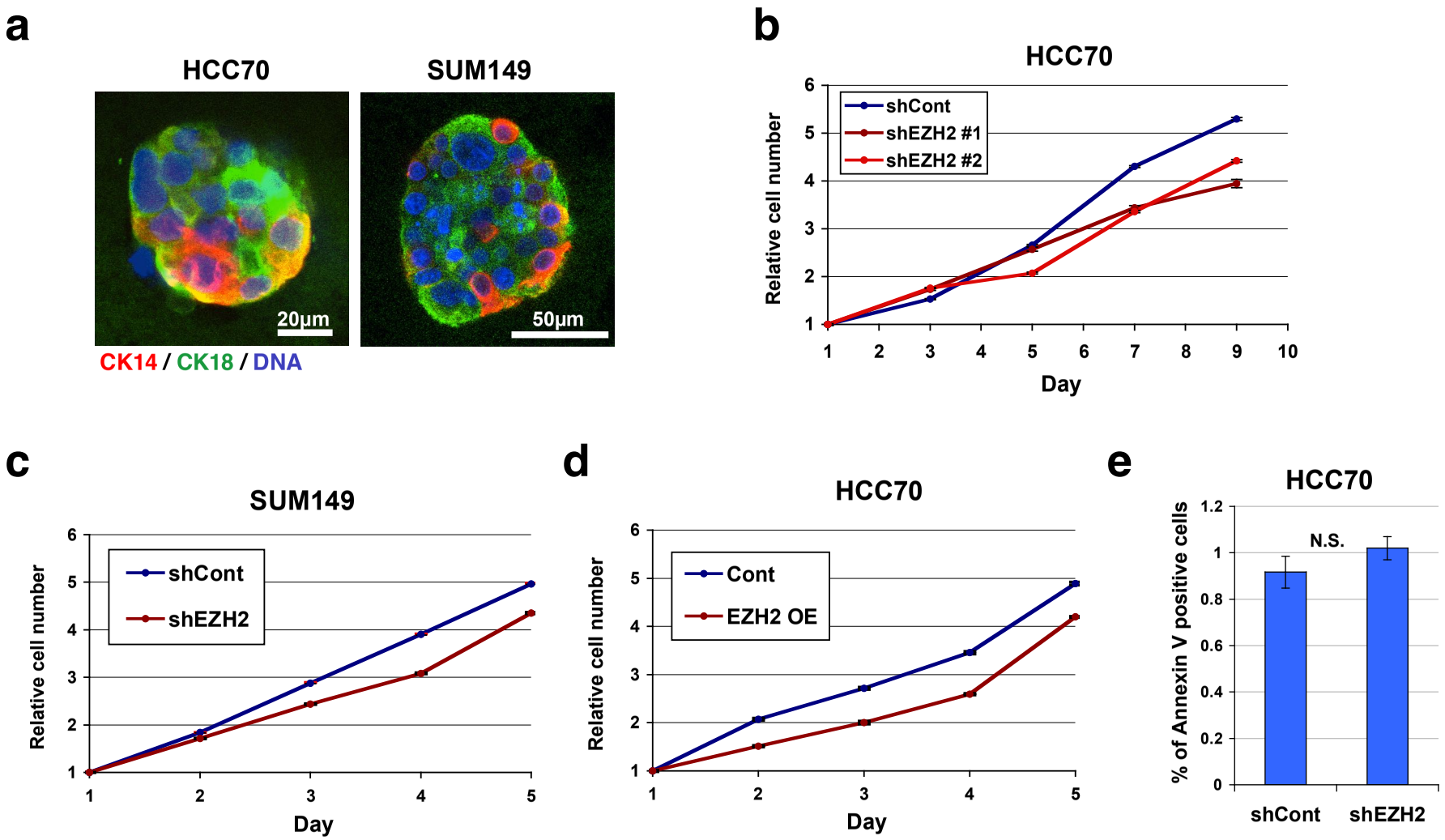
# Figure S4


a
1211 breast tumors
Gene sets
Luminal A
Luminal B
Norm
Cldn
HER2
Basal-like
# of genes:
Basal-like 817
Claudin-low 1308
HER2 347
Luminal 781
b
9 254
7 88
3 359
4 309
1 415
2 61
11 83
10 252
8 477
12 410
5 166
6 111
Downregulated (P < 0.05)
N.S.
Upregulated (P < 0.05)
c
ITGB8
NFE2L3
PARP1
TRIM29
DSC2
VGLL1
ETV6
MMP7
FOXC1
NFIB
SLC34A2
NDRG2
CXCL5
CD133
Relative expression (log2)
-1 <
>1
d
HCC70 shEZH2
Repressed
Activated
Basal-like
HER2-like
Luminal
Claudin-low
Gene Set
20 10 0 10
-log10 P-value
e
HCC70 shEZH2
HCC70 EZH2 OE
Repressed
Activated
Development
Proliferation
Basal-like
Luminal
Cldn/Norm
Gene Set
-log10 P-value

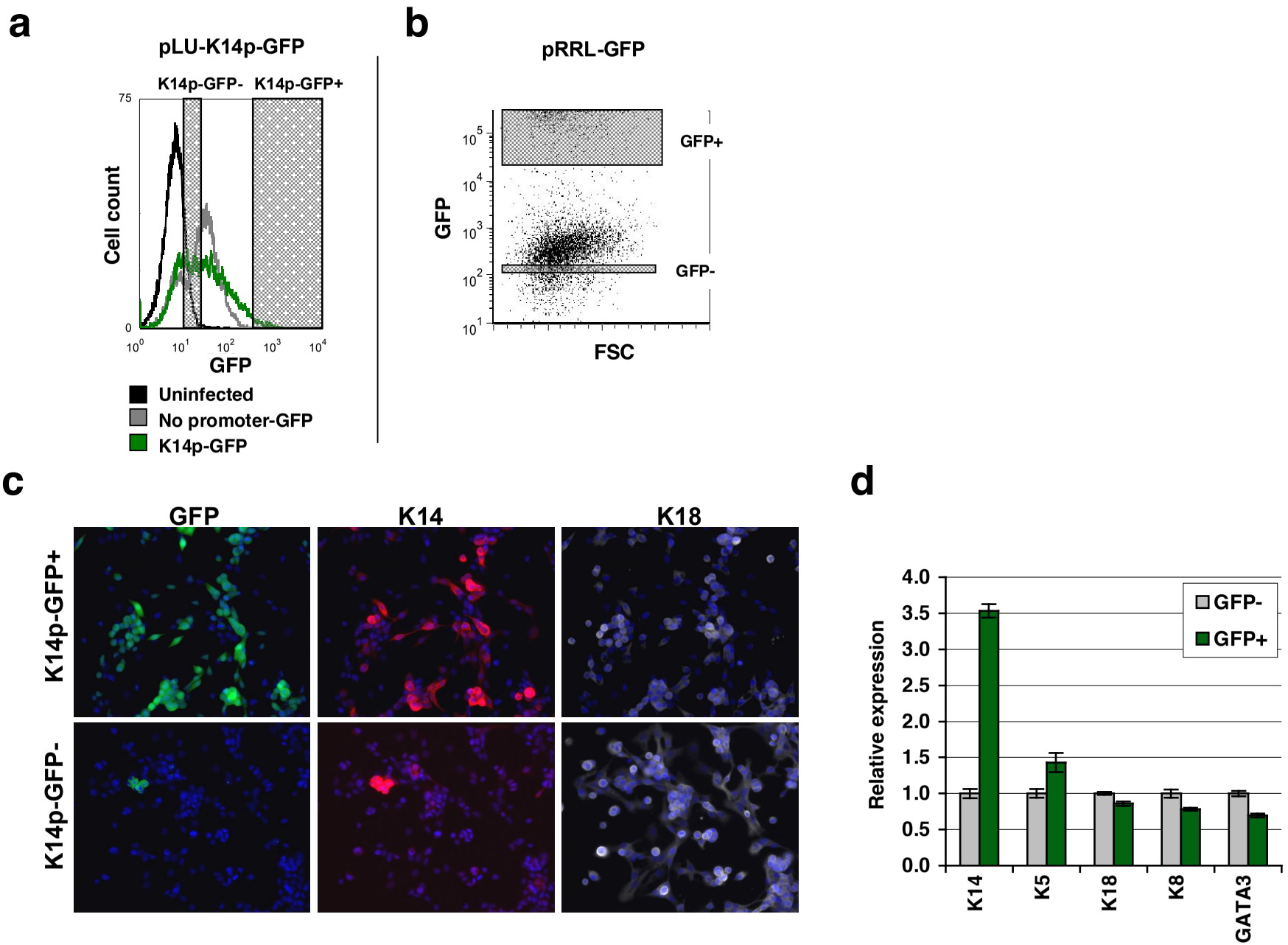
# Figure S5


a
HCC70 shCont
HCC70 shEZH2 #1
HCC70 shEZH2 #2
HCC70shEED
32.2%
22.2%
20.2%
19.0%
CK18
CK14
b
MDA468 shCont
MDA468 shEZH2
49.7%
36.4%
CK18
CK14

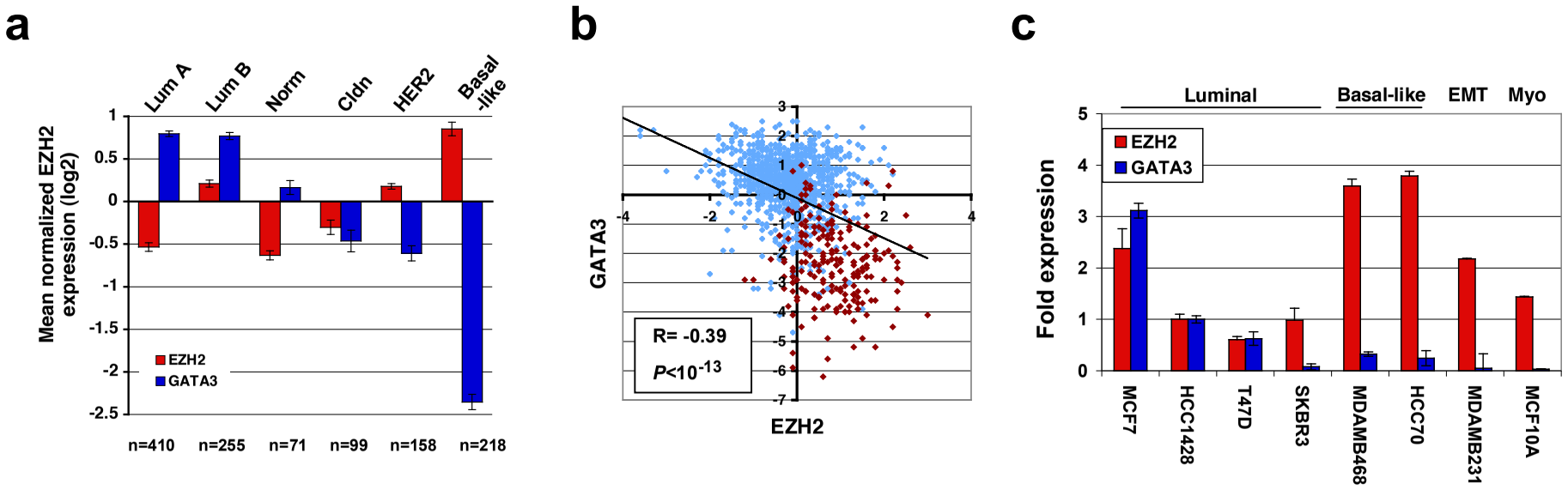
# Figure S6


a
HCC70
SUM149
20μm
50μm
CK14 / CK18 / DNA
b
HCC70
shCont
shEZH2 #1
shEZH2 #2
Relative cell number
Day
c
SUM149
shCont
shEZH2
Relative cell number
Day
d
HCC70
Cont
EZH2 OE
Relative cell number
Day
e
HCC70
N.S.
% of Annexin V positive cells
shCont
shEZH2

# Figure S7


a
pLU-K14p-GFP
K14p-GFP-
K14p-GFP+
75
0
Cell count
$10^0$ $10^1$ $10^2$ $10^3$ $10^4$
GFP
Uninfected
No promoter-GFP
K14p-GFP
b
pRRL-GFP
$10^5$
$10^4$
$10^3$
$10^2$
$10^1$
GFP
GFP+
GFP-
FSC
c
GFP
K14
K18
K14p-GFP+
K14p-GFP-
d
Relative expression
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
GFP-
GFP+
K14
K5
K18
K8
GATA3

# Figure S8


a
Lum A
Lum B
Norm
Cldn
HER2
Basal -like
Mean normalized EZH2 expression (log2)
EZH2
GATA3
n=410
n=255
n=71
n=99
n=158
n=218
b
GATA3
EZH2
R= -0.39
P<10-13
c
Luminal
Basal-like
EMT
Myo
Fold expression
EZH2
GATA3
MCF7
HCC1428
T47D
SKBR3
MDAMB468
HCC70
MDAMB231
MCF10A
d
ER+ PR+ HER2-
Basal-like triple-negative
GATA3

# Chapter 2 *:* Regulation of differentiation transitions through asymmetric divisions in basal-like breast cancer

Status: submitted to Nature Cell Biology, in review (2016)

R.Z. Granit, E Carmon, Y Fixler, S Dalin, R Moran, T Sella, A Sonnenblick, D Katz, U Lehmann, K Paz, F Piccioni, A Regev, D.E. Root and I Ben-Porath

**Regulation of differentiation transitions through asymmetric divisions in basal-like breast cancer**

Roy Zvi Granit[1], Einat Carmon[2], Yaakov Fixler[1], Simona Dalin[1], Rimmon Moran[1], Tamar Sella[3], Amir Sonnenblick[4], Daniela Katz[4], Ulrich Lehmann[5], Keren Paz[6], Federica Piccioni[7], Aviv Regev[7,8], David E. Root[7] and Ittai Ben-Porath[1,*]

[1] Department of Developmental Biology and Cancer Research, Institute for Medical Research – Israel-Canada, The Hebrew University–Hadassah Medical School, Jerusalem, Israel

[2] Department of Surgery, Hadassah–Hebrew University Medical Center, Jerusalem, Israel

[3] Department of Radiology, Hadassah–Hebrew University Medical Center, Jerusalem, Israel

[4] Sharett Institute of Oncology, Hadassah–Hebrew University Medical Center, Jerusalem, Israel

[5] Institute of Pathology, Medizinische Hochschule Hannover, Hannover, Germany

[6] Champions Oncology Inc., Baltimore, MD, USA

[7] Broad Institute of MIT and Harvard, Cambridge, MA, USA

[8] Howard Hughes Medical Institute and David H. Koch Institute of Integrative Cancer Biology, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

*** Correspondence:** Ittai Ben-Porath, E-mail: ittaibp@mail.huji.ac.il

**Keywords:** basal-like breast cancer, differentiation, asymmetric divisions, mammary gland, EZH2, FOXA1, Notch

**Running title:** Regulation of differentiation transitions in breast cancer

**Abstract**

Differentiation events contribute to cellular heterogeneity within tumors and influence disease progression and response to therapy. Here we dissect the mechanisms controlling intratumoral heterogeneity within basal-like breast cancers. We show that cancer cells can transition between a differentiation state related to that of normal luminal progenitors and a state closer to that of mature luminal cells, and that this occurs through asymmetric cell divisions. The Polycomb factor EZH2 and the Notch pathway act to increase the rates of symmetric divisions that produce progenitor-like cells, while the FOXA1 transcription factor promotes asymmetric divisions that reduce the numbers of such cells. Through functional screening, we identified a group of regulators that control cancer cell differentiation state and the relative proportions of tumor cell subpopulations. Our findings highlight the regulation of asymmetric cell divisions as a mechanism controlling intratumoral heterogeneity, and identify molecular pathways that control breast cancer cellular composition.

## Introduction

Cellular heterogeneity is observed in many tumor types and represents a major hurdle for effective therapy[1,2]. Changes in differentiation state are thought to contribute to intratumoral heterogeneity. In breast cancer, subpopulations of stem-like tumor cells possessing self-renewal and differentiation capacity have been identified[3-5]. However, recent studies suggest that, rather than following a unidirectional differentiation hierarchy, cancer cells may be able to transition between stem-like and differentiated identities, and can also adopt intermediate states[2,6-8]. In some cases cancer cell populations can reach an equilibrium between phenotypically different subpopulations, suggesting the existence of forces controlling the overall composition of the tumor[9,10].

Differentiation transitions in breast cancers may occur along various axes[11]. One of these is the epithelial to mesenchymal transition (EMT). It is thought that bi-directionality of EMT, and intermediate epithelial-mesenchymal states, are important for the roles of this axis in metastasis and stemness[12-14]. Lineage differentiation programs active in the normal mammary gland also influence breast cancer cell differentiation state[11,15,16]. Normal mammary cells can differentiate along an axis leading from a basal identity (associated with mammary stem cell potential), to a luminal progenitor state and, in turn, to a mature luminal state[15]. These differentiation states are reflected in breast cancer subtypes: luminal, estrogen receptor (ER)-positive tumors express a gene signature associated with mature differentiated luminal cells, while basal-like breast cancers, which largely overlap with the triple-negative pathological group, express the signature of luminal progenitor cells, and likely originate from them[17-19].

Basal-like tumor cells may thus have a higher level of plasticity allowing them to transition between basal, progenitor and luminal states. A hallmark of these tumors is the expression of

basal keratins such as K5 and K14, usually in a heterogeneous pattern, and often together with luminal keratins[16,20], consistent with the notion of such plasticity. These tumors are also enriched for cells in a partially mesenchymal state[14]. A detailed delineation of the differentiation states of subpopulations within basal-like tumors is, however, currently lacking. In addition, while various regulators of normal mammary differentiation have been implicated in controlling the differentiation state of breast cancers cells[5,11,15], there is little understanding of how the activity of these regulators generates intratumoral phenotypic heterogeneity and influences the overall cellular composition of tumors.

Here we study lineage-associated differentiation transitions within basal-like tumors and the manner by which they give rise to cellular heterogeneity. We show that basal-like tumors harbor cell subpopulations displaying an enhanced progenitor-like identity, as well as cells displaying a more mature luminal state. Tumor cells can transition between these states through asymmetric divisions, whose rates determine the overall composition of the tumor cell population. We identify regulators that control the rates of asymmetric divisions and the proportions of cancer cell subpopulations, and demonstrate that tumor cell composition can be modulated by pharmacologic inhibition of specific pathways.

## Results

### Heterogeneity in differentiation state of basal-like breast cancer cells

We examined the degree of cellular heterogeneity in a collection of triple-negative breast cancers (**Supplementary Table 1**) by co-staining for the luminal cytokeratin K18, the basal cytokeratin K14, and the mesenchymal marker vimentin (VIM). As expected, tumors exhibited heterogeneous staining patterns, with cells, often in close proximity, displaying highly differential expression levels of these markers (**Fig. 1a,b**). Nearly all tumors contained a mixture

of cells expressing K18 or K14, and most tumors contained cells that co-expressed both of these markers ($K18^+K14^+$ cells) (**Fig. 1a,b**). Vimentin-expressing carcinoma cells were also detected in most tumors, and these cells often co-expressed K18, but rarely K14 (**Fig. 1a,b**).

This heterogeneity was observed also in basal-like cell lines, specifically those maintaining epithelial identity and the overall molecular characteristics of basal-like breast cancers (termed "Basal A")[21,22]. FACS analysis of K18, K14 and vimentin expression showed three prominent subpopulations in these lines: $K18^+$, $K18^+K14^+$ and $K18^+VIM^+$ (**Fig. 1c,d**). Cell lines representing other disease subtypes were more homogenous, containing either $K18^+$ cells (luminal and $HER2^+$ lines), $K18^-K14^+$ cells (normal basal/myoepithelial lines) or $K18^-VIM^+$ cells (mesenchymal, "Basal B" lines) (**Fig. 1c,d**).

To assess whether the cell subpopulations found in basal-like breast cancer lines represent distinguishable differentiation states, we sorted $K18^+$, $K18^+K14^+$ and $K18^+VIM^+$ cells from the HCC70 and MDA-MB-468 lines, and analyzed their transcription profiles (**Fig. 1e**). Interestingly, the expression profiles of the three subpopulations were distinct from one another, but were similar between the two cell lines (**Supplementary Fig. 1a**). We next analyzed the level of expression in each subpopulation of previously established gene signatures of normal human mammary cell types[17]. These included the signatures of mature luminal cells, luminal progenitor cells, and basal cells (which are known to be enriched for mammary stem cells, MaSCs), as well as the signature of mammary cells that have undergone an EMT[11,23] **(Supplementary Table 2)**. To do this we assessed the correlation between the up- and down-regulated genes in the expression profile of each subpopulation and each of these signatures. We found that the expression profile of $K18^+$ cells showed the highest correlation with the signature of normal mature luminal cells, while that of the $K18^+K14^+$ subpopulation showed the highest correlation

with the signature of luminal progenitors (**Fig. 1f**, **Supplementary Fig. 1b** and **Supplementary Table 3**). The $K18^+VIM^+$ subpopulation showed the highest correlation with the EMT signature (**Fig. 1f**), consistent with these cells possessing a partially mesenchymal identity. The profile of $K18^+VIM^+$ cells was also correlated with the basal/MaSC signature, yet to a lower degree than to the EMT signature, likely reflecting the partially mesenchymal nature of normal basal cells[24]. This analysis indicates that the basal-like breast cancer cell population contains cells that exist in distinct states: $K18^+VIM^+$ cells are partially mesenchymal, $K18^+$ cells possess a relatively differentiated luminal identity, and $K18^+K14^+$ cells are in a state related to that of luminal progenitors.

To further establish the link between the $K18^+K14^+$ profile and the luminal progenitor state, we isolated $K18^+$ and $K18^+K14^+$ cells from human triple-negative tumors grown as patient derived xenografts (PDXs) in immunocompromised mice, and tested by qRT-PCR the expression levels of representative genes from the luminal progenitor signature. $K18^+K14^+$ cells from two independently isolated PDXs expressed higher levels of progenitor-signature genes than $K18^+$ cells (**Fig. 1g** and **Supplementary Fig. 1c**). Together these analyses indicate that cells in basal-like breast cancers can display distinct identities along the mesenchymal–epithelial and basal–luminal differentiation axes, and that the $K18^+K14^+$ profile indicates an increased molecular link to luminal progenitor cells.

**$K18^+K14^+$ cells possess enhanced tumorigenic traits**

We next assessed whether the $K18^+K14^+$ profile is indicative of tumorigenic properties distinct from those of $K18^+$ cells. Relying on the finding that all $K14^+$ cells in basal-like lines also express K18 (**Fig. 1d**), we introduced a reporter vector expressing GFP under the control of the K14 promoter[25] into basal-like HCC70. We then isolated $GFP^{high}$ cells, enriched for the

$K18^+K14^+$ population, as well as $GFP^-$ cells (**Supplementary Fig. 2**). Transplantation of these cells in limiting dilutions in mouse mammary glands revealed that the $K18^+K14^+$-enriched cell fraction had an 11-fold higher tumor initiation capacity, and gave rise to faster growing tumors, compared to the $GFP^-$ fraction (**Fig. 2a,b**). To test the representation of $K18^+K14^+$ cells in a spontaneous metastasis model, we injected unsorted MDA-MB-468 cells, labelled with GFP for detection purposes, to mouse mammary glands. We found that in spontaneous lung metastases formed by these tumor xenografts the percentage of $K18^+K14^+$ cells was substantially higher than in the primary tumors, such that 95% of seeded cells displayed this profile (**Fig. 2c**). We also injected unsorted MDA-MB-468 cells into mouse tail veins for direct metastatic seeding in the lungs, and found that the percentage of $K18^+K14^+$ cells among seeded cells rose from 43% two hours after injection to 98% ten days later (**Fig. 2d**), suggesting that this state is either selected for or adopted during colonization. Together these results indicate that the $K18^+K14^+$ cell subpopulation possesses enhanced tumor seeding and growth capabilities, and a potential advantage during metastasis formation.

**Asymmetric divisions contribute to differentiation transitions**

We next asked whether transitions of tumor cells between differentiation states, underlying the observed heterogeneity, could be directly detected. Asymmetric cell divisions, in which the two daughter cells differ in phenotype from each other, provide a mechanism through which such transitions could take place[26]. To test whether asymmetric divisions occur in basal-like breast cancer cell populations, we treated sparsely seeded HCC70 cells with BrdU for one hour to label cells in S-phase, and fixed the cells 18 hours later to allow generation of BrdU-labelled mitotic cell pairs (**Fig. 3a** and **Supplementary Fig. 3a**). We then scored $BrdU^+$ doublets for the expression of K14 and K18.

We found that the majority of divisions were symmetric, such that daughter cells were either both $K18^+$ or both $K18^+K14^+$ (**Fig. 3b,c**). However, ~14% of cell divisions were asymmetric, giving rise to one $K18^+$ and one $K18^+K14^+$ daughter cell (**Fig. 3b,c**). We verified that asymmetric cell doublets indeed represent products of one mitosis rather than randomly positioned cells by mixing cells expressing different fluorescent markers and scoring the numbers of cell pairs of different colors, which was found to be negligible (**Supplementary Fig. 3b**). Furthermore, daughter cells of the same mitosis could be identified by their matching levels and patterns of BrdU staining, which was distinct in each pair (**Fig. 3b**). These findings indicate that asymmetric divisions occur within basal-like breast cancer cell populations, allowing phenotypic transitions. Interestingly, we did not detect asymmetric divisions in which one daughter cell expressed vimentin and the other did not (**Supplementary Fig. 3c**), suggesting that transitions to the partially mesenchymal state do not occur through asymmetric divisions, or that vimentin is not an appropriate marker for the detection of such divisions.

**Polycomb repressive complex 2 maintains $K18^+K14^+$ cell numbers by inhibiting asymmetric divisions**

We previously reported that shRNA silencing of EZH2, the catalytic subunit of the Polycomb repressive complex 2 (PRC2), or of EED, another critical PRC2 component, leads to reduced numbers of $K18^+K14^+$ cells[25]. We found that treatment of HCC70 cells with the EZH2 pharmacologic inhibitor, GSK-126[27], had a similar effect, and decreased the fraction of $K18^+K14^+$ cells in a dose-dependent manner, down to approximately half of its original size (~13% of total from ~29%) (**Fig. 3d**). As expected, GSK-126 treatment reduced global H3K27me3 levels in the cells, but it did not significantly affect cell viability (**Supplementary Fig. 4a,b**). EZH2 thus promotes the $K18^+K14^+$ state, consistent with its role in maintaining

luminal progenitor identity in the normal breast[28,29].

To test whether PRC2 influences rates of asymmetric divisions, we compared the percentages of symmetric and asymmetric divisions in cells in which EZH2 or EED was silenced. Strikingly, the percentages of asymmetric divisions in EZH2- and EED-silenced cells was more than double of that of control cells, while the percentage of symmetric $K18^+K14^+$ divisions was proportionally decreased (**Fig. 3e**). These changed proportions would give rise to fewer $K18^+K14^+$ daughter cells than in the control cells. Treatment of the cells with GSK-126 also reduced the ratio of symmetric $K18^+K14^+$ divisions relative to asymmetric divisions, increasing the latter (**Fig. 3f**). In addition, GSK-126 treated cells had a higher overall proportion of symmetric $K18^+$ divisions, further increasing the production of $K18^+$ cells. Together these results indicate that PRC2 promotes symmetric $K18^+K14^+$ divisions and inhibits asymmetric divisions.

**FOXA1 limits $K18^+K14^+$ cell numbers by promoting asymmetric divisions**

FOXA1 is a master transcriptional regulator controlling luminal differentiation and the function of the estrogen receptor[30,31]. FOXA1 expression levels are typically high in tumors of the luminal subtypes, and it is mutated in ~2% of these tumors[32,33]. Basal-like tumors and cell lines express lower levels of FOXA1, yet these are variable, and are negatively correlated with K14 and EZH2 (**Supplementary Fig. 5a–c**). Both FOXA1 and GATA3, another central regulator of luminal differentiation[34,35], are negatively regulated by EZH2[25] (**Supplementary Fig. 5d**). We found that silencing of FOXA1 in HCC70 cells leads to a substantial increase in the fraction of $K18^+K14^+$ cells (**Fig. 3g** and **Supplementary Fig. 5e**), indicating that FOXA1 limits the numbers of these cells. Strikingly, FOXA1-silenced cells displayed a dramatic increase in the proportion of $K18^+K14^+$ symmetric divisions, while asymmetric divisions were reduced to ~3%, an effect opposite of that of EZH2 or EED silencing (**Fig. 3h**). GATA3 silencing influenced the

proportion of divisions in a similar manner (**Fig. 3h**). These results indicate that FOXA1 and GATA3 promote asymmetric divisions at the expense of $K18^+K14^+$ symmetric divisions, exerting a role consistent with their function in the normal tissue and opposing the function of PRC2.

**A functional screen identifies regulators of heterogeneity**

To identify additional components of the regulatory network controlling the composition of basal-like breast cancers we conducted a functional screen. We chose 177 candidate transcription factors and signaling molecules that are either known regulators of differentiation in the normal or cancerous breast, or show high and specific expression in basal-like tumors versus other subtypes (**Supplementary Table 4**). Each gene was silenced using 2 shRNA lentiviral vectors whose knockdown efficiency was previously validated, or up to 16 vectors for which knockdown efficiency was unknown, a total of 534 shRNAs (**Supplementary Table 4**). We infected HCC70 cells seeded in 96-well plates with shRNA-expressing viruses, including several control shRNAs. Following selection, we stained the cells and quantified the composition of each infected population by FACS (**Fig. 4a,b**).

Using this method we identified candidate genes whose silencing increased or decreased the fraction of $K18^+K14^+$ cells relative to cells infected with control vectors (**Fig. 4b,c** and **Supplementary Fig. 6**). shRNA vectors silencing the gene encoding K14 itself (*KRT14*), used as a positive control, showed the expected reduction in the number of $K18^+K14^+$ cells (**Fig. 4b**). Among the genes whose targeting by shRNA led to reductions in the percentage of $K18^+K14^+$ cells were those encoding the SOX9 transcription factor, a known regulator of mammary progenitor identity[36], as well as the KLF5, FOXQ1, ID3 and NFIB transcription factors, and FZD7, a Wnt receptor known to regulate stem cell identity in the normal and cancerous breast[37]

(**Fig. 4c**). Silencing of the RBPJ transcription factor (also known as CBF1), the mediator of Notch signaling[38], also reduced $K18^+K14^+$ cell numbers, suggesting a role for this pathway in controlling heterogeneity (**Fig. 4b,c**). As expected, hairpins silencing FOXA1 led to an increase in the $K18^+K14^+$ fraction (**Fig. 4b,c**), as did silencing of the genes encoding the IFI16 transcription factor, the negative regulator of Wnt signaling SFRP1, and the LYN and FYN kinases (**Fig. 4c**). Together these findings implicate a group of regulators in controlling the numbers of $K18^+K14^+$ cells in basal-like breast cancer cell populations.

We next conducted gene expression profiling of cells in which the regulators affecting $K18^+K14^+$ numbers were silenced, and analyzed the changes in lineage-associated gene signatures in each of the shRNA-knockdown lines. To reduce shRNA-specific effects we averaged the expression profiles of the two shRNAs targeting each regulator. We found that regulators could be divided into five main groups, based on the effect they had on the expression of the three lineage-associated signatures (**Fig. 4d**). Cluster A contained genes (including EZH2, EED and FOXQ1) whose silencing led to reduced expression of the luminal progenitor signature, accompanied by increased expression of the mature luminal signature (**Fig. 4d**). These factors thus repress luminal differentiation, and promote progenitor identity (**Fig. 4e**). Cluster B contained genes, including SOX9, RBPJ, NFIB and SFRP1, whose silencing led to a more pronounced reduction in the expression of the progenitor signature, but led to increased expression of the basal signature, indicating that these regulators promote progenitor identity and repress basal identity (**Fig. 4d,e**). Cluster D factors, which included GATA3, FOXA1, had the expected effects upon silencing – increased expression of progenitor and basal genes and reduced expression of mature luminal genes; interestingly this cluster included also HIF1$\alpha$ (**Fig. 4d,e**). These results uncover groups of regulators that exert distinct effects on the expression of lineage-associated expression

programs, and control the composition of the cancer cell population.

**The Notch pathway regulates $K18^+K14^+$ numbers and asymmetric divisions**

The functional screen revealed that silencing of RBPJ, a transcription factor that plays a central role in Notch signaling, causes a substantial reduction in the numbers of the $K18^+K14^+$ cells (**Fig. 5a**), suggesting that the Notch pathway influences the composition of the cancer cell population. Expression of Notch transcriptional targets was dramatically reduced in RBPJ-silenced basal-like cells, indicating ongoing Notch activity in these cells (**Fig. 5b**). Indeed, silencing of Notch1 had a similar effect on population composition and target gene expression (**Fig. 5c,d**). Furthermore, treatment of the cells with the γ-secretase inhibitor (GSI) LY-411575, which inhibits Notch, also reduced $K18^+K14^+$ numbers and Notch target genes (**Fig. 5e,f**), as well the luminal progenitor signature (**Fig. 4d**) without substantially affecting cell viability (**Supplementary Fig. 7a**). Conversely, overexpression of the active Notch1 intracellular domain (NICD), led to an increase in expression of K14 and of Notch targets (**Fig. 5g**). These findings indicate that Notch pathway activity promotes the $K18^+K14^+$ identity. Interestingly, RBPJ silencing led to a dramatic decrease in tumor formation by the HCC70 cells (**Supplementary Fig. 7b**), highlighting a pro-tumorigenic function of this transcription factor.

We next tested whether Notch regulates asymmetric division rates. We treated cells with GSI for three days, and then scored symmetric versus asymmetric divisions. The number of symmetric $K18^+K14^+$ divisions was substantially reduced in GSI treated cells, while the number of asymmetric divisions was increased (**Fig. 5h**). Notch activity thus promotes the numbers of $K18^+K14^+$ cells in a manner similar to that of EZH2, suppressing asymmetric divisions.

In light of the effects of the Notch pathway on $K18^+K14^+$ cell numbers and asymmetric division rates, and the known functions of this pathway in differentiation[38], we examined which of the

other identified regulators influences Notch activity. We found that the expression of the Notch signature, defined as genes whose expression was changed upon GSI treatment (**Supplementary Table 2**), was reduced in cells silenced for EED or EZH2, or treated with the EZH2 inhibitor GSK-126 (**Supplementary Fig. 7c**). qRT-PCR analysis indicated that, indeed, GSK-126-treated cells expressed lower levels of Notch receptors and targets (**Fig. 5i**). EZH2 levels were, however, unchanged in GSI-treated or Notch1-silenced cells (**Supplementary Fig. 7d**), suggesting that EZH2 acts upstream to Notch.

Cells silenced for the expression of NFIB, a pro-tumorigenic transcription factor that is often translocated in adenoid cystic breast carcinomas[39,40], also showed a substantial reduction in $K18^+K14^+$ cell numbers (**Fig. 4c and Fig. 5j**). Consistent with this, high NFIB levels were observed preferentially in basal-like breast cancers, and in normal luminal progenitor and basal cells (**Supplementary Fig. 8a,b**). NFIB-silenced cells displayed significantly reduced expression of the Notch signature and targets, but EZH2 levels were unchanged (**Fig. 5k** and **Supplementary Fig. 7c**), suggesting that its regulation of Notch was achieved independently of EZH2.

To further study the links between EZH2, NFIB and Notch, we tested the effects of co-inhibition of these factors. Inhibition of Notch by GSI treatment reduced $K18^+K14^+$ cell numbers to lower levels than did EZH2 inhibition by GSK-126, while treatment with both drugs reduced these numbers only slightly more than GSI treatment alone, a change that was not statistically significant (**Fig. 5j**). This suggests that while Notch is a mediator of EZH2 regulation of $K18^+K14^+$ cell numbers, it likely receives inputs from additional sources. NFIB-silenced cells also showed a significant reduction in $K18^+K14^+$ cell numbers upon GSI treatment, while GSK-126 treatment, alone or combined with the GSI had a more modest effect (**Fig. 5j**). Together

these results indicate that Notch is a central regulator of $K18^{+}K14^{+}$ cell numbers and of asymmetric division rates, and is positively regulated by both EZH2 and NFIB (**Fig. 5l**).

**Discussion**

Accumulating evidence indicates that non-genetic mechanisms substantially contribute to cellular heterogeneity within tumors, and thereby to disease progression and resistance to therapy[1,2,6,41-44]. The ability of cells to transition between differentiation states and adopt intermediate states can generate a wide spectrum of cell identities, offering a range of functional phenotypes. Population-averaged assessments of tumor differentiation can thus represent a mixture of widely different cellular differentiation states, and therefore analysis at the whole tumor level may obscure critical aspects of phenotypic heterogeneity.

Our results provide a novel view of the mechanisms regulating the cellular composition of basal-like breast cancers, tumors whose effective treatment is currently challenging. At the whole tumor level, basal-like breast cancers exhibit a gene expression program that is related to that of luminal progenitor cells[17]. However, their high level of cellular heterogeneity suggests that cells within these tumors can transition between states that are closer to the "mean" progenitor-related state or more distant from it. This heterogeneity is maintained, at least in part, in cell lines derived from this disease (often termed Basal A lines). We therefore began our analysis by asking whether expression of these known lineage markers reflected distinct differentiation states. Our analysis of the expression profiles of isolated cell subpopulations revealed that cells co-expressing K18 and vimentin show enrichment for an EMT signature, indicating that this subpopulation indeed resides in a partially-mesenchymal state. We also show that the expression profile of cells co-expressing K14 and K18 is more closely correlated with that of luminal progenitors than that of other subpopulations, while cells expressing only K18 show a higher

correlation with the signature of mature luminal cells. The $K18^+$ subpopulation is not, however, fully luminally differentiated – these cells do not express the estrogen receptor and therefore likely less differentiated than cells in luminal tumors. The association of the $K18^+K14^+$ profile with normal progenitor identity is consistent with previous demonstrations of basal keratin expression in progenitor cells[17,22,45]. Other markers used to define luminal progenitors, such as CD49f, EpCAM and ALDH[5,46,47] may allow for further enrichment for progenitor-like cells from these tumors. Recent studies in fact suggest that several types of progenitors exist in the luminal compartment[48,49], and therefore distinct progenitor-like subtypes may exist in tumors, with the $K18^+K14^+$ profile enriching for a particular subset.

Our experiments indicate that $K18^+K14^+$ cells possess an advantage in tumorigenicity and metastasis. This is consistent with previous work showing that $K14^+$ cells are present in the invasive edges and metastases of breast cancers and have enhanced collective invasiveness[50], and with work showing increased expression of basal and progenitor markers in individually seeded metastatic cells[51]. The progenitor-related differentiation state of $K18^+K14^+$ cells may thus provide distinct tumorigenic traits *in vivo* that could affect disease progression, and the distinction between such cells and previously described mesenchymal-like cancer stem cells is consistent with the now well-founded notion that more than one type of stem- or progenitor-like cells exists in breast cancer[5].

The scarcity of cells co-expressing K14 and vimentin, both in patient tumors and in cells lines, suggests that the mesenchymal–epithelial differentiation axis, controlled by the EMT program, can be viewed as distinct from the basal–luminal differentiation axis, active in the normal breast, and that cancer cells can transition along either axis[11]. Emphasizing this distinction, cell lines of the Basal B group, often used as a model for basal-like disease, are in fact mesenchymal in nature

and do not capture the largely epithelial identity of these carcinomas, unlike Basal A lines which maintain overall epithelial identity and retain heterogeneity[22].

The processes allowing cancer cells to transition from one state to another are overall poorly characterized and it is not clear whether such transitions are generally dependent on cell division. Asymmetric divisions represent one mechanism tying cell division to differentiation, through the production of two daughter cells each displaying a distinct state[26]. The existence of asymmetric divisions has been shown in cell populations of several cancer types, including glioma, colon and breast[52-54]. We directly show asymmetric divisions occurring within basal-like cell populations, producing one $K18^+K14^+$ daughter cell and one $K18^+$ daughter cell. Polycomb and Notch repress asymmetric divisions and increase symmetric $K18^+K14^+$ divisions, while FOXA1 and GATA3 promote asymmetric divisions. The relative balance in the activity of these factors thus determines the equilibrium of cell subpopulation fractions, and their roles are also consistent with their functions in the normal mammary gland: EZH2 and Notch promote progenitor identity[28,29,55] while FOXA1 and GATA3 promote luminal differentiation[30,34,35]. Our ability to detect asymmetric divisions involving K14 may be related to the observation that cytokeratin filaments disintegrate during cell divisions and are distributed between daughter cells prior reassembly[56]. Indeed, asymmetric keratin distribution has also been observed in colon cancer cells[54]. We were not able to directly determine the identities of the mother cells of the three different division types detected in the population, and therefore it remains to be further studied whether transitions preferably occur towards one or another identity.

Our functional screen allowed us to identify several regulators that increase or decrease $K18^+K14^+$ cell numbers in the cancer cell population. Some of these, such as SOX9, FZD7[36,37] have been shown to regulate mammary progenitor and stem cell identity, while most others are

known to control differentiation in different settings. Analysis of the effects of these regulators on the mammary lineage-associated expression signatures allowed their clustering into five groups, exerting distinct changes. This supports a model in which the combined action of a network of regulators drives the overall levels of expression of differentiation signatures. The direct interactions among these regulators require further analysis, but we uncover some of these relationships. We show that Notch, a known regulator of mammary differentiation and of division symmetry in a variety of differentiation settings[38,55,57], is an important component of this network. Notch acts downstream to EZH2, an interaction previously described[58]. NFIB also positively regulates Notch, a novel functional relationship. Previous studies have implicated Notch and its target and negative regulator Numb in regulation of asymmetric divisions, and p53 and miR-34 have been linked to their activity[54,59,60]. These components could, potentially, also influence Notch activity in the context described here.

Together, our findings shed new light on the manner by which cellular heterogeneity and differentiation state in basal-like breast cancer is regulated, and highlight several specific factors that influence population composition. We focused on cell-intrinsic mechanisms regulating cellular identity; however, extrinsic signals, such as ligand binding, extracellular matrix composition, hypoxia and metabolic state are likely to play important roles in influencing cell identity, potentially influencing some of regulators described here. Importantly we show that manipulation of tumor composition can be achieved by pharmacologic means, specifically inhibitors of EZH2 and Notch, currently undergoing clinical evaluation[27,61]. This suggests that approaches aimed at modulating cancer cell differentiation state and thereby the cellular composition of tumors may be considered as a component of treatment that could influence responses to cytotoxic therapy.

**Acknowledgements**

We thank N. E. Kidess-Bassir for histological preparation. This study was supported by grants from the Israel Cancer Research Foundation (I.B.-P.), the Israel Science Foundation (2309/15), the Lower Saxony – Israel Research Cooperation Program (I.B.-P. and U.L.), the Joint Hebrew University-Hadassah Research Fund, the Israel Cancer Association (I.B.-P. and D.K.), the Jacob and Lena Joels Memorial Foundation Senior Lectureship for Excellence in the Life and Medical Sciences (I.B.-P.) the United States Agency for International Development American Schools and Hospitals Abroad Program (I.B.-P.), the Janey and Albert Sweet Scholarships (R.Z.G.), the Leonard and Faigel Shapiro Fellowship of the Canadian Friends of the Hebrew University (R.Z.G) and the Bester Scholarship Fund (R.Z.G.).

**Author Contributions**

R.Z.G. and I.B.-P. designed the study and wrote the manuscript, R.Z.G. conducted the experiments, Y.F., S.D. and R.M. assisted in experimentation and data analysis, E.C., T.S., A.S., D.K., U.L. and K.P. obtained and analyzed patient tumor samples, F.P., A.R. and D.E.R. participated in functional screening and data analysis.

**Competing financial Interests**

K.P. is a shareholder at Champions Oncology. Other authors declare no conflict of interest.

**References**


1. Marusyk, A., Almendro, V., and Polyak, K. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* **12:**323-334.
2. Meacham, C.E., and Morrison, S.J. (2013) Tumour heterogeneity and cancer cell plasticity. *Nature* **501:**328-337.
3. Lobo, N.A., Shimono, Y., Qian, D., and Clarke, M.F. (2007) The biology of cancer stem cells. *Annu Rev Cell Dev Biol* **23:**675-699.
4. Condiotti, R., Guo, W., and Ben-Porath, I. (2014) Evolving views of breast cancer stem cells and their differentiation States. *Crit Rev Oncog* **19:**337-348.
5. Brooks, M.D., Burness, M.L., and Wicha, M.S. (2015) Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer. *Cell Stem Cell* **17:**260-271.
6. Roesch, A., Fukunaga-Kalabis, M., Schmidt, E.C., Zabierowski, S.E., Brafford, P.A., Vultur, A., Basu, D., Gimotty, P., Vogt, T., and Herlyn, M. (2010) A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141:**583-594.
7. Chaffer, C.L., Marjanovic, N.D., Lee, T., Bell, G., Kleer, C.G., Reinhardt, F., D'Alessio, A.C., Young, R.A., and Weinberg, R.A. (2013) Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. *Cell* **154:**61-74.
8. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., Louis, D.N., Rozenblatt-Rosen, O., Suva, M.L., Regev, A., and Bernstein, B.E. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344:**1396-1401.
9. Quintana, E., Shackleton, M., Foster, H.R., Fullen, D.R., Sabel, M.S., Johnson, T.M., and Morrison, S.J. (2010) Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. *Cancer Cell* **18:**510-523.
10. Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146:**633-644.
11. Granit, R.Z., Slyper, M., and Ben-Porath, I. (2014) Axes of differentiation in breast cancer: untangling stemness, lineage identity, and the epithelial to mesenchymal transition. *Wiley Interdiscip Rev Syst Biol Med* **6:**93-106.

12. Tam, W.L., and Weinberg, R.A. (2013) The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med* **19:**1438-1449.
13. Tsai, J.H., Donaher, J.L., Murphy, D.A., Chau, S., and Yang, J. (2012) Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* **22:**725-736.
14. Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M., Concannon, K.F., Donaldson, M.C., Sequist, L.V., Brachtel, E., Sgroi, D., Baselga, J., Ramaswamy, S., Toner, M., Haber, D.A., and Maheswaran, S. (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339:**580-584.
15. Visvader, J.E. (2009) Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev* **23:**2563-2577.
16. Prat, A., and Perou, C.M. (2011) Deconstructing the molecular portraits of breast cancer. *Mol Oncol* **5:**5-23.
17. Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., Feleppa, F., Huschtscha, L.I., Thorne, H.J., Fox, S.B., Yan, M., French, J.D., Brown, M.A., Smyth, G.K., Visvader, J.E., and Lindeman, G.J. (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15:**907-913.
18. Keller, P.J., Arendt, L.M., Skibinski, A., Logvinenko, T., Klebba, I., Dong, S., Smith, A.E., Prat, A., Perou, C.M., Gilmore, H., Schnitt, S., Naber, S.P., Garlick, J.A., and Kuperwasser, C. (2011) Defining the cellular precursors to human breast cancer. *Proc Natl Acad Sci U S A* **109:**2772-2777.
19. Molyneux, G., Geyer, F.C., Magnay, F.A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., Reis-Filho, J.S., and Smalley, M.J. (2010) BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* **7:**403-417.
20. Badve, S., Dabbs, D.J., Schnitt, S.J., Baehner, F.L., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R., Palacios, J., Rakha, E.A., Richardson, A.L., Schmitt, F.C., Tan, P.H., Tse, G.M., Weigelt, B., Ellis, I.O., and Reis-Filho, J.S. (2011)

Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. *Mod Pathol* **24:**157-167.

21. Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., Lapuk, A., Wang, N.J., Kuo, W.L., Stilwell, J.L., Pinkel, D., Albertson, D.G., Waldman, F.M., McCormick, F., Dickson, R.B., Johnson, M.D., Lippman, M., Ethier, S., Gazdar, A., and Gray, J.W. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10:**515-527.
22. Prat, A., Karginova, O., Parker, J.S., Fan, C., He, X., Bixby, L., Harrell, J.C., Roman, E., Adamo, B., Troester, M., and Perou, C.M. (2013) Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat* **142:**237-255.
23. Taube, J.H., Herschkowitz, J.I., Komurov, K., Zhou, A.Y., Gupta, S., Yang, J., Hartwell, K., Onder, T.T., Gupta, P.B., Evans, K.W., Hollier, B.G., Ram, P.T., Lander, E.S., Rosen, J.M., Weinberg, R.A., and Mani, S.A. (2010) Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A* **107:**15449-15454.
24. Ye, X., Tam, W.L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E., and Weinberg, R.A. (2015) Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525:**256-260.
25. Granit, R.Z., Gabai, Y., Hadar, T., Karamansha, Y., Liberman, L., Waldhorn, I., Gat-Viks, I., Regev, A., Maly, B., Darash-Yahana, M., Peretz, T., and Ben-Porath, I. (2013) EZH2 promotes a bi-lineage identity in basal-like breast cancer cells. *Oncogene* **32:**3886-3895.
26. Morrison, S.J., and Kimble, J. (2006) Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441:**1068-1074.
27. McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., 3rd, Diaz, E., LaFrance, L.V., Mellinger, M., Duquenne, C., Tian, X., Kruger, R.G., McHugh, C.F., Brandt, M., Miller, W.H., Dhanak, D., Verma, S.K., Tummino, P.J., and Creasy, C.L. (2012) EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492:**108-112.

28. Pal, B., Bouras, T., Shi, W., Vaillant, F., Sheridan, J.M., Fu, N., Breslin, K., Jiang, K., Ritchie, M.E., Young, M., Lindeman, G.J., Smyth, G.K., and Visvader, J.E. (2013) Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. *Cell Rep* **3:**411-426.
29. Michalak, E.M., Nacerddine, K., Pietersen, A., Beuger, V., Pawlitzky, I., Cornelissen-Steijger, P., Wientjens, E., Tanger, E., Seibler, J., van Lohuizen, M., and Jonkers, J. (2013) The Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal progenitor pool. *Stem Cells*.
30. Bernardo, G.M., and Keri, R.A. (2012) FOXA1: a transcription factor with parallel functions in development and cancer. *Biosci Rep* **32:**113-130.
31. Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43:**27-33.
32. Nakshatri, H., and Badve, S. (2009) FOXA1 in breast cancer. *Expert Rev Mol Med* **11:**e8.
33. Robinson, J.L., Holmes, K.A., and Carroll, J.S. (2013) FOXA1 mutations in hormone-dependent cancers. *Front Oncol* **3:**20.
34. Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006) GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127:**1041-1055.
35. Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., Lindeman, G.J., and Visvader, J.E. (2007) Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9:**201-209.
36. Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zurrer-Hardi, U., Bell, G., Tam, W.L., Mani, S.A., van Oudenaarden, A., and Weinberg, R.A. (2012) Slug and sox9 cooperatively determine the mammary stem cell state. *Cell* **148:**1015-1028.
37. Chakrabarti, R., Wei, Y., Hwang, J., Hang, X., Andres Blanco, M., Choudhury, A., Tiede, B., Romano, R.A., DeCoste, C., Mercatali, L., Ibrahim, T., Amadori, D., Kannan, N., Eaves, C.J., Sinha, S., and Kang, Y. (2014) DeltaNp63 promotes stem cell activity in

mammary gland development and basal-like breast cancer by enhancing Fzd7 expression and Wnt signalling. *Nat Cell Biol* **16:**1004-1015, 1001-1013.

38. Koch, U., Lehal, R., and Radtke, F. (2013) Stem cells living with a Notch. *Development* **140:**689-704.
39. Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., Crowley, D., Barretina, J., Garraway, L., Meyerson, M., and Jacks, T. (2011) Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev* **25:**1470-1475.
40. Persson, M., Andren, Y., Mark, J., Horlings, H.M., Persson, F., and Stenman, G. (2009) Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proc Natl Acad Sci U S A* **106:**18740-18744.
41. Almendro, V., Cheng, Y.K., Randles, A., Itzkovitz, S., Marusyk, A., Ametller, E., Gonzalez-Farre, X., Munoz, M., Russnes, H.G., Helland, A., Rye, I.H., Borresen-Dale, A.L., Maruyama, R., van Oudenaarden, A., Dowsett, M., Jones, R.L., Reis-Filho, J., Gascon, P., Gonen, M., Michor, F., and Polyak, K. (2014) Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Rep* **6:**514-527.
42. Su, Y., Subedee, A., Bloushtain-Qimron, N., Savova, V., Krzystanek, M., Li, L., Marusyk, A., Tabassum, D.P., Zak, A., Flacker, M.J., Li, M., Lin, J.J., Sukumar, S., Suzuki, H., Long, H., Szallasi, Z., Gimelbrant, A., Maruyama, R., and Polyak, K. (2015) Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. *Cell Rep* **11:**1549-1563.
43. Wagenblast, E., Soto, M., Gutierrez-Angel, S., Hartl, C.A., Gable, A.L., Maceli, A.R., Erard, N., Williams, A.M., Kim, S.Y., Dickopf, S., Harrell, J.C., Smith, A.D., Perou, C.M., Wilkinson, J.E., Hannon, G.J., and Knott, S.R. (2015) A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520:**358-362.
44. Allen, E., Mieville, P., Warren, C.M., Saghafinia, S., Li, L., Peng, M.W., and Hanahan, D. (2016) Metabolic Symbiosis Enables Adaptive Resistance to Anti-angiogenic Therapy that Is Dependent on mTOR Signaling. *Cell Rep* **15:**1144-1160.

45. Chakrabarti, R., Wei, Y., Romano, R.A., DeCoste, C., Kang, Y., and Sinha, S. (2012) Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. *Stem Cells* **30:**1496-1508.
46. Visvader, J.E., and Stingl, J. (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* **28:**1143-1158.
47. Smalley, M.J., Kendrick, H., Sheridan, J.M., Regan, J.L., Prater, M.D., Lindeman, G.J., Watson, C.J., Visvader, J.E., and Stingl, J. (2012) Isolation of mouse mammary epithelial subpopulations: a comparison of leading methods. *J Mammary Gland Biol Neoplasia* **17:**91-97.
48. Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C.J., and Stingl, J. (2012) Phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res* **14:**R134.
49. Tornillo, G., and Smalley, M.J. (2015) ERrrr...where are the progenitors? Hormone receptors and mammary cell heterogeneity. *J Mammary Gland Biol Neoplasia* **20:**63-73.
50. Cheung, K.J., Gabrielson, E., Werb, Z., and Ewald, A.J. (2013) Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155:**1639-1651.
51. Lawson, D.A., Bhakta, N.R., Kessenbrock, K., Prummel, K.D., Yu, Y., Takai, K., Zhou, A., Eyob, H., Balakrishnan, S., Wang, C.Y., Yaswen, P., Goga, A., and Werb, Z. (2015) Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526:**131-135.
52. Cicalese, A., Bonizzi, G., Pasi, C.E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Di Fiore, P.P., and Pelicci, P.G. (2009) The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. *Cell* **138:**1083-1095.
53. Sugiarto, S., Persson, A.I., Munoz, E.G., Waldhuber, M., Lamagna, C., Andor, N., Hanecker, P., Ayers-Ringler, J., Phillips, J., Siu, J., Lim, D.A., Vandenberg, S., Stallcup, W., Berger, M.S., Bergers, G., Weiss, W.A., and Petritsch, C. (2011) Asymmetry-defective oligodendrocyte progenitors are glioma precursors. *Cancer Cell* **20:**328-340.
54. Bu, P., Chen, K.Y., Chen, J.H., Wang, L., Walters, J., Shin, Y.J., Goerger, J.P., Sun, J., Witherspoon, M., Rakhilin, N., Li, J., Yang, H., Milsom, J., Lee, S., Zipfel, W., Jin,

M.M., Gumus, Z.H., Lipkin, S.M., and Shen, X. (2013) A microRNA miR-34a-regulated bimodal switch targets Notch in colon cancer stem cells. *Cell Stem Cell* **12:**602-615.

55. Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.L., Oakes, S.R., Lindeman, G.J., and Visvader, J.E. (2008) Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. *Cell Stem Cell* **3:**429-441.
56. Lane, E.B., Goodman, S.L., and Trejdosiewicz, L.K. (1982) Disruption of the keratin filament network during epithelial cell division. *EMBO J* **1:**1365-1372.
57. Harrison, H., Farnie, G., Brennan, K.R., and Clarke, R.B. (2010) Breast cancer stem cells: something out of notching? *Cancer Res* **70:**8973-8976.
58. Gonzalez, M.E., Moore, H.M., Li, X., Toy, K.A., Huang, W., Sabel, M.S., Kidwell, K.M., and Kleer, C.G. (2014) EZH2 expands breast stem cells through activation of NOTCH1 signaling. *Proc Natl Acad Sci U S A* **111:**3098-3103.
59. Bu, P., Wang, L., Chen, K.Y., Srinivasan, T., Murthy, P.K., Tung, K.L., Varanko, A.K., Chen, H.J., Ai, Y., King, S., Lipkin, S.M., and Shen, X. (2016) A miR-34a-Numb Feedforward Loop Triggered by Inflammation Regulates Asymmetric Stem Cell Division in Intestine and Colon Cancer. *Cell Stem Cell* **18:**189-202.
60. Tosoni, D., Zecchini, S., Coazzoli, M., Colaluca, I., Mazzarol, G., Rubio, A., Caccia, M., Villa, E., Zilian, O., Di Fiore, P.P., and Pece, S. (2015) The Numb/p53 circuitry couples replicative self-renewal and tumor suppression in mammary epithelial cells. *J Cell Biol* **211:**845-862.
61. Andersson, E.R., and Lendahl, U. (2014) Therapeutic modulation of Notch signalling--are we there yet? *Nat Rev Drug Discov* **13:**357-378.
62. Hu, Y., and Smyth, G.K. (2009) ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* **347:**70-78.
63. Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* **2:**666-673.
64. Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I., Fan, C., Zhang, X., He, X., Pavlick, A., Gutierrez, M.C., Renshaw, L., Larionov, A.A., Faratian, D., Hilsenbeck, S.G., Perou, C.M., Lewis, M.T., Rosen, J.M., and Chang, J.C. (2009) Residual breast cancers after

conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A* **106:**13820-13825.

## Methods

### Human tumor samples

Breast cancer sample sections were obtained from the Fox Chase Cancer Center (FCCC) Philadelphia, USA. Clinical details are found in Supplementary Table 1. A patient derived xenograft (PDX) triple-negative breast tumor sample was obtained from Champions Oncology Inc. as part of Champions Personalized Oncology Service, and grown as passage 4 in the mammary glands of immunocompromised NSG mice. Patients signed an Informed Consent allowing the use of their tissue and its derived xenografts as well as their medical information for further research. A second sample was grown in NSG mice after isolation from a biopsy obtained from a lung metastasis of a patient with recurrent triple-negative metaplastic disease treated at Hadassah Medical Center with the approval of the Institutional Review Board (Helsinki Committee) and under informed consent.

### Cell culture and infections

Cell lines were obtained from ATCC. HCC70 cells were grown in RPMI medium containing 10% FBS, and MDA-MB-468 cells in Leibovitz L15 medium containing 10% FBS and supplemented with penicillin and streptomycin. Other cell lines were grown according to ATCC media protocols. The EZH2 inhibitor GSK-126 (Cayman Chemicals) and the γ-secretase inhibitor LY-411575 (Sigma-Aldrich) were dissolved in DMSO and added to the media in indicated concentrations for three days. Gene silencing was performed using pLKO-puro shRNA vectors (detailed in Supplementary Table 4). The K14p-GFP reporter lentivirus was previously described[25]. For notch intracellular domain overexpression we used the pMIG-NICD retroviral construct (kindly provided by Michael Berger). Standard virus generation and infection

procedures were used, with packaging conducted by transfection into 293T cells with the pHRΔ8.2 and pCMV-VSV-G vectors. Cell lines were authenticated by short tandem repeat profiling.

**Immunostaining and FACS**

Immunohistological staining was conducted on paraffin embedded sections using standard procedures. For immunofluorescence we used antibodies against K14 (RB-9020, Thermo), K18 (MS-142, Thermo) and vimentin (GP-53, Progen), and secondary antibodies from Jackson ImmunoResearch Laboratories. Images were collected on an Olympus Fv10I confocal microscope. Cell subpopulation numbers in human tumor sections were scored visually using nuclear morphology for identification of carcinoma cells, and the median percentages of ten microscopic fields from each tumor were used. For immunohistochemical staining we used HRP secondary antibodies (ImmPRESS reagent kit, Vector) followed by peroxidase substrate kit DAB (SK-4100, Vector), and images were collected on an Olympus CX41 microscope using Nikon DS-Fi1 camera. For staining of cells for differentiation markers by FACS, the cells were detached by trypsin-EDTA, fixed and permeabilized for 10 min in 100% methanol on ice, and stained with antibodies against K18 (sc-31700, Santa Cruz), K14, and vimentin (NCL-L-VIM-V9, Novocastra), followed by conjugated secondary antibodies. Stained cells were analyzed on a MACSQuant Analyzer (Miltenyi Biotec). Isolation of stained cell subpopulations or GFP-expressing cells was done on an Aria III sorter (BD biosciences). For quantification of H3K27me3 levels cells were stained with an antibody against H3K27me3 (07-449, Millipore). Annexin V (Biolegend) staining and quantification was used to to assess apoptosis following inhibitor treatment.

**Tumor cell transplantation**

For limiting dilution tumor seeding, HCC70 cells infected with the K14p-GFP vector were sorted for $GFP^{high}$ and $GFP^{-}$ fractions by FACS, and cells were counted and injected in 10μl culture medium containing 25% Matrigel (BD-biosciences) into both #4 mammary glands of 6-week-old female NOD-SCID mice. Tumors were measured externally by calipers and followed up to 5 months. Lack of tumor formation was validated by excision of the mammary gland. The proportion of tumor initiating cells was calculated by the ELDA algorithm[62]. To evaluate the metastatic presence of $K18^{+}K14^{+}$ cells, $1x10^{6}$ GFP-labelled MDA-MB-468 cells were suspended in 25% matrigel in media, injected into #4 mammary glands of 6-week-old female NSG mice and allowed to form tumors for 3 months. Primary tumors were then excised, sectioned and stained for K14, K18 and GFP (ab6673 Abcam). For lung seeding experiments, $1x10^{6}$ GFP-labelled MDA-MB-468 cells were suspended in 100μl F12 media and injected into the tail vein of NSG female mice. Lungs were excised and stained for K18, K14 and GFP at indicated times after injection. The percentage of $K18^{+}K14^{+}$ cells among $GFP^{+}$ cells was scored by image analysis using the CellProfiler software. For growth of patient derived tumors, samples were minced and implanted into both #4 mammary glands of 6-week-old female NSG mice in 100μl 50% Matrigel (BD biosciences). Tumors were extracted 3–9 weeks after transplantation, minced and suspended in media containing collagenase A (Roche, 1.5mg/ml), incubated for 1 hour in a 37°C rocker, filtered through a 40μM strainer (BD) and stained and sorted to isolate $K18^{+}$ and $K18^{+}K14^{+}$ cell fractions. The joint Institutional Animal Care and Use Committee of the Hebrew University and Hadassah Medical Center approved the study protocol for animal welfare. The Hebrew University is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.

**Cell division symmetry assay**

HCC70 cells were filtered twice through a 40μm strainer and seeded sparsely on chambered cover glasses (Lab-Tek, Nunc) coated with 2% gelatin (Sigma-Aldrich). Cells were allowed to adhere for 24 hours and then were treated for 1 hour with 10μg/ml BrdU (Sigma-Aldrich). 18 hours later the cells were fixed with cold ethanol and treated with 2M HCl. Residual acid was neutralized with 0.1M sodium tetraborate ($Na_2B_4O_7$) pH=8.5, and cells were stained for BrdU (sc-56258 Santa Cruz), K14 and K18. Experiments testing the effect of gene silencing were conducted 4 days after infection with the appropriate shRNA lentivirus. Treatment of cells with inhibitors was conducted for 3 days prior to the assay. Stained cells were imaged and scored using a confocal microscope. Only cell doublets that showed identical BrdU stain pattern were scored. In rare cases of ambiguity, a threshold of 70% of total K14 staining intensity in one daughter cells was used as a criterion for asymmetric division. To assess the occurrence of random placement of BrdU-labelled cell pairs that were not mitotic products, we seeded a 1:1 mix of HCC70 cells infected with pLeGO-C2 expressing mCherry (Addgene #27339) and pLeGO-Cer2 expressing Cerulean fluorescent protein (Addgene #27338). Cells were visualized 2 and 18 hours after seeding and the numbers of same and different color pairs were scored.

**Functional screening**

177 candidate regulators were chosen for functional screening. 534 concentrated shRNA viruses targeting these genes as well as 6 non-targeting control viruses were obtained from The RNAi Consortium at the Broad Institute (Supplementary Table 4). For genes with pre-validated knockdown efficiency data, two individual shRNAs sequences were used, for genes in which knockdown efficiency of vectors was unknown all available shRNAs were screened, a range of 2–16. A set of six different control shRNA viruses were included in each 96 well plate. $1.5x10^4$

HCC70 cells were seeded per well in 96-well plates, infected 24 hours post seeding, placed under puromycin selection (2μg/ml) 48 hours later, and and fixed and stained for K14 and vimentin 7 days after seeding. Seeding and staining were conducted using a Microlab STAR Line robotic liquid handler (Hamilton). Cells then underwent FACS analysis in 96-well format on a MACSQuant Analyzer (Miltenyi Biotec), and the percentages of $K14^+$ and $VIM^+$ cells were determined. Cells were treated with the Fixable Viability Dye eFluor 450 (eBiosicences), and >3,000 live cells per well were scored. Each plate was screened at least twice to identify genes whose silencing altered $K18^+K14^+$ numbers relative to controls, which were defined as those for which two shRNAs or more caused $K18^+K14^+$ cell percentages to be greater or lower than the range observed in control shRNA infected cells, using a 0.95 confidence interval. We conducted a secondary validation screen by testing 2 hairpins for each of the first-round positive genes (43 genes, and 4 controls) in a single plate, repeated in three independent infection-staining experiments. A final 15 genes whose silencing consistently increased or decreased $K18^+K14^+$ cell numbers were identified. We found that no seed sequences were shared between shRNAs that influenced $K18^+K14^+$ cell numbers, reducing the likelihood of seed-specific effects.

**RNA extraction, qRT-PCR and expression profiling**

RNA extraction was conducted using ISOLATE II RNA Mini Kit (Bioline) and qRT–PCR analyses were performed according to standard procedures using the iScript cDNA synthesis kit (Bio-Rad) and iTaq Universal SYBR mix (Bio-Rad). Results were normalized to levels of HPRT1. To extract RNA from fixed and stained cell subpopulations isolated by FACS from cell lines and the PDX tumor, we used Recover All Total Nucleic Acid Isolation Kit (Ambion). RNAs extracted from each subpopulation were pooled from two independent sorts of each cell line, and libraries were prepared and sequenced using Illumina's RNA sequencing protocol

(HiSeq). Reads were mapped to Human genome (Ensembl GRCh37) using TopHat2. Read quantification and normalization were done using Cuffdiff to produce gene level normalized expression values (fragments per kb of exon per million). Expression data for each subpopulation was normalized relative to the mean expression level of each gene across the subpopulations in each cell line. Principal component analysis of subpopulations was computed in R using the prcom function and visualized using Plot.ly. For expression profiling of HCC70 cells silenced for regulators identified in the functional screen, we followed an adapted CEL-Seq protocol[63]. Cells were grown in 96-well plates and lysed seven days after infection. RNA was extracted using the Single Cell RNA Purification Kit (Norgen) and 3' cDNA was synthesized and barcoded in batches of 12 samples, followed by RNA synthesis and amplification by *in vitro* transcription[63]. Each pooled RNA sample was used to generate libraries for paired-end sequencing by HiSeq (Illumina). Individual sample profiles were established by de-convolution of barcodes using the CEL-Seq-pipeline (https://github.com/yanailab/CEL-Seq-pipeline.git). Expression data was deposited in GEO accession number GSEXXX (pending).

**Gene expression analysis**

To measure the degree of expression of gene signatures in mRNA expression profiles we determined the correlation coefficient between each expression profile and each signature taking into account both up- and down-regulated genes[11,64]. Lineage gene signatures were obtained from Ref (17) and the EMT signature was adapted from Ref (11). The Notch signature was generated by extracting the genes increased or decreased two-fold or more in HCC70 cells treated with the LY-411575 GSI. Up-regulated genes in each signature were assigned a value of 1 and down-regulated genes a value of -1; the Spearman correlation coefficient between the tested gene expression profile and these values was then calculated using MatLab (MathWorks)

producing the *R* value, indicating the degree of positive or negative correlation. To examine the effects of silencing of the regulators identified in the screen on lineage signature expression, the expression profiles of each shRNA-expressing line were normalized to the median of the expression profiles of cells infected with four different control shRNAs. To reduce the impact of off-target effects, the expression profiles of cells carrying the two different shRNAs were averaged. The correlation with each signature was then calculated for each silenced regulator. The five groups of regulators were identified by hierarchical clustering of the correlation coefficients using GENE-E (Broad Institute, USA).

**Statistics**

Unless otherwise noted all *P* values were calculated using two-sided Student's *t* test. The tumor diversity index was calculated using the Shannon–Wiener index ($H' = -\sum_{i=0}^{R} pi \times pi$) based on the relative abundance cell populations in a given sample. Graphs and Chi-square tests were produced using Prism GraphPad.

**Primers**

| **Gene** | **FW** | **Rev** |
|---|---|---|
| EZH2 | GCGCGGGACGAAGAATAATCAT | TACACGCTTCCGCCAACAAACT |
| FOXA1 | AGACACGCAGGAGGCCTACTC | CATGTTGCCGCTCGTAGTCA |
| GATA3 | CTCATTAAGCCCAAGCGAAG | TCCTCCAGAGTGTGGTTGTG |
| HES1 | GCCTATTATGGAGAAAAGACG | CTATCTTTCTTCAGAGCATCC |
| HES4 | ATGACCGTGAGACACCTGC | AGACACTCGTGGAAGCCG |
| HES5 | TCAGCCCCAAAGAGAAAAAC | GCTTCAGCTGCTCGATGCT |
| HEY1 | TGGATCACCTGAAAATGCTG | CGAAATCCCAAACTCCGATA |
| HPRT1 | TGACACTGGCAAAACAATGCA | GGTCCTTTTCACCAGCAAGCT |

| | | |
|---|---|---|
| KLK10 | CTCTGGCGAAGCTGCTG | ATAGGCTTCGGGGTCCAA |
| KRT14 | GACCATTGAGGACCTGAGGA | CATACTTGGTGCGGAAGTCA |
| KRT15 | AGCCCAGAATGCGACTACAG | GCATTGTCGATCTCCAGGAT |
| KRT18 | GAGGCTGAGATCGCCACCTA | CCAAGGCATCACCAAGATTAAAG |
| NDRG2 | GAGATATGCTCTTAACCACCC | GCTGCCCAATCCATCCAA |
| NFIB | TTTGTGTCCAGCCACATCAT | GTGGCTTGGACTTCCTGATT |
| NOTCH1 | GTGACTGCTCCCTCAACTTCAAT | CTGTCACAGTGGCCGTCACT |
| NOTCH3 | CTACAATGGTGATAACTGTGAG | CAGTCATCCTCATTAATCTCG |
| PROM1 | ACCGACTGAGACCCAACATC | TGAACAGCACCTTGAAGAGCT |
| RBPJ | GAAGATGGCGCCTGTTGT | TACTGTTTGATCCCCTCGCT |
| TTYH1 | CACCAGTTGGTGGCACTG | GAGTAGCAGGAAGAGCAGGC |
| VGLL1 | GGCTCAGTTCACTATAAGAA | TATTTCCCAGGTGTCTCTAA |
| VIM | GCAGAAGAATGGTACAAATCCAAG | GTGAGGGACTGCACCTGTCTC |

**Figure Legends**

**Figure 1. Basal-like breast cancers harbor cell subpopulations with distinct differentiation states. (a)** Co-staining of K18 (green), K14 (red) and vimentin (white) in sections of normal human breast, and a luminal and triple-negative breast tumor. Double-positive $K18^+K14^+$ cells appear as yellow/orange. Scale bar = 50μm. **(b)** Percentages of cells stained positive for K18, K14 and vimentin and their combinations in a panel of 45 triple-negative breast tumors. The intratumoral Shannon diversity index value for each sample is shown below. Samples are shown in the order of their $K18^+K14^+$ content. **(c)** FACS analysis of K18, K14 and vimentin expression in indicated cell lines. Gates indicate $K18^+K14^+$ cells (blue) and $K18^+VIM^+$ cells (red), K18 staining for each sample is shown below, with gate indicating $K18^+$ range. **(d)** Percentages of cell subpopulations in cell lines representing different breast cancer subtypes, as analyzed by FACS. Ba.B – Basal B, HER2 – HER2 overexpressing, Ba/Myo - immortalized basal/myoepithelial. **(e)** FACS isolation of subpopulations of HCC70 cells subjected to mRNA-seq (left), and mRNA levels of K18, K14 and vimentin in the profiled subpopulations (right). Values are normalized to the levels in the $K18^+$ subpopulation. **(f)** Correlation values of the expression profiles of the $K18^+$, $K18^+K14^+$ and $K18^+VIM^+$ subpopulations, sorted as in (e) from HCC70 and MDA-MB-468 cells, with indicated lineage-associated gene signatures. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$ of Spearman correlation. **(g)** Section of patient-derived breast tumor xenograft co-stained for K18, K14 and vimentin (left), and qRT-PCR analysis of genes representing the luminal progenitor signature in $K18^+$ and $K18^+K14^+$ cell populations isolated from the tumor (right). Values indicate mean of three replicates ± s.e.m. Scale bar = 10μm. * $P <$ 0.05, Student's *t*-test.

**Figure 2. $K18^+14^+$ cells display increased tumorigenicity. (a)** Numbers of tumors formed by $GFP^-$ and $GFP^{high}$ subpopulations sorted from HCC70 cells expressing a K14 promoter-driven GFP reporter, after injection into mouse mammary glands (6 injections each) in limiting dilutions. Left, injected cell numbers. Bottom – calculated tumor-initiating cell (TIC) fraction in each population. ** $P < 0.005$ by ELDA algorithm. **(b)** Growth curves of tumors formed by $GFP^-$ and $GFP^{high}$ sorted HCC70 cells injected into the mammary glands of NOD-SCID mice. Values indicate mean volumes of $n = 9$ tumors ± s.e.m. **(c)** Percentage of $K18^+K14^+$ cells found in primary GFP-labelled MDA-MB-468 xenograft tumors and in lung metastases originating from them. Values indicate mean of tumors ± s.e.m. Lines connect tumors and lungs of same mice. Images show representative section of tumor section and metastatic cell in lung. **(d)** Images (left) and quantification (right) of the fraction of $K18^+K14^+$ cells in lung metastases seeded by MDA-MB-468 cell injection into mouse tail-veins, at the indicated times following injection. Values indicate mean fraction in lung three samples in each time point ± s.e.m. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$ by Student's *t*-test in panels b-d.

**Figure 3. Differentiation transitions occur through asymmetric cell divisions and are regulated by developmental transcription factors. (a)** Diagram illustrating the labelling procedure for identification of symmetric and asymmetric divisions. Single cells were seeded and treated with BrdU to label replicating cells, and cells were fixed and stained 18 hours later to identify cell doublets representing mitotic products. **(b)** Cell doublets stained for K18 (green), K14 (red) and BrdU (white) 18 hours following BrdU treatment. Images show representative symmetric divisions in which both mitotic products are $K18^+$ (top row) or both are $K18^+K14^+$ (middle row), and asymmetric divisions in which one daughter is $K18^+$ and the other $K18^+K14^+$ (bottom row). **(c)** Distribution of symmetric (S) and asymmetric (AS) cell divisions in HCC70

cells, scored from cells stained as in (b). *n* indicates the number of cell doublets scored. **(d)** FACS analysis of HCC70 cell subpopulation percentages following treatment with the EZH2 inhibitor GSK-126 (10 μM), or with DMSO (left). Gates indicate percentages of $VIM^+$ (red) and $K14^+$ (blue) cells, with all cells expressing $K18^+$. Right: $K14^+$ and $VIM^+$ population percentages following treatment with indicated inhibitor concentrations (right) for 3 days. Values indicate mean of three replicates ± s.e.m. * $P < 0.05$, Student's *t*-test. **(e)** Distribution of cell division types, as in (c), in HCC70 cells infected with shRNAs targeting EZH2 or EED, or with a control shRNA (shCont). *** $P < 0.0005$, chi-square. **(f)** Distribution of cell division types in HCC70 cells after treatment with GSK-126 or DMSO for three days. *** $P < 0.0005$ chi-square. **(g)** FACS analysis of cell subpopulation percentages in HCC70 cells expressing a control shRNA (shCont) or two different shRNAs targeting FOXA1. (**h)** Distribution of cell division types in HCC70 cells expressing control (shCont), FOXA1 or GATA3 shRNAs. *** $P < 0.0005$, chi-square.

**Figure 4. Functional screen identifies regulators of phenotypic heterogeneity. (a)** Diagram of screening procedure for the identification of regulators of tumor cell population composition. 517 shRNAs against 177 candidate regulators, as well as 6 different control shRNAs, were introduced into HCC70 in 96 well format. Effects on subpopulation fraction sizes were scored by FACS analysis following K18, K14 and vimentin co-staining. Genes whose silencing increased or decreased $K18^+K14^+$ cell numbers were retested in a validation screen and individually, and their effects on gene expression were tested by mRNA-seq. **(b)** Representative results of functional screen showing effects of shRNAs analyzed in one plate. Values indicate percentages of $K14^+$ cells in HCC70 cell infected with individual shRNA viruses (dots) each targeting the indicated candidate gene, or non-targeting controls (blue), and the median effect of shRNAs for

each gene (bars). The gray area marks the range of values measured in cells infected with control vectors, considered as baseline. Red dots indicate hairpins against genes whose silencing increased or decreased $K14^+$ cell numbers. >2 hairpins were screened for genes in which hairpins were not pre-validated for knockdown efficiency. **(c)** Relative size of $K14^+$ cell fraction in HCC70 cells silenced for the indicated regulators identified in the screen, shown normalized to control shRNA median value. Dots indicate individual shRNAs. Cells maintained K18 expression. **(d)** Correlations of gene expression profiles of cells silenced for each of the identified regulators (columns) with normal mammary lineage gene signatures. Red indicates increase in the expression of gene signature (positive correlation), blue indicates a decrease in signature expression (negative correlation), and white indicates no change. Regulators are clustered into five groups based on similarity of their effects on the expression signatures. **(e)** Diagram illustrating the effect of regulators in each of the clusters on each of the lineage signatures, derived from (d). ML – Mature luminal, LP – Luminal progenitor, Ba/MaSC – Basal/MaSC. Red arrows indicate positive regulation, blue lines indicate repression.

**Figure 5. The Notch pathway regulates $K18^+K14^+$ numbers and asymmetric divisions. (a)** FACS analysis of HCC70 cell population composition following infection with an shRNA targeting RBPJ (shRBPJ) or a control shRNA (shCont). **(b)** Relative expression of RBPJ, lineage markers and Notch targets assessed by qRT-PCR in HCC70 cells infected with shCont or shRBPJ vectors. Values indicate mean of three replicates ± s.e.m. **(c)** FACS analysis of HCC70 subpopulations after infection with shCont or shNOTCH1. **(d)** Relative expression of NOTCH1, lineage markers and Notch targets assessed by qRT-PCR in HCC70 cells infected with shCont or shNOTCH1 vectors. **(e)** FACS analysis of HCC70 subpopulations after treatment for 3 days with the γ-secretase inhibitor (GSI) LY-411575 (200nM) or DMSO, and quantification of $K14^+$

percentages following treatment with indicated inhibitor concentrations (right). Values indicate mean of three replicates ± s.e.m. **(f)** Relative expression of NOTCH1, lineage markers and Notch targets assessed by qRT-PCR in HCC70 cells treated with LY-411575. **(g)** Relative expression of GFP, Notch targets and lineage markers assessed by qRT-PCR in HCC70 infected with a NICD1 overexpression retroviral vector. **(h)** Distribution of cell division types in HCC70 cells after three days of treatment with LY-411575 GSI or DMSO. *n* = cell doublets scored. **(i)** Relative expression of EZH2, lineage markers and Notch targets assessed by qRT-PCR in HCC70 cells treated with GSK-126. **(j)** $K14^+$ cell percentages in HCC70 cells expressing an shRNA against NFIB, treated with GSK-126 for EZH2 inhibition, treated with the LY-411575 (GSI) for Notch inhibition, or treated with combinations of these inhibitors. Cells expressing shCont were used as controls for shNFIB cells in all experiments. Values indicate mean of FACS-scored values of three replicates ± s.e.m. **(k)** Relative expression of NFIB, lineage markers and Notch targets assessed by qRT-PCR in HCC70 cells infected with shCont or shNFIB vectors. **(l)** Diagram illustrating the effects of indicated regulators on the composition of basal-like breast cancer cell populations. EZH2, NFIB and NOTCH increase $K18^+K14^+$ symmetric division rates and generate higher progenitor like content, while FOXA1 and GATA3 increase asymmetric divisions, reducing the content of these cells. Throughout, * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$. Student's *t*-test except (h) chi-square.

**Supplementary Figure Legends**

**Supplementary Figure 1. Expression profiles of basal-like breast cancer subpopulations. (a)** Principle component analysis of the expression profiles of subpopulations isolated from HCC70 and MDA-MB-468 cells stained for K18, K14 and vimentin. Corresponding populations from the two cell lines are marked with the same color. **(b)** Relative mRNA expression levels of representative genes included in lineage expression signatures in sorted sub-populations of HCC70 cells. Values are normalized to levels in $K18^+$ cells. **(c)** qRT-PCR analysis of genes representing the luminal progenitor signature in $K18^+$ and $K18^+K14^+$ cell populations isolated from a second PDX tumor. Values indicate mean of three replicates ± s.e.m. * $P < 0.05$, ** $P < 0.005$, Student's *t*-test.

**Supplementary Figure 2. K14 reporter plasmid allows enrichment for $K18^+K14^+$ cells. (a)** Diagram of the K14-promoter driven GFP reporter vector. **(b)** FACS analysis displaying the expression of GFP under the control of the K14 promoter in HCC70 cells, and the gating used to sort $GFP^-$ and $GFP^{high}$ cell populations. **(c)** Images of $GFP^{high}$ sorted HCC70 cells stained for K14 to assess the purity of the separated populations (left), and quantification of $K14^+$ cells in sorted $GFP^-$ and $GFP^{high}$ cell populations (right). **(d)** Relative mRNA levels of K14 and additional differentiation markers in $GFP^-$ and $GFP^{high}$ cell populations assessed by qRT-PCR. Values indicate mean of three replicates. * $P$ value < 0.05, ** $P$ value < 0.005

**Supplementary Figure 3. Detection of mitotic cell pairs and asymmetric divisions. (a)** Determination of time-point for detection of mitotic cell doublets following BrdU labelling of cells. HCC70 cells were sparsely seeded, treated with BrdU for one hour and fixed stained for BrdU at the indicated subsequent time points. BrdU-labelled single cells and doublets were scored. No labelled doublets are detected 6 hours after treatment, and the largest number was

detected 18 hours after treatment. **(b)** Images of HCC70 cells labelled with Cerulean- or mCherry-expressing viruses, mixed together in a 1:1 ratio, sparsely seeded, and imaged 2 hours later (left). Cell doublets were allowed to replicate, and were imaged again 18 hours after seeding (center), and the numbers of same-color and different-color doublets were quantified (right), demonstrating rarity of randomly appearing doublets that are not mitotic products. **(c)** Distribution of symmetric and asymmetric cell divisions involving vimentin expression. HCC70 cells were stained for vimentin following the BrdU labelling pulse-chase protocol (right), and mitotic product doublets were scored for vimentin expression. All divisions were found to be symmetric, of two $VIM^+$ or two $VIM^-$ products (left).

**Supplementary Figure 4. The EZH2 inhibitor GSK-126 reduces H3K27me3 levels with limited effect on cell viability. (a)** Quantification of H3K27me3 levels in HCC70 cells treated with the indicated doses of GSK-126 (left) measured by FACS as shown on the right. a.u. – arbitrary units. **(b)** Percentage of Annexin V positive cells measured by FACS two days after treatment with the indicated concentrations of GSK-126.

**Supplementary Figure 5. FOXA1 expression in basal-like breast cancers. (a)** Relative expression of FOXA1 in breast cancers of different subtypes included in the TCGA collection. LumA = Luminal A, $n = 332$; LumB = Luminal B, $n = 217$; HER2 $n = 119$; Basal-like $n = 169$. Values are normalized to mean of all samples = 0. **(b)** Pearson correlation between the mRNA level of FOXA1 and that of EZH2, K14 or GATA3 in 169 basal-like breast tumors included in the TCGA database. Values indicate correlation coefficients. **(c)** Immunohistochemical staining of MCF7 and HCC70 xenograft tumors for FOXA1, showing expression of FOXA1 in the basal-like HCC70 tumors. Scale bar = 100μm. **(d)** Relative mRNA levels of EZH2, K14 and the luminal regulators FOXA1 and GATA3 in HCC70 cells infected with shCont or shEZH2

vectors. **(e)** Relative mRNA levels of FOXA1 and lineage markers assessed by qRT-PCR in HCC70 cells infected with shCont or shFOXA1 vectors.

**Supplementary Figure 6. High throughput screen for identification of factors regulating tumor cell population composition.** Grouped presentation of results of the first round of the. Shown are relative numbers of $K14^+$ cells as measured by FACS in HCC70 cells expressing shRNAs (dots) against all genes included in the screen (x axis). Values indicate Z-scores calculated and normalized per each plate. Non-targeting control shRNAs included in all careen plates are labelled green. Horizontal bars indicate the median values for the shRNAs targeting each gene. Blue area marks the range of Z score ± 1 measured in cells infected with control vectors, which was considered as baseline.

**Supplementary Figure 7. The Notch pathway promotes tumor growth and is regulated by EZH2 and NFIB. (a)** Percentage of Annexin V positive cells measured by FACS two days after treatment with the indicated concentrations of the γ-secretase inhibitor LY-411757. **(b)** Growth curves of xenograft tumors formed by HCC70 cells infected with shCont or shRBPJ vectors and injected into the mammary glands of NOD-SCID mice. Values indicate mean tumor volume ± s.e.m. *n* = number of tumors in each group. **(c)** Effects of silencing of indicated regulators on the Notch expression signature, shown as the significance level of the positive correlation between the gene expression profiles of indicated cells and the Notch signature. **(d)** Relative mRNA expression of EZH2 in HCC70 cells silenced for NOTCH1 or treated with GSI.

**Supplementary Figure 8. NFIB expression in breast cancers and in normal mammary lineages. (a)** Relative expression of NFIB in breast cancers of different subtypes included in the TCGA collection. LumA = Luminal A, *n* = 332; LumB = Luminal B, *n* = 217; HER2 *n* = 119; Basal-like *n* = 169. Values are normalized to mean across all samples = 0. **(b)** Relative mRNA

levels of NFIB in different cell types of the normal breast lineage as profiled by Lim et al.[17]. Circles represent individual samples, vertical line indicates the mean of samples ± s.d.

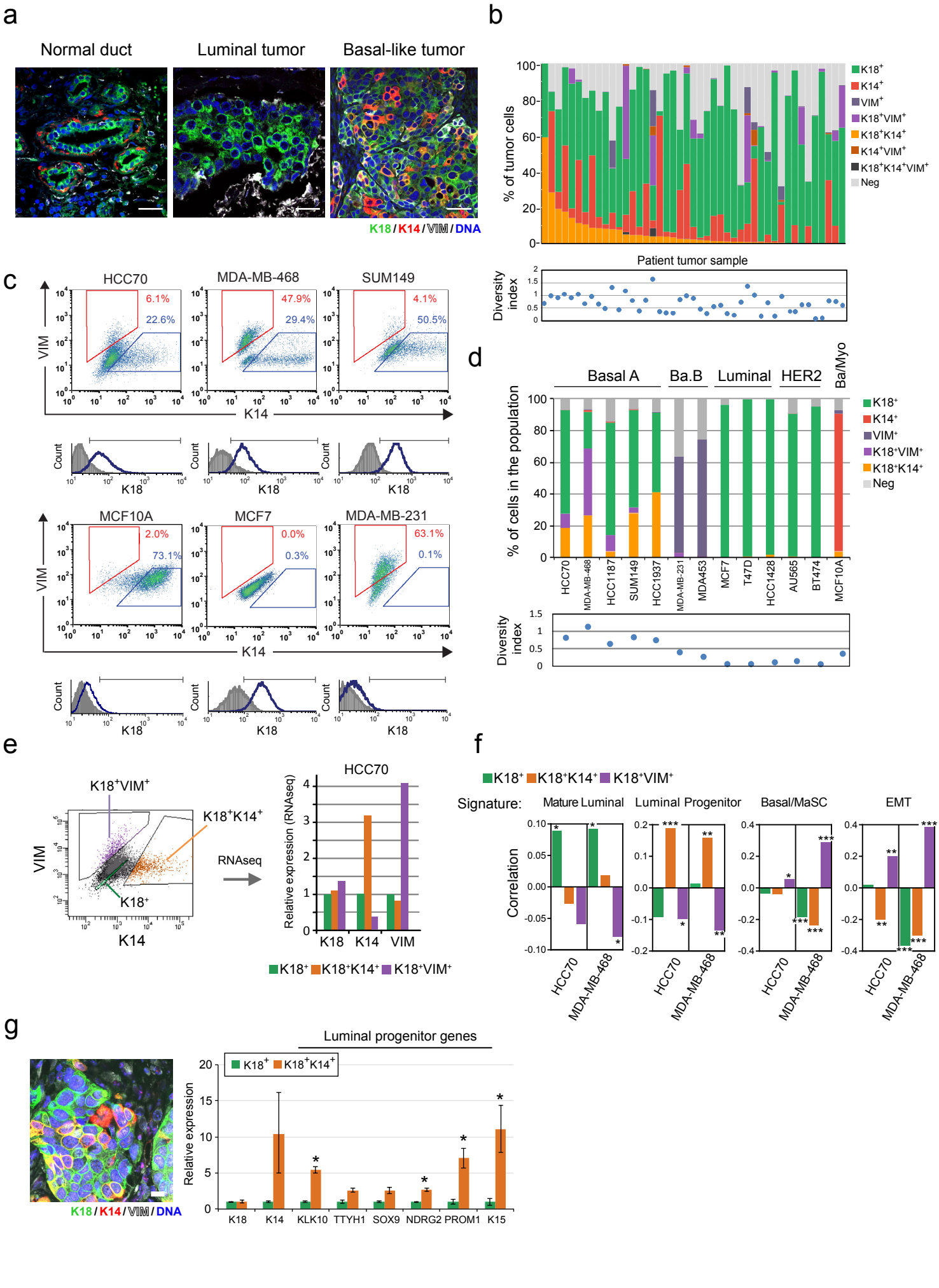
**Supplementary Table 1. Patient tumor samples. (a)** Tumor sections stained and scored for K18, K14 and vimentin expression. **(b)** Patient derived tumor xenografts grown for isolation of cell subpopulations.

**Supplementary Table 2. Signature genes. (a)** Normal breast lineage signatures. **(b)** EMT signature. **(c)** GSI-treated signature.

**Supplementary Table 3. Signature genes showing positive correlations in the expression profiles of subpopulations.** Listed are genes included in the indicated lineage signatures whose expression positively correlated with the expression profiles of subpopulations from both HCC70 and MDA-MB-468 lines.

**Supplementary Table 4. shRNA vectors used in the study.**

# Figure 1


a
Normal duct
Luminal tumor
Basal-like tumor
K18 / K14 / VIM / DNA
b
% of tumor cells
K18+
K14+
VIM+
K18+VIM+
K18+K14+
K14+VIM+
K18+K14+VIM+
Neg
Patient tumor sample
Diversity index
c
HCC70
MDA-MB-468
SUM149
MCF10A
MCF7
MDA-MB-231
6.1%
22.6%
47.9%
29.4%
4.1%
50.5%
2.0%
73.1%
0.0%
0.3%
63.1%
0.1%
VIM
K14
Count
K18
d
Basal A
Ba.B
Luminal
HER2
Ba/Myo
% of cells in the population
HCC70
MDA-MB-468
HCC1187
SUM149
HCC1937
MDA-MB-231
MDA453
MCF7
T47D
HCC1428
AU565
BT474
MCF10A
Diversity index
e
K18+VIM+
K18+K14+
K18+
VIM
K14
RNAseq
HCC70
Relative expression (RNAseq)
K18
K14
VIM
f
K18+ K18+K14+ K18+VIM+
Signature:
Mature Luminal
Luminal Progenitor
Basal/MaSC
EMT
Correlation
HCC70
MDA-MB-468
g
K18 / K14 / VIM / DNA
Luminal progenitor genes
K18+ K18+K14+
Relative expression
K18
K14
KLK10
TTYH1
SOX9
NDRG2
PROM1
K15

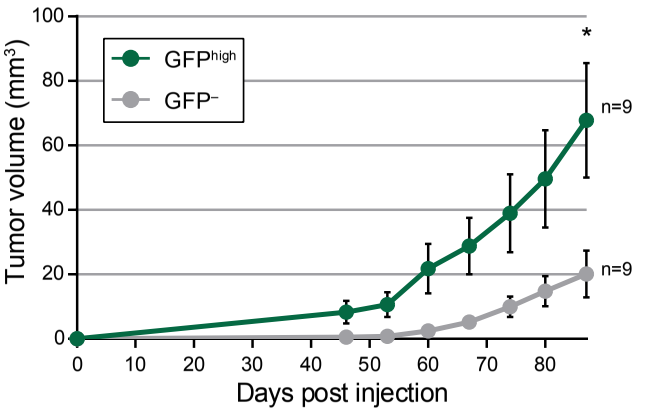
# Figure 2

a

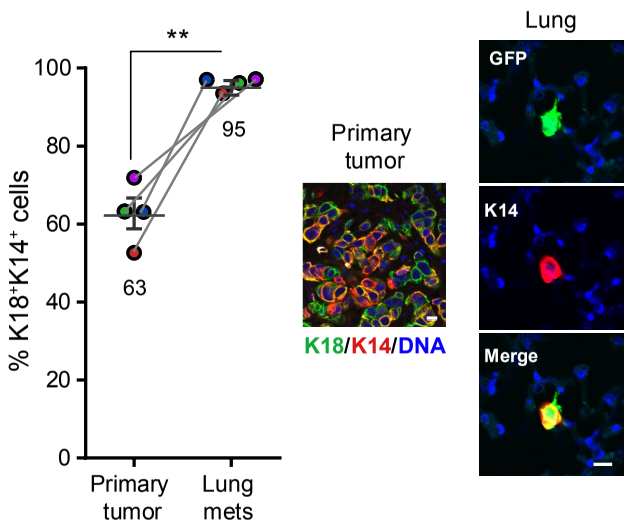
HCC70 K14p-GFP

| Cells injected | GFP⁻ | GFP$^{high}$ |
|---|---|---|
| $10^6$ | 6/6 | 6/6 |
| $10^4$ | 3/6 | 6/6 |
| $10^3$ | 0/6 | 3/6 |
| TIC frequency: | 1:16,085 | 1:1,424 |
| | ** | |

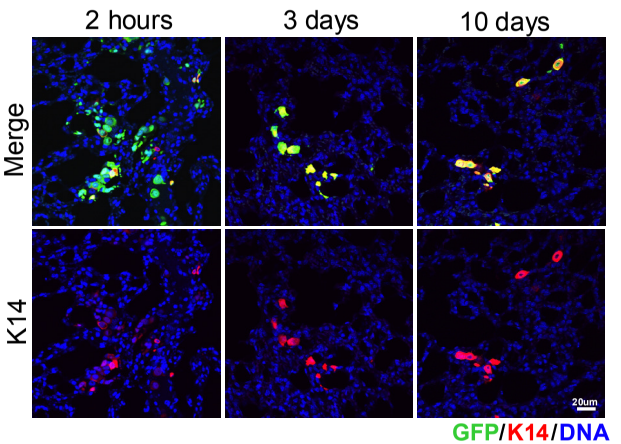
b




c




d



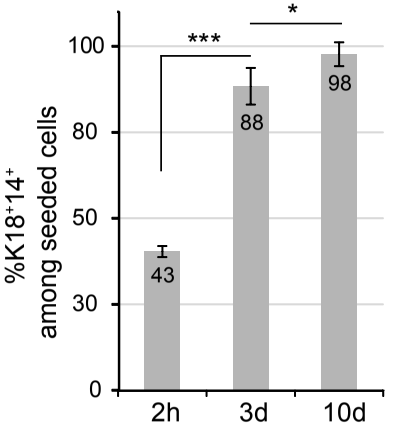



Figure 3

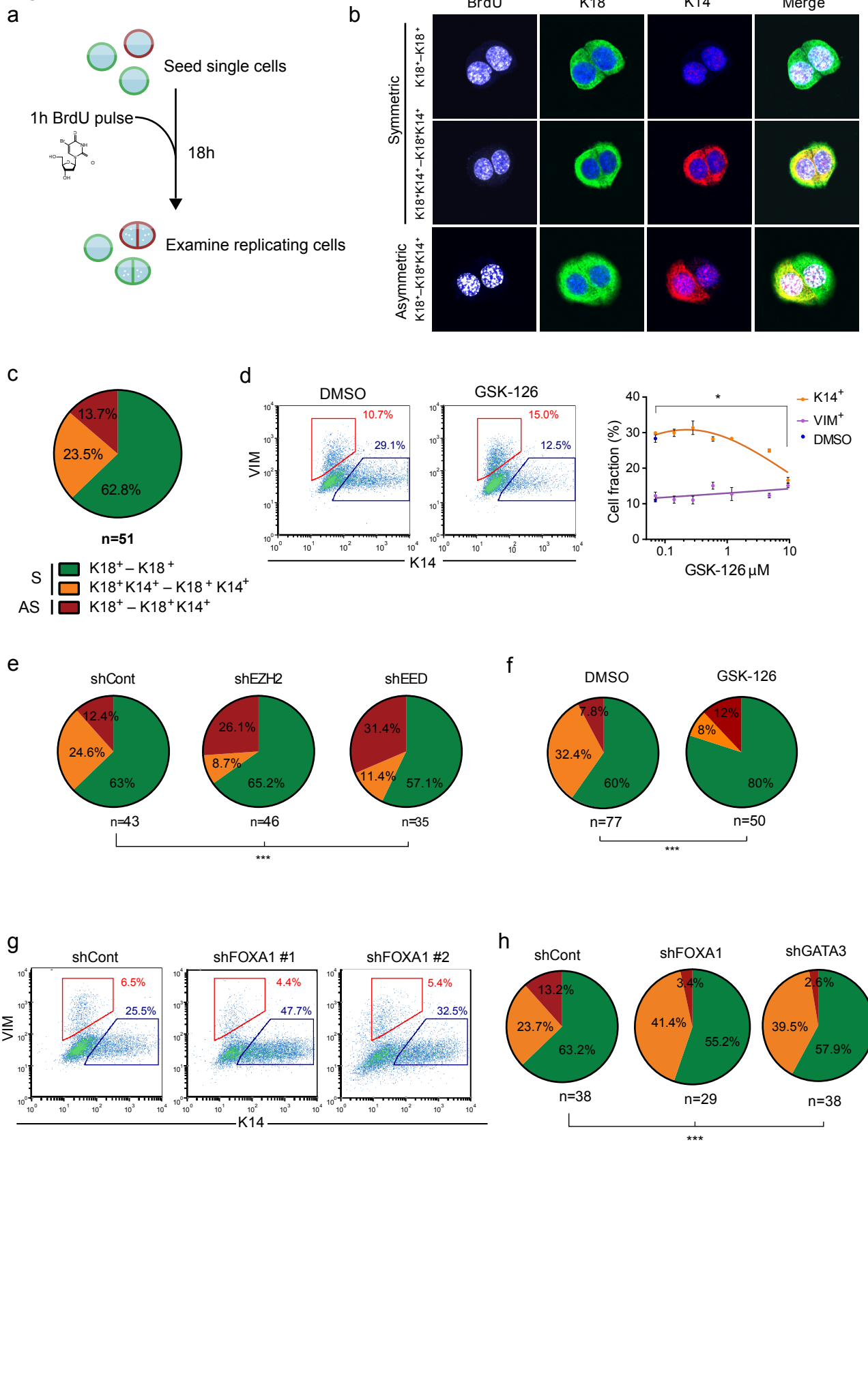
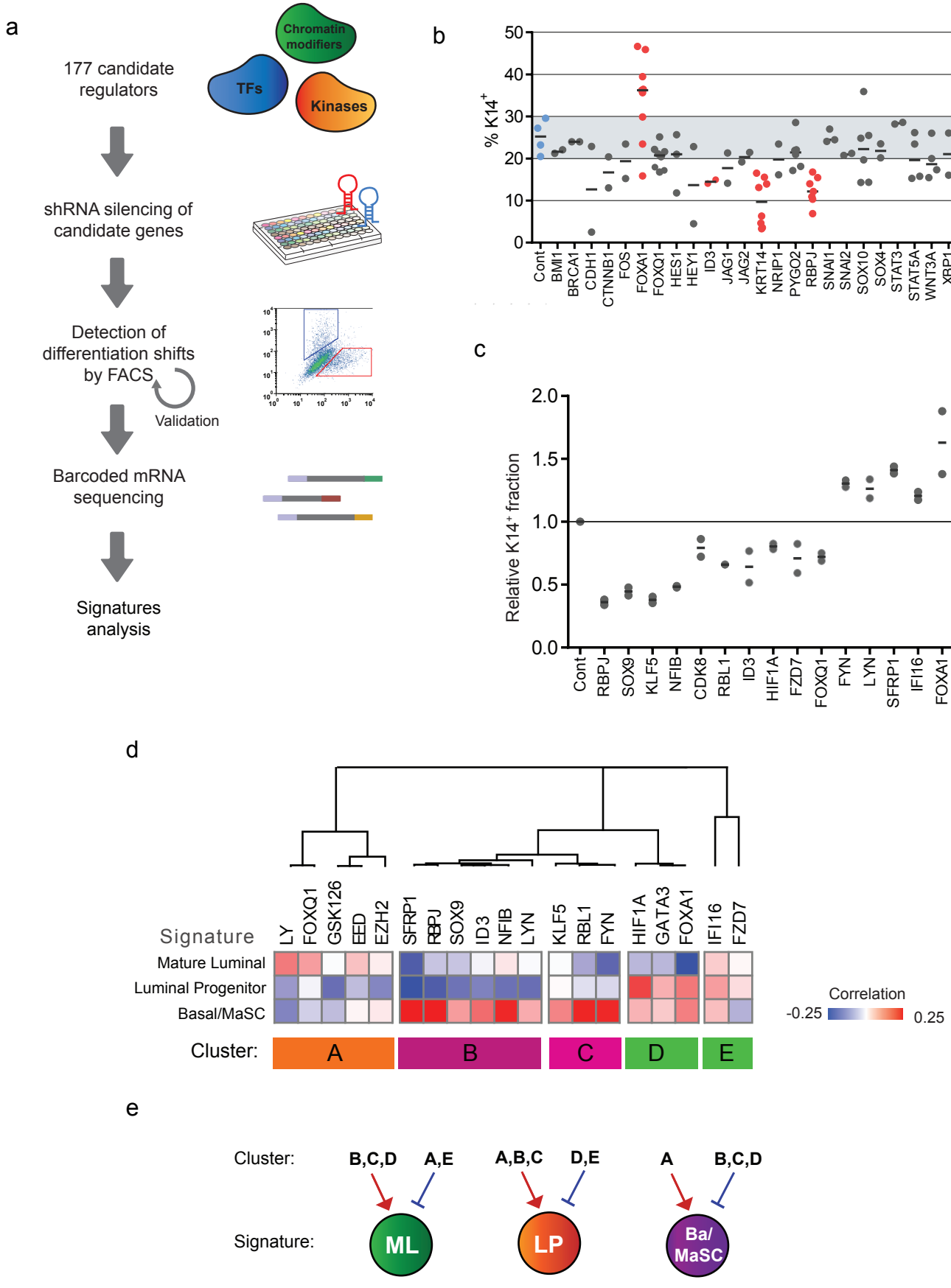

a
Seed single cells
1h BrdU pulse
18h
Examine replicating cells
b
BrdU
K18
K14
Merge
Symmetric
K18+–K18+
K18+K14+–K18+K14+
Asymmetric
K18+–K18+K14+
c
13.7%
23.5%
62.8%
n=51
S
AS
K18+ – K18+
K18+K14+ – K18+K14+
K18+ – K18+K14+
d
DMSO
GSK-126
10.7%
29.1%
15.0%
12.5%
VIM
K14
Cell fraction (%)
GSK-126 μM
K14+
VIM+
DMSO
e
shCont
shEZH2
shEED
12.4%
24.6%
63%
26.1%
8.7%
65.2%
31.4%
11.4%
57.1%
n=43
n=46
n=35
***
f
DMSO
GSK-126
7.8%
32.4%
60%
12%
8%
80%
n=77
n=50
***
g
shCont
shFOXA1 #1
shFOXA1 #2
6.5%
25.5%
4.4%
47.7%
5.4%
32.5%
VIM
K14
h
shCont
shFOXA1
shGATA3
13.2%
23.7%
63.2%
3.4%
41.4%
55.2%
2.6%
39.5%
57.9%
n=38
n=29
n=38
***

Figure 4


a
177 candidate regulators
Chromatin modifiers
TFs
Kinases
shRNA silencing of candidate genes
Detection of differentiation shifts by FACS
Validation
Barcoded mRNA sequencing
Signatures analysis
b
% K14+
Cont
BMI1
BRCA1
CDH1
CTNNB1
FOS
FOXA1
FOXQ1
HES1
HEY1
ID3
JAG1
JAG2
KRT14
NRIP1
PYGO2
RBPJ
SNAI1
SNAI2
SOX10
SOX4
STAT3
STAT5A
WNT3A
XBP1
c
Relative K14+ fraction
Cont
RBPJ
SOX9
KLF5
NFIB
CDK8
RBL1
ID3
HIF1A
FZD7
FOXQ1
FYN
LYN
SFRP1
IFI16
FOXA1
d
Signature
Mature Luminal
Luminal Progenitor
Basal/MaSC
LY
FOXQ1
GSK126
EED
EZH2
SFRP1
RBPJ
SOX9
ID3
NFIB
LYN
KLF5
RBL1
FYN
HIF1A
GATA3
FOXA1
IFI16
FZD7
Cluster:
A
B
C
D
E
Correlation
-0.25
0.25
e
Cluster:
B,C,D
A,E
A,B,C
D,E
A
B,C,D
Signature:
ML
LP
Ba/ MaSC

# Figure 5

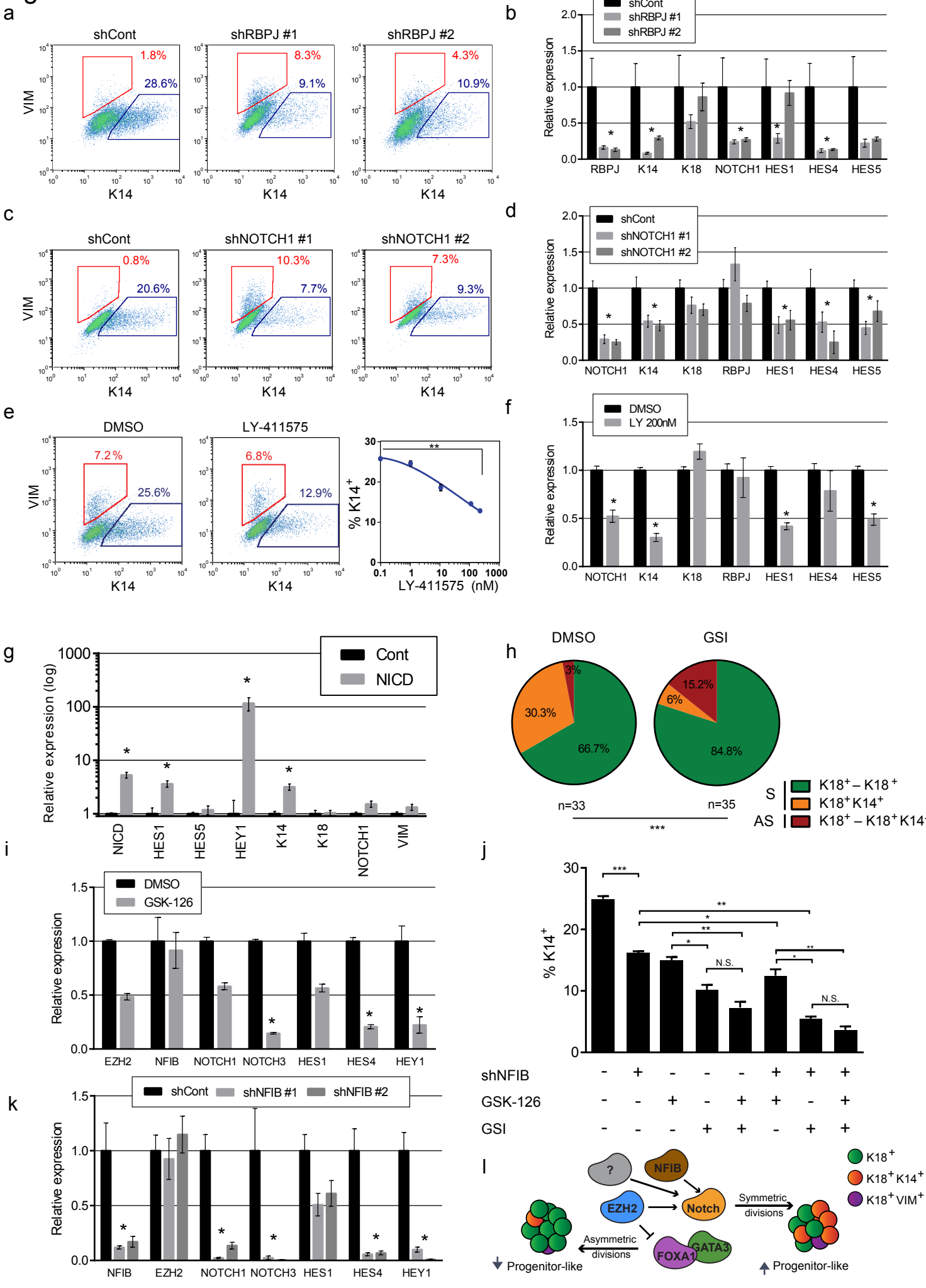
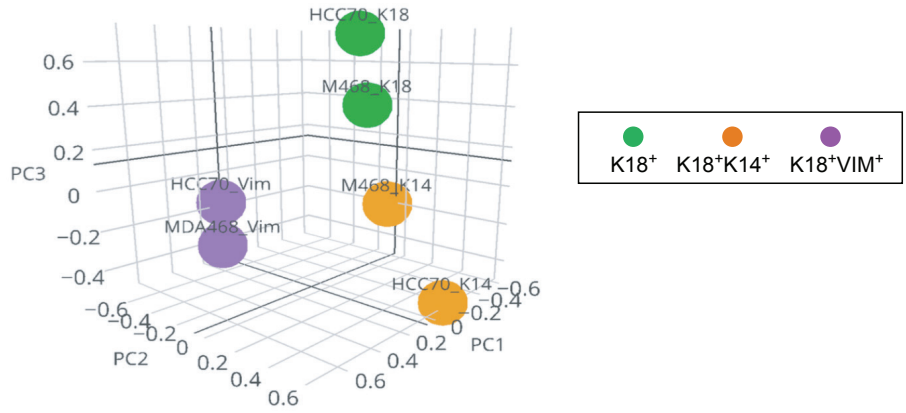
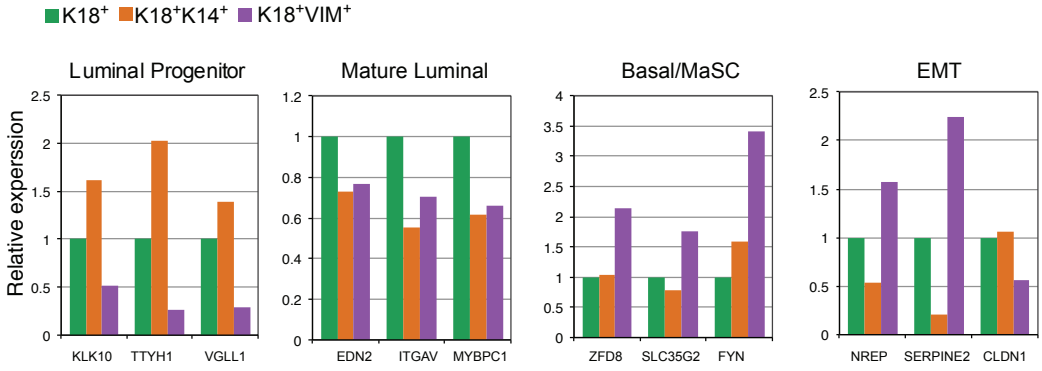

a
shCont
shRBPJ #1
shRBPJ #2
1.8%
28.6%
8.3%
9.1%
4.3%
10.9%
VIM
K14
b
shCont
shRBPJ #1
shRBPJ #2
Relative expression
RBPJ
K14
K18
NOTCH1
HES1
HES4
HES5
c
shCont
shNOTCH1 #1
shNOTCH1 #2
0.8%
20.6%
10.3%
7.7%
7.3%
9.3%
d
shCont
shNOTCH1 #1
shNOTCH1 #2
e
DMSO
LY-411575
7.2%
25.6%
6.8%
12.9%
% K14+
LY-411575 (nM)
f
DMSO
LY 200nM
g
Cont
NICD
Relative expression (log)
h
DMSO
GSI
3%
30.3%
66.7%
15.2%
6%
84.8%
n=33
n=35
S
AS
K18+ – K18+
K18+K14+
K18+ – K18+K14+
i
DMSO
GSK-126
EZH2
NFIB
NOTCH1
NOTCH3
HES1
HES4
HEY1
j
% K14+
N.S.
shNFIB
GSK-126
GSI
k
shCont
shNFIB #1
shNFIB #2
l
?
NFIB
EZH2
Notch
GATA3
FOXA1
Symmetric divisions
Asymmetric divisions
Progenitor-like
Progenitor-like
K18+
K18+K14+
K18+VIM+

Figure S1

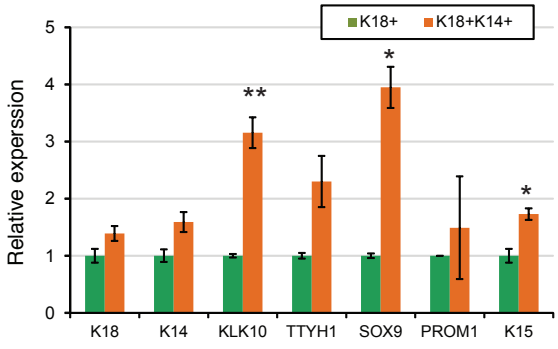
a



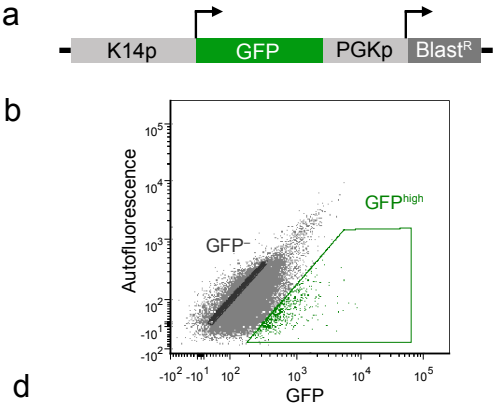

b




c

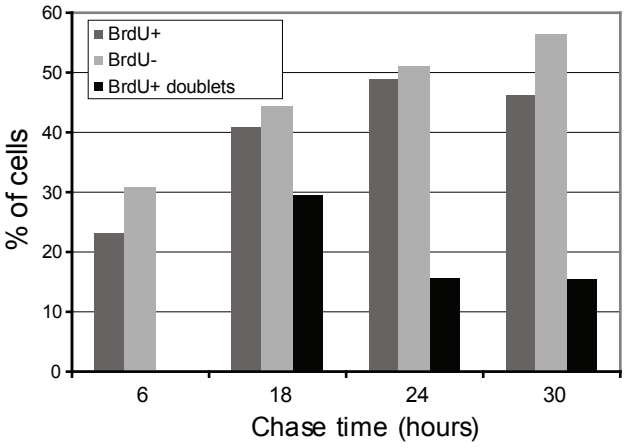


# Figure S2

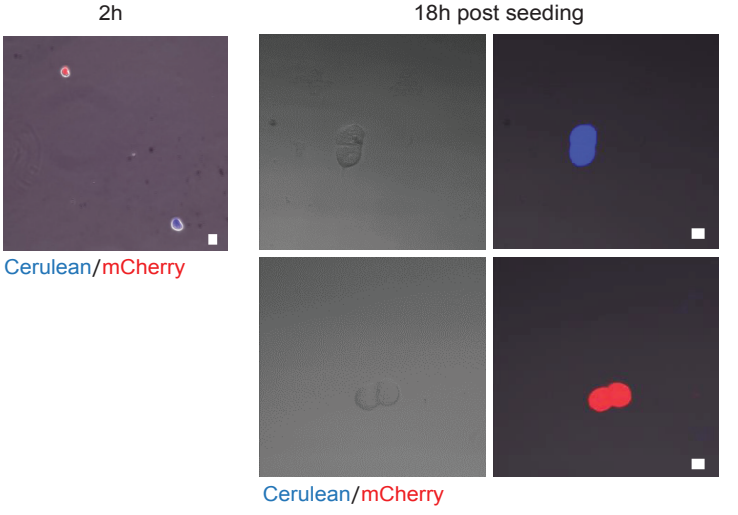

a
K14p
GFP
PGKp
Blast^R
b
Autofluorescence
GFP
GFP⁻
GFP^high
c
GFP
Merge
GFP/K14/DAPI
%K14 positive
GFP⁻
GFP high
d
Relative expression
GFP⁻
GFP^high
**
*
K14
K5
K18
K8
CDH1
VIM

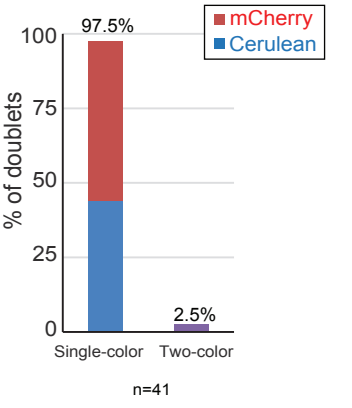
# Figure S3

a

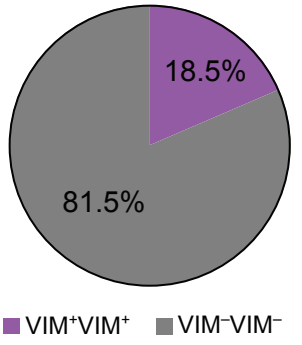



b




c



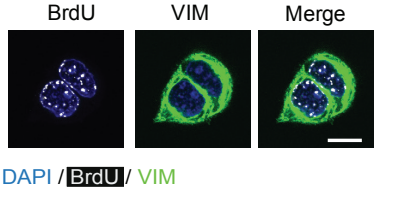
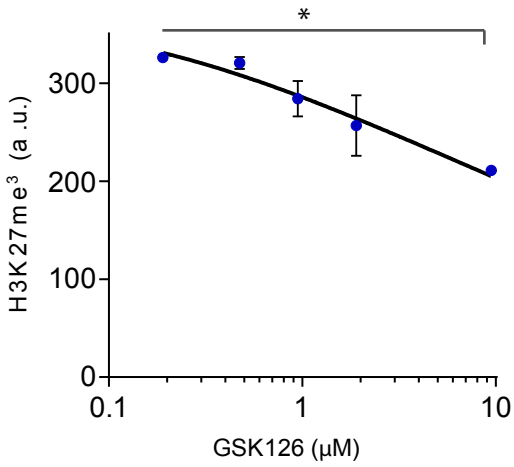
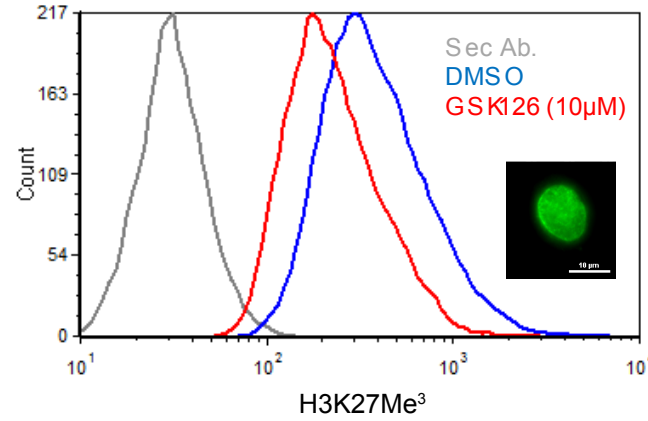



Figure S4

a


*
H3K27me3 (a.u.)
300
200
100
0
0.1
1
10
GSK126 (μM)


217
163
109
54
0
Count
Sec Ab.
DMSO
GSK126 (10μM)
10 μm
10^1
10^2
10^3
10^4
H3K27Me3

b

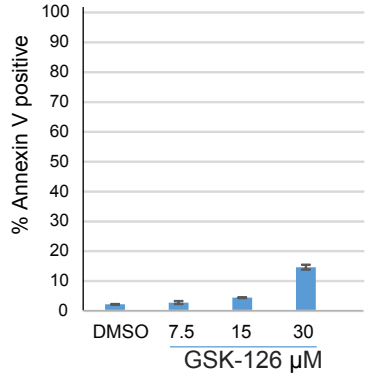

100
90
80
70
60
50
40
30
20
10
0
% Annexin V positive
DMSO
7.5
15
30
GSK-126 μM

Figure S5

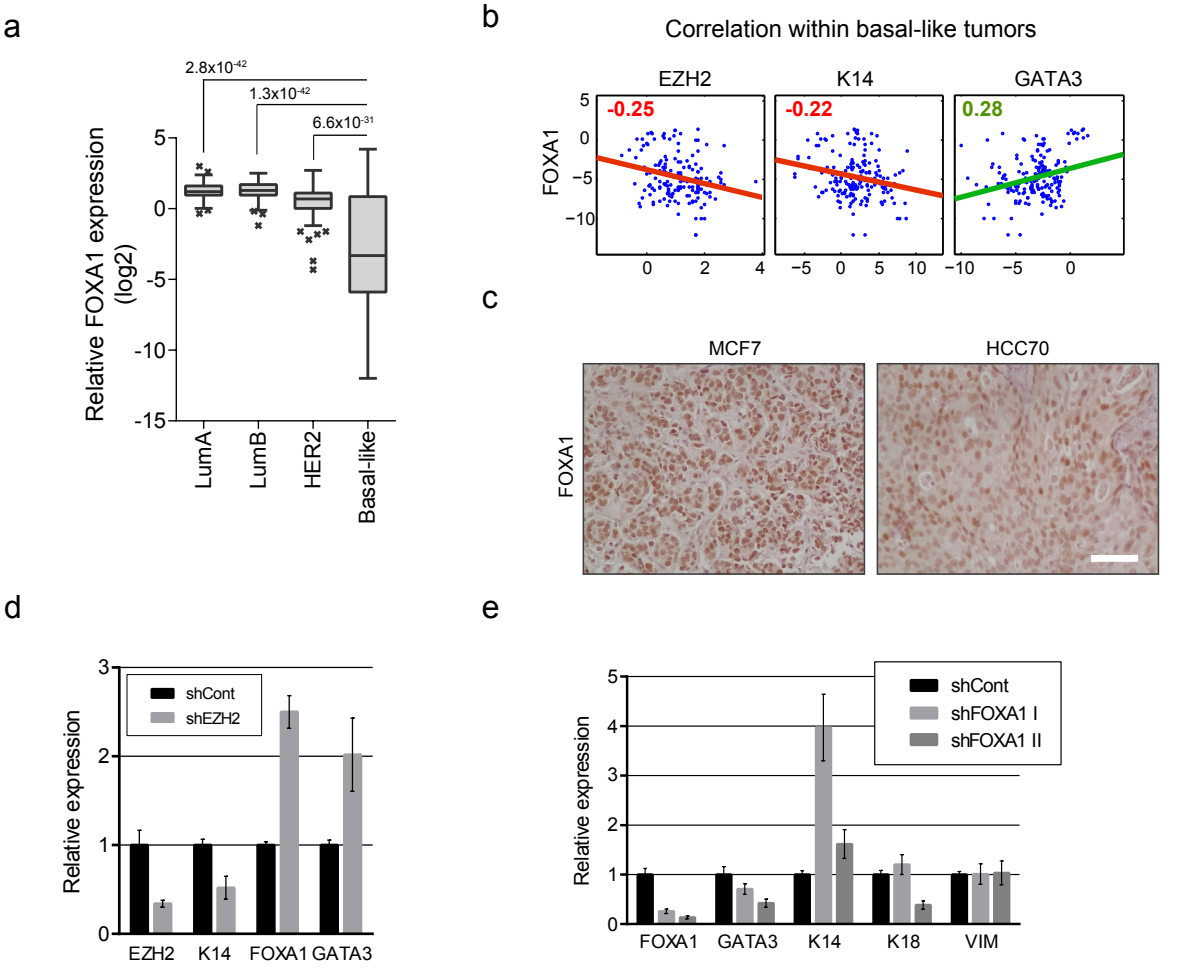
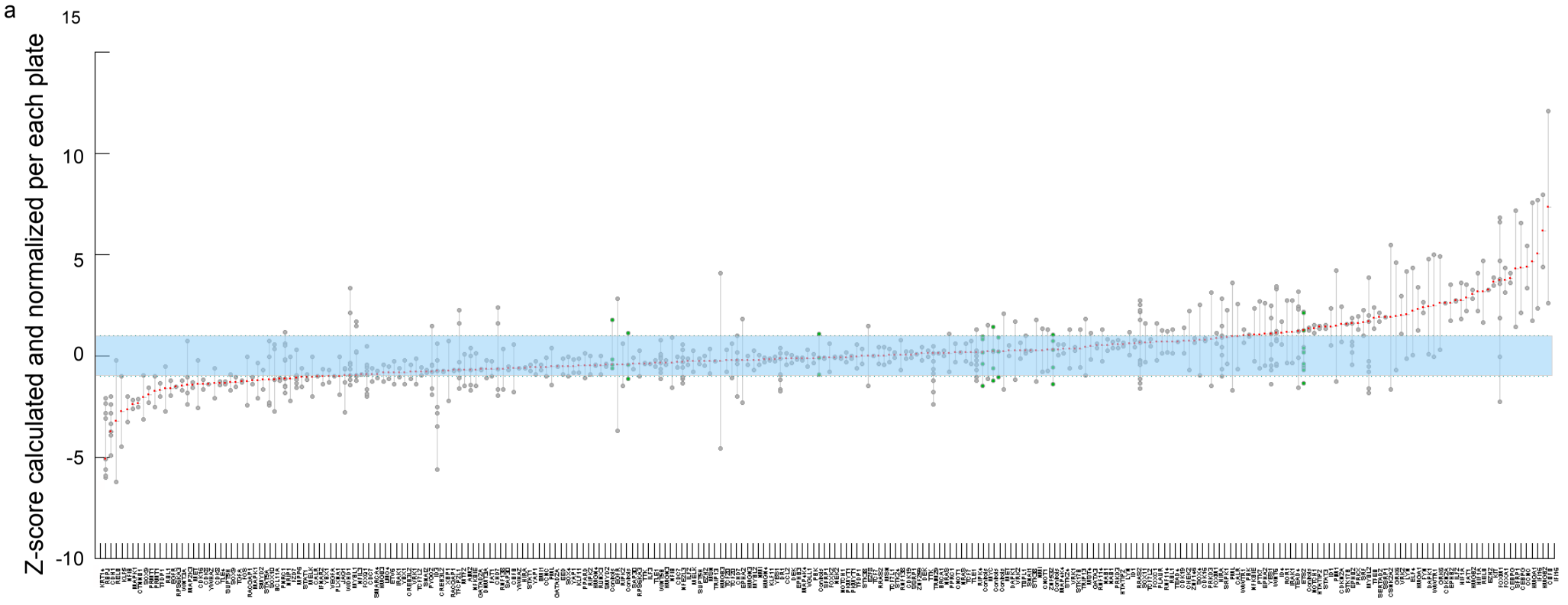

a
2.8x10-42
1.3x10-42
6.6x10-31
Relative FOXA1 expression (log2)
5
0
-5
-10
-15
LumA
LumB
HER2
Basal-like
b
Correlation within basal-like tumors
EZH2
K14
GATA3
FOXA1
-0.25
-0.22
0.28
c
MCF7
HCC70
FOXA1
d
shCont
shEZH2
Relative expression
EZH2
K14
FOXA1
GATA3
e
shCont
shFOXA1 I
shFOXA1 II
Relative expression
FOXA1
GATA3
K14
K18
VIM

Figure S6


a
15
10
5
0
-5
-10
Z-score calculated and normalized per each plate

# Figure S7

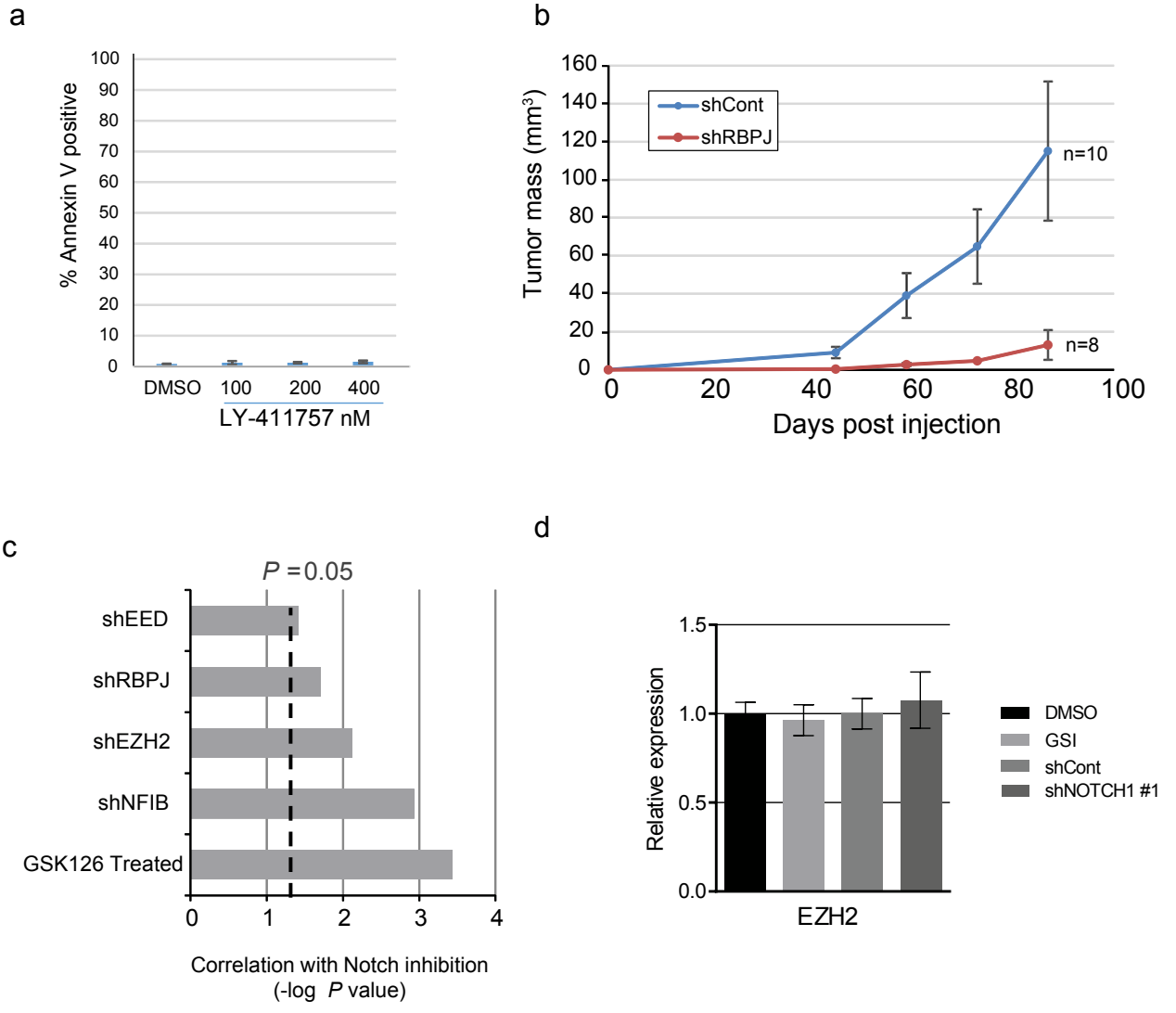
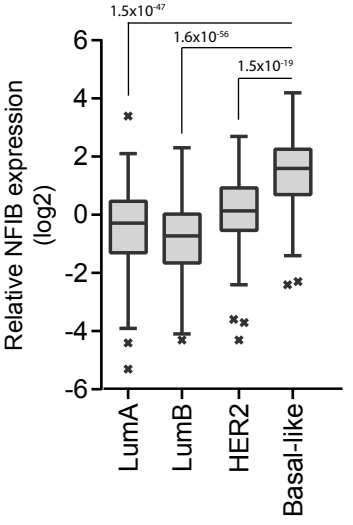
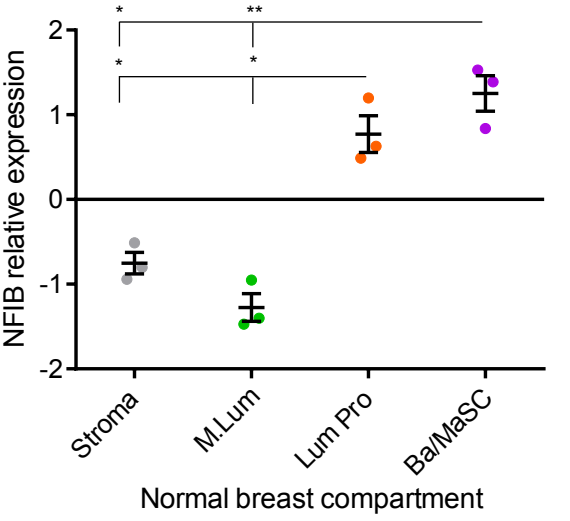

a
% Annexin V positive
100
90
80
70
60
50
40
30
20
10
0
DMSO
100
200
400
LY-411757 nM
b
Tumor mass (mm3)
160
140
120
100
80
60
40
20
0
shCont
shRBPJ
n=10
n=8
0
20
40
60
80
100
Days post injection
c
P =0.05
shEED
shRBPJ
shEZH2
shNFIB
GSK126 Treated
0
1
2
3
4
Correlation with Notch inhibition
(-log P value)
d
Relative expression
1.5
1.0
0.5
0.0
EZH2
DMSO
GSI
shCont
shNOTCH1 #1

Figure S8

a


1.5x10-47
1.6x10-56
1.5x10-19
Relative NFIB expression (log2)
6
4
2
0
-2
-4
-6
LumA
LumB
HER2
Basal-like

b


*
**
*
*
NFIB relative expression
2
1
0
-1
-2
Stroma
M.Lum
Lum Pro
Ba/MaSC
Normal breast compartment

# Chapter 3: RNF20 and histone H2B ubiquitylation exert opposing effects in basal-like versus luminal breast cancer

Status: accepted to Cell Death & Differentiation (2016)

O Tracic*, R.Z. Granit*, Ioannis S. Pateras, Hadas Mansury, Bella Maly, Vassilis G. Gorgoulis, Eli Pikarsky, Ittai Ben-Porath & Moshe Oren

* Equal contribution

**RNF20 and histone H2B ubiquitylation exert opposing effects in basal-like versus luminal breast cancer**

Ohad Tarcic[1,*], Roy Z. Granit[2,*], Ioannis S. Pateras[3], Hadas Masury[2], Bella Maly[4], Yaara Zwang[5], Yosef Yarden[5], Vassilis G. Gorgoulis[3,6,7], Eli Pikarsky[8], Ittai Ben-Porath[2#] & Moshe Oren[1#]

1. Department of Molecular Cell Biology, The Weizmann Institute, Rehovot 7610001, Israel
2. Department of Developmental Biology and Cancer Research, Institute for Medical Research – Israel-Canada, Hadassah School of Medicine, The Hebrew University of Jerusalem, Jerusalem 91120, Israel
3. Molecular Carcinogenesis Group, Dept. of Histology-Embryology, School of Medicine, University of Athens, 11527 Athens, Greece
4. Department of Pathology, Hadassah Medical Center, Jerusalem 91120, Israel
5. Department of Biological Regulation, The Weizmann Institute, Rehovot 7610001, Israel
6. Biomedical Research Foundation, Academy of Athens, 11527 Athens, Greece
7. Biomedical Research Foundation, Academy of Athens, 11527 Athens, Greece
8. Faculty Institute for Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, M13 9WL, UK
9. Department of Immunology & Cancer Research and Department of Pathology, Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel

* These authors contributed equally

# Corresponding Authors: Ittai Ben-Porath, E-mail: ittaibp@mail.huji.ac.il, and Moshe Oren, E-mail: moshe.oren@weizmann.ac.il

Keywords: Breast cancer, H2Bub1, RNF20, estrogen receptor, cytokines

Running title: RNF20 has opposing roles in breast cancer subtypes

**Abstract**

Breast cancer subtypes display distinct biological traits that influence their clinical behavior and response to therapy. Recent studies have highlighted the importance of chromatin structure regulators in tumorigenesis. The RNF20-RNF40 E3 ubiquitin ligase complex monoubiquitylates histone H2B to generate H2Bub1, while the deubiquitinase (DUB) USP44 can remove this modification. We found that RNF20 and RNF40 expression and global H2Bub1 are relatively low, and USP44 expression is relatively high, in basal-like breast tumors compared to luminal tumors. Consistent with a tumor-suppressive role, silencing of RNF20 in basal-like breast cancer cells increased their proliferation and migration, and their tumorigenicity and metastatic capacity, partly through upregulation of inflammatory cytokines. In contrast, in luminal breast cancer cells, RNF20 silencing reduced proliferation, migration and tumorigenic and metastatic capacity, and compromised estrogen receptor transcriptional activity, indicating a tumor promoting role. Notably, the effects of USP44 silencing on proliferation and migration in both cancer subtypes were opposite to those of RNF20 silencing. Hence, RNF20 and H2Bub1 have contrasting roles in distinct breast cancer subtypes, through differential regulation of key transcriptional programs underpinning the distinctive traits of each subtype.

## Introduction

Ductal carcinomas of the breast are grouped into several main subtypes, each displaying distinct phenotypic features and clinical behavior (Rakha, El-Sayed et al. 2009; Prat and Perou 2011; TCGA 2012). Estrogen receptor-positive (ER+) tumors of the Luminal A and B subtypes are often responsive to anti-estrogenic treatment and display favorable clinical outcomes. In contrast, triple-negative tumors of the basal-like subtype are typically poorly differentiated and highly aggressive (Rakha, El-Sayed et al. 2009; Prat, Adamo et al. 2013). While the mutation profiles associated with these subtypes have been characterized extensively (TCGA 2012; Ellis and Perou 2013), the underlying molecular programs dictating their distinct traits are still not fully understood.

Post-translational histone modifications impinge on all aspects of chromatin function, and are extensively implicated in cancer development (Campos and Reinberg 2009). Monoubiquitylation of histone H2B (H2Bub1) on Lys120 is executed mainly by the E3 ubiquitin ligase complex comprised of RNF20 and RNF40 (Osley 2006; Kim, Guermah et al. 2009). Conversely, this modification can be erased by deubiquitylating enzymes (DUBs). Numerous DUBs have been reported to deubiquitylate H2Bub1 (Fuchs and Oren 2014), most notably USP22 (Zhang, Pfeiffer et al. 2008), which is often upregulated in cancer and is part of a cancer gene signature associated with stem cell-like features and bad prognosis (Glinsky, Glinskii et al. 2006), and USP44 (Fuchs, Shema et al. 2012), reported to contribute to breast cancer aggressiveness (Liu, Sun et al. 2015). Modulation of H2Bub1 levels can upregulate the expression of specific gene sets in a cell type-specific manner, while repressing other gene sets (Shema, Tirosh et al. 2008; Fuchs, Shema et al. 2012). Regulation of transcription through H2Bub1 can occur through recruitment of SWI/SNF complexes, interactions with TFIIS, and, potentially, additional mechanisms (Shema, Kim et al. 2011).

Reduced global levels of H2Bub1 and of *RNF20* and *RNF40* mRNA, relative to corresponding non-cancerous tissue, were observed in various tumor types, including breast cancer (Chernikova, Razorenova et al. 2012; Hahn, Dickson et al. 2012; Urasaki, Heath et al. 2012; Thompson, Guppy et al. 2013; Wang, Yang et al. 2013; Bedi, Scheel et al. 2014; Tarcic, Pateras et al. 2016), suggesting that this reduction provides an advantage for tumor growth. Such a reduction could be driven by a variety of mechanisms,

including *RNF20* promoter hypermethylation (Shema, Tirosh et al. 2008) or increased expression of pertinent DUBs (Sussman, Stanek et al. 2013; Fuchs and Oren 2014; Liu, Sun et al. 2015; Wang, Zhu et al. 2015). We previously reported that RNF20 silencing promotes the migration of non-transformed human mammary epithelial MCF10A cells, and facilitates transformation of mouse cells, consistent with a tumor-suppressive role (Shema, Tirosh et al. 2008). Moreover, reduced levels of RNF20 and H2Bub1, observed in colitis and colorectal cancer in mice and humans, can increase the expression of cytokines via the NF-κB pathway, promoting chronic inflammation and colitis-associated colorectal cancer (Tarcic, Pateras et al. 2016). Together, these findings suggest a role for H2Bub1 in cancer prevention and for RNF20 as a tumor suppressor.

Other studies, however, suggested pro-tumorigenic activities of RNF20. Thus, it was shown that the SMURF2 E3 ubiquitin ligase, a direct negative regulator of RNF20 and thereby of H2Bub1, is downregulated in many breast tumors, while RNF20 protein is upregulated (Blank, Tang et al. 2012). Likewise, a positive contribution of H2Bub1 to cancer was demonstrated in leukemias involving *MLL* gene rearrangements (Wang, Kawaoka et al. 2013).

In the present study we sought to resolve the potential discrepancy between findings indicating a tumor-suppressive role of RNF20 and H2Bub1 in breast cancer (Prenzel, Begus-Nahrmann et al. 2011; Urasaki, Heath et al. 2012; Bedi, Scheel et al. 2014) and those suggesting a tumor-promoting role (Blank, Tang et al. 2012; Duan, Huo et al. 2016). We report that the effects of RNF20 on mammary tumorigenesis are subtype-dependent: whereas RNF20 exhibits tumor-suppressive features in basal-like breast cancer cells, it supports the tumorigenicity of luminal breast cancer cells. In basal-like cancers cells, RNF20 suppresses the NF-κB-dependent expression of cytokines, known to contribute to the growth of this tumor subtype. In contrast, RNF20 enhances the expression of targets of the estrogen receptor, the main pathway driving luminal breast cancer growth. In agreement, H2Bub1 levels tend to be higher in luminal tumors as compared to basal-like tumors. Hence, the opposing effects of RNF20, and most probably H2Bub1, on the different breast cancer subtypes are mediated, at least in part, by differential regulation of genes whose function is closely associated with subtype identity.

## Results

### Levels of H2Bub1 and its regulators differ between breast cancer subtypes

To obtain a detailed view of the expression of H2Bub1 regulators in breast cancer, we examined the mRNA levels of *RNF20*, *RNF40, SMURF2, USP22* and *USP44* in published gene expression profiles of large cohorts of breast cancers that include the major subtypes of the disease (TCGA 2012). We found that *RNF20* and *RNF40* mRNA levels were significantly lower in basal-like tumors relative to tumors of the luminal A and B subtypes (Figure 1A). In contrast, *SMURF2 and USP44*, predicted to downregulate H2Bub1, displayed the opposite pattern, showing increased levels in basal-like tumors relative to luminal tumors. *USP22* expression was comparable in the different subtypes (data not shown).

The combination of low RNF20/RNF40 and high SMURF2/USP44 expression in basal-like tumors suggested that H2Bub1 levels may be lower in these tumors than in luminal tumors. Indeed, analysis of a breast cancer tissue microarray stained with antibodies against H2Bub1 and total H2B, confirmed that triple negative (TN) cancers, which include basal-like tumors (Prat, Adamo et al. 2013), display, on average, weaker H2Bub1 staining than ER-positive tumors, representing the luminal subtypes (Figure 1B,C).

Interestingly, H2Bub1 levels showed an opposite association with clinical outcome in patients carrying tumors of the different subtypes. Among patients carrying ER+ tumors, those with high H2Bub1 staining had a shorter mean survival time than those with low H2Bub1, while an opposite trend was observed in patients with TN tumors (Figure 1D). To further support these distinct associations, we interrogated the TCGA breast cancer mRNA database, defining an H2Bub1 level-predictive score based on the relative levels of *RNF20*, *RNF40* and *SMURF2* mRNA in each tumor. This analysis, which allowed a better definition of the different subtypes, revealed a significant association of the H2Bub1-high scoring basal-like tumors with improved patient survival, whereas luminal A tumors displayed an opposite trend (Figure 1E and Supplemental Table 1). A similar trend, albeit weaker, was observed in the METABRIC dataset (Supplemental Figure S1 and Supplemental Table 1).

Together, these results indicate that basal-like tumors tend, on average, to have lower H2Bub1 levels than other subtypes, and that within these tumors low H2Bub1 is associated with poor prognosis. In contrast, luminal tumors tend to have more H2Bub1, and within these tumors it is high H2Bub1 that is associated with poorer outcome. These distinct associations might potentially account for the apparent discrepancies in previous reports, which did not segregate the analyzed tumors into molecular subtypes.

**H2Bub1 modulation exerts opposing effects on the proliferation and motility of basal-like versus luminal breast cancer cells**

To experimentally address the possibility that H2Bub1 exerts opposing effects in the different breast cancer subtypes, we investigated the effects of RNF20 and H2Bub1 manipulation on the properties of cells representing these subtypes. Consistent with our observations in human tumors, basal-like cell lines, including HCC1937 and MDA-MB-468, expressed lower levels of *RNF20* and *RNF40* mRNA (shCon, Figure 2A, and Figure S2A) and protein (Figure S2B) than luminal lines, including T47D and MCF7. The immortalized normal mammary epithelial cell lines hTERT-HME and 184A1 expressed intermediate levels of these transcripts (Figure 2A). Consistent with this, tumors that developed upon transplantation of basal-like cancer cell lines into mouse mammary glands showed weaker H2Bub1 staining than luminal cancer-derived tumors (Figure 2B).

Next, we silenced RNF20 in representative luminal and basal-like cancer lines by stable infection with a lentivirus expressing RNF20 shRNA (Figure 2A and S2B). RNF20 silencing reduced H2Bub1 levels in all tested lines (Figure S2B). Importantly, it increased the proliferation rates of the basal-like HCC1937 and MDA-MB-468 cells (Figure 2C). In contrast, the proliferation of MCF7 and T47D cells was decreased upon RNF20 silencing, indicating that it supports the proliferation of these luminal cancer cells. Furthermore, while the migration rate (assessed by a gap closure “scratch” assay) of RNF20-silenced HCC1937 cells was increased, MCF7 and T47D cells displayed reduced migration (Figure 2D,E).

In agreement, overexpression of RNF20 together with RNF40 had the opposite effect in both cell types: it suppressed the proliferation and migration of HCC1937 cells, yet enhanced MCF7 cell proliferation (Figure 2F,G). Furthermore, silencing of the H2Bub1

DUB USP44 (Fuchs, Shema et al. 2012), which increased H2Bub1 levels (Figure S2C), inhibited HCC1937 cell proliferation and migration, but promoted MCF7 cell proliferation (Figure 2F,G). This further implies that the effects of RNF20 manipulation are due to altered H2Bub1 levels. Thus, RNF20 and H2Bub1 support the proliferation and motility of cultured luminal breast cancer cells, but restrict these traits in basal-like cancer cells.

Interestingly, in normal mammary epithelium-derived hTERT-HME cells, silencing of RNF20 increased both the proliferation and gap closure rates (Figure 2C,D), in agreement with our previous findings in non-transformed MCF10A cells (Shema, Tirosh et al. 2008), and resembling its effects in the basal-like cancer lines. To further assess the impact of RNF20 depletion on the features of non-transformed human mammary epithelial cells, we transiently depleted *RNF20* mRNA from 184A1 cells and tested their ability to resume DNA synthesis after EGF starvation (Zwang, Sas-Chen et al. 2011). Strikingly, in this experimental setting, RNF20 silencing bypassed the requirement for EGF in order for the cells to enter S-phase effectively (Figure S2D), implying that RNF20 depletion reduces the dependence of those cells on growth factor signaling.

**RNF20 promotes the growth of luminal tumors but limits basal-like tumor growth**

To further explore the impact of RNF20 on basal-like versus luminal cancer, we compared the tumorigenic capacity of control and RNF20-silenced cells, injected into the mammary glands of immunocompromised NSG mice. Remarkably, relative to control cells, RNF20-silenced basal-like MDA-MB-468 cells gave rise to faster growing mammary tumors, which developed at an earlier onset (Figure 3A,B). As expected, the RNF20-silenced tumors displayed reduced H2Bub1 (Figure S3A,B).

Since MCF7 cells display very low orthotopic tumorigenicity, we infected them with a retrovirus expressing oncogenic mutant *H-RAS* immediately prior to injection. In contrast to MDA-MB-468 cells, RNF20-silenced MCF7-RAS cells yielded tumors that grew more slowly than controls, and in some cases gave rise to only residual growth (Figure 3C,D). Concordantly, the abundance of lung disseminating cells, indicative of metastatic potential, changed upon RNF20 silencing in a manner corresponding to primary tumor

size (Figure 3E,F). These findings further support the conjecture that RNF20 is tumor suppressive in basal-like cancer, but pro-tumorigenic in luminal cancer.

**RNF20 silencing augments the expression of cytokine genes in basal-like breast cancer cells but reduces the expression of ER-target genes in luminal cancer cells**

Pro-inflammatory cytokines can enhance the growth and migration of many types of cancer cells. Notably, the relative expression of many cytokine-encoding genes is higher in basal-like tumors than in luminal tumors (Hartman, Poage et al. 2013) (Figure S4A). Indeed, the basal-like cell lines expressed far more *IL8* and *IL6* mRNA than the luminal lines (Figure 4A). Notably, RNF20 silencing further increased *IL8* and *IL6* expression in the basal-like cells (Figure 4A). Moreover, in agreement with earlier observations (Tarcic, Pateras et al. 2016), RNF20 silencing decreased the repressive chromatin modification H3K9me3 on the promoters of these genes (Figure S4B). Consistent with this, we found that the levels of *IL6*, *IL8* and *CXLC1* mRNA in individual breast tumors are inversely correlated with *RNF20* and *RNF40* mRNA levels, but positively correlated with *SMURF2* mRNA (Figure 4B). RNF20 silencing augmented cytokine gene expression also in luminal cancer cells; however, the relative levels of these transcripts, even after RNF20 silencing, were much lower than in the basal-like cells (Figure 4A), probably rendering them inconsequential. Hence in basal-like breast cancer cells, RNF20, and presumably H2Bub1, repress cytokine-encoding genes.

Luminal tumors and cell lines typically express the estrogen receptor alpha (encoded by the *ESR1* gene) and its transcriptional targets, such as the progesterone receptor (*PGR*) gene (Figure S4A), and are dependent on estrogen receptor (ER) signaling for proliferation, survival and tumor progression. H2Bub1 was previously shown to promote ER target gene expression (Prenzel, Begus-Nahrmann et al. 2011; Bedi, Scheel et al. 2014), in conjunction with the histone chaperone SUPT6H (Bedi, Scheel et al. 2014). In agreement, RNF20 silencing in luminal cancer cells decreased the expression of the ER targets *PGR* and *CXCL12* (Figure 4C), which can drive breast cancer progression, proliferation and migration (Ramos, Grochoski et al. 2011; Yasmin, Siraj et al. 2015), as well as of *FOXA1*, encoding the pioneer transcription factor regulating ER function (Carroll, Liu et al. 2005). The same genes were downregulated also in RNF20-silenced

MCF7-RAS cells (Figure S5A). Furthermore, in human breast cancers, expression of ER target genes correlates positively with *RNF20* and *RNF40* mRNA, but negatively with *SMURF2* mRNA (Figure 4D). Thus, in luminal breast cancer cells, RNF20 and H2Bub1 promote ER target gene activation.

Overall, these results suggest that RNF20 and H2Bub1 affect the expression of genes that execute subtype-specific functions, possibly underpinning the opposite effects of this pathway on proliferation and tumorigenicity in the different subtypes.

We also observed differential effects of RNF20 silencing on several additional genes. Interestingly, RNF20 silencing upregulated the mRNA of the Polycomb factor and oncogene EZH2 (Kleer, Cao et al. 2003; Pietersen, Horlings et al. 2008) in HCC1937 cells, but decreased it in MCF7 cells (Figure S5B), consistent with their respective changes in in proliferation and motility. EZH2 may thus also contribute to the differential effects of shRNF20 on the proliferation and tumorigenicity of these cells. In addition, RNF20 knockdown elevated fibronectin (*FN1*) mRNA levels in HCC1937 but not MCF7 cells (Figure S5C), suggesting that some of the HCC1937 cell population might have transitioned towards a more mesenchymal-like differentiation state.

**RNF20 restricts NF-κB activity in basal-like breast cancer cells and promotes ER pathway activity in luminal cancer cells**

To further elucidate the opposite impact of RNF20 on basal-like versus luminal breast cancer features, we investigated the links between RNF20 and key signaling pathways in both tumor-derived cell types.

NF-κB is a major regulator of cytokine gene transcription. Recently, we showed that RNF20 and H2Bub1 decrease NF-κB activity, restricting inflammatory cytokine secretion (Tarcic, Pateras et al. 2016). We therefore asked whether changes in NF-κB activity might mediate the anti-proliferative effects of RNF20 in basal-like cells.

Silencing of the p65 subunit of NF-κB (Figure S6A) downregulated *IL6* mRNA in HCC1937 cells, and compromised the increase in *IL6* mRNA observed upon RNF20-silencing (Figure 5A). Importantly, p65 silencing markedly attenuated the proliferation of these cells, and combined silencing of *RNF20* and p65 failed to stimulate their proliferation beyond that observed in control cells (shCon+siLacZ) (Figure 5B). Hence,

the positive effects of *RNF20* depletion in these basal-like cancer cells rely, at least in part, on NF-κB.

As expected, transient silencing of *ESR1* (Figure S6B) strongly decreased *PGR* expression in MCF7 cells (Figure 5C). However, no additional decrease in *PGR* mRNA was seen when both *ESR1* and *RNF20* were silenced (Figure 5C), indicating that the transcriptional effect of RNF20 on *PGR* is dependent on ER activity. Concordantly, the inhibitory effect of *ESR1* silencing on MCF7 cell proliferation was not significantly augmented by simultaneous silencing of *RNF20* (Figure 5D). These findings support the hypothesis that RNF20 promotes luminal breast cancer cell proliferation, at least partly, by enhancing ER activity.

In sum, our findings suggest that H2Bub1 can possess tumor suppressive features in basal-like breast cancer, in part by restricting the activity of NF-κB and the expression of pro-inflammatory cytokines, but might exert tumor promoting effects in luminal cancers, presumably by augmenting estrogen receptor transcriptional signaling.

## Discussion

Earlier studies have provided apparently conflicting answers to the question whether H2Bub1 suppresses or promotes breast cancer (Prenzel, Begus-Nahrmann et al. 2011; Blank, Tang et al. 2012; Urasaki, Heath et al. 2012; Bedi, Scheel et al. 2014). We now report that both answers are correct, in a manner dictated by the particular subtype of breast cancer.

In triple-negative/basal-like tumors, RNF20, RNF40 and H2Bub1 levels tend to be low and USP44 expression is relatively high, similar to what has been observed in several other types of cancer and consistent with a tumor suppressive role. Concordantly, the prognosis of basal-like breast cancer patients with lower H2Bub1 tends to be worse than that of patients with high H2Bub1, indicating that in this type of disease H2Bub1 may indeed restrict tumor progression. In contrast, luminal tumors tend to display relatively higher levels of RNF20, RNF40 and H2Bub1, consistent with the tumor-promoting effects observed in this subtype. Indeed, luminal breast cancer patients with low H2Bub1 have a better prognosis than those with higher levels. Notably, a recent study, published while this paper was under revision, also supports a cancer-promoting role of RNF20 and RNF40 in luminal breast cancer (Duan, Huo et al. 2016).

Our findings imply that the differential effects of RNF20 on the growth and metastatic potential of luminal and basal-like tumors are dictated by the differences in the key molecular mechanisms driving each tumor subtype. While in luminal breast cancer estrogen receptor activity is a central driver, in basal-like/triple negative breast cancer NF-κB and cytokines are important contributors (Yamamoto, Taguchi et al. 2013). In agreement with previous work (Prenzel, Begus-Nahrmann et al. 2011), we found that RNF20 supports the efficient transcription of ER target genes. In contrast, H2Bub1 downregulation augments NF-κB activity and elevates the expression of pro-inflammatory cytokines, as recently observed also in an *in vivo* model of intestinal inflammation (Tarcic, Pateras et al. 2016). In basal-like cancers, cytokines such as IL6 and IL8 are abundantly expressed, and promote tumor aggressiveness in both humans and mice (Aceto, Duss et al. 2012; Prat, Adamo et al. 2013; Barbie, Alexe et al. 2014; Kim, Ouzounova et al. 2015). Moreover, in these tumors, NF-κB and inflammatory cytokines can enhance tumor growth by eliciting stem cell-like features and increasing the fraction

of tumor-initiating cells within the population (Kendellen, Bradford et al. 2014; Kim, Yang et al. 2015). Our observation that *EZH2* mRNA increases in RNF20-silenced basal-like cells is also consistent with their enhanced proliferation, and suggests that their differentiation state may be partly altered. However, the specific roles played by EZH2 in this context will require further analysis. Likewise, downregulation of *CXCL12* mRNA upon RNF20 silencing in luminal breast cancer cells, may contribute to their attenuated proliferation and migration.

One potential implication of these findings is that luminal tumors with low H2Bub1 might express lower levels of ER target genes, and instead might recruit alternative signaling pathways to drive tumor progression. If indeed so, anti-estrogenic therapy might be less effective in such cases than in patients with higher H2Bub1. On the other hand, basal-like and triple-negative breast cancers with low H2Bub1 could depend on the production of inflammatory cytokines more than those with high H2Bub1, and may thus respond better to anti-inflammatory treatments or inhibition of cytokine signaling (e.g. STAT3 inhibitors).

Altogether, our study demonstrates that the impact of H2Bub1 on cancer is context-dependent, and greatly depends on the particular molecular pathways that drive a given cancer. H2Bub1 is required for double strand DNA break repair (Moyal, Lerenthal et al. 2011; Nakamura, Kato et al. 2011), and its partial loss promotes genomic instability (Chernikova, Dorth et al. 2010; Chernikova, Razorenova et al. 2012); this might suggest that H2Bub1 should be a tumor suppressor by default. Likewise, the ability of H2Bub1/RNF20 to selectively support the expression of cancer-inhibitory genes such as *TP53*, while repressing proto-oncogenes such as *c-MYC* and *c-FOS* (Shema, Tirosh et al. 2008), is also consistent with a generic tumor suppressor function of this chromatin modification. Indeed, in many cancer types in which H2Bub1 has been analyzed by immunohistochemistry, advanced disease is associated with decreased H2Bub1. Our observation that RNF20 silencing promotes the proliferation and migration of non-transformed mammary cells is consistent with this. Nevertheless, in cancer types where H2Bub1 supports the efficient execution of transcriptional programs that underpin the cancerous behavior, such as luminal breast cancer and leukemias with *MLL* gene rearrangements (Blank, Tang et al. 2012; Wang, Kawaoka et al. 2013), the generic tumor

suppressor features of H2Bub1 may be overridden by the stronger tumor-specific selective advantages offered by H2Bub1, resulting in a drive to retain high H2Bub1 levels.

## Materials and Methods

### Breast cancer gene and protein expression and survival analyses

Breast cancer mRNA expression data was downloaded from the TCGA portal (http://cancergenome.nih.gov/). A list of included samples is presented in Supplemental Table 1. Expression of each gene was normalized to the mean across samples and $\log_2$ transformed. Breast cancer subtypes were predicted as previously described, using the PAM50 gene predictor (Parker, Mullins et al. 2009). Gene-expression correlations were calculated using Matlab employing the Pearson coefficient. Breast tumor microarrays were obtained from the Fox Chase Cancer Center (Philadelphia, USA). Consecutive sections were subjected to immunohistochemical staining for H2Bub1 and H2B and quantified as previously described (Tarcic, Pateras et al. 2016). To generate Kaplan-Meier survival plots, tumors were divided into H2Bub1 high and low groups based on the ratio between H2Bub1 and H2B across samples, samples with ratio higher then the median were considered 'H2Bub1 high', and those below 'H2Bub1 low'. *P* value was calculated using Gehan-Breslow-Wilcoxon test. To generate Kaplan-Meier plots from TCGA expression data, mRNA level of key H2Bub1 regulator has examined in individual samples. Those that displayed low expression (Z-score < –1) of RNF20 or RNF40 and highly expressed (Z-score > 1) SMURF2 were considered to low H2Bub1 pathway; samples that did not meet this criteria were considered H2Bub1 high.

### Cell culture and gene perturbation

MCF7 cells were cultured at $37^0$C in complete DMEM medium (Biological Industries, Israel) supplemented with penicillin, streptomycin and 5% heat-inactivated fetal calf serum. HCC1937 cells were cultured at $37^0$C in complete RPMI medium (Biological Industries, Israel) supplemented with penicillin and streptomycin and 5% heat-inactivated fetal calf serum. T47D cells were cultured at $37^0$C in complete RPMI medium (Biological Industries, Israel) supplemented with penicillin and streptomycin, as well as 5 μg/mL insulin and 5% heat-inactivated fetal calf serum. MDA-MB-468 cells were cultured at $37^0$C in Leibovitz L15 (Biological Industries, Israel) containing 10% FBS and supplemented with penicillin and streptomycin. Cell lines were obtained from ATCC, and authenticated by STR profiling. HME primary mammary epithelial cells expressing

the hTERT gene (Elenbaas, Spirio et al. 2001), were obtained from Robert Weinberg. Normal 184A1 mammary epithelial cells (Zwang, Sas-Chen et al. 2011) were obtained from Yosef Yarden. Both cell lines were grown in MEGM media (Lonza) mixed 1:1 with DMEM:F12, supplemented with insulin, EGF and hydrocortisone. siRNA transfection, RNA extraction, qRT-PCR Western blot analysis and ChIP assay were performed as previously described (Tarcic, Pateras et al. 2016). For stable RNA silencing, the pLKO.1-puro lentiviral vector was used. We employed standard virus generation and infection procedures, packaging with the pHRΔ8.2 and pCMV-VSV-G vectors. BrdU incorporation analysis in cells exposed to EGF pulses was as previously described (Zwang, Sas-Chen et al. 2011).

**Proliferation and "scratch" assays**

For proliferation assays, cells were seeded at $1x10^6$ cells/10 cm plate. 4 days later, cells were harvested and counted. For scratch assays, cells were seeded in a culture insert (Ibidi). The insert was removed 24h following seeding, and migration of cells was monitored at the indicated times.

**Tumor xenografts**

For xenograft implantation, $1x10^6$ viable GFP-labeled cells were injected in 20µl culture medium containing 25% Matrigel (BD-biosciences) into both #4 mammary glands of 6-week-old female NSG mice. Tumor growth during the experiment was measured by palpation, and at the end of the experiment excised tumors were weighed. Tumors were formalin-fixed and paraffin-embedded for immunohistochemistry. To assess metastatic burden, lungs were pictured and the GFP-positive area was quantified and normalized to the total lung area. All experiments involving animals were performed under approval of the Hebrew University Animal Use and Care Committee.

**Immunohistochemistry**

For IHC analysis, the following antibodies were used: anti-histone H2B (ab52484, Abcam, and 07-371, Millipore) and anti-H2Bub1 (56, 05-1312, Millipore). IHC was performed on paraffin-embedded tissues, employing heat-mediated antigen retrieval in

10mM citric acid (pH6.0). The UltraVision LP Detection System was used (#TL-060-HD, Thermo Scientific, Bioanalytica, Greece) according to the manufacturer's instructions. Evaluation was performed by measuring the H2Bub1/H2B ratio as previously described (Tarcic, Pateras et al. 2016).

**Antibodies**

The following primary antibodies were used for Western blot analysis and ChIP: anti-H2B (07-371, Millipore), anti-RNF20 (ab32639, Abcam), anti-GAPDH (MAB374, Millipore), anti-H3K9me3 (ab8898, Abcam), anti H3 (ab1791, Abcam). The anti-H2Bub1 antibody was previously described (Minsky, Shema et al. 2008).

**Primers**

For mRNA amplification the following primers were used:

IL-8: Fw 5'GGCAGCCTTCCTGATTTCTG

Rev 5'CTTGGCAAAACTGCACCTTCA

RNF20: Fw 5'GAACAGCGACTCAACCGACA

Rev 5'GGAATTCACCCGTTCTAGGACTT

GAPDH: Fw 5'AGCCTCAAGATCATCAGCAATG

Rev 5'CACGATACCAAAGTTGTCATGGAT

IL-6: Fw 5'AGCCCTGAGAAAGGAGACATGTA

Rev 5'TCTGCCAGTGCCTCTTTGCT

p65: Fw 5'CTCCGCGGGCAGCAT

Rev 5'TCCTGTGTAGCCATTGATCTTGAT

PGR: Fw 5' GTCCTTACCTGTGGGAGCTG

Rev 5' CGATGCAGTCATTTCTTCCA

ESR1: Fw 5' AGCTACTGTTTGCTCCTAACTTGCT

Rev 5' CCACCATGCCCTCTACACATT

EZH2: Fw 5' GCGCGGGACGAAGAATAATCAT

Rev 5' TACACGCTTCCGCCAACAAACT

Cxcl12: Fw 5' CGTCAAGCATCTCAAAATTCTCA

Rev 5' CAGCCGGGCTACAATCTGA

FN1: Fw 5’ GACTGGGACGTTTTATCAAATTGG

Rev 5’ TTGAGCTTGGATAGGTCTGTAAAGG

For ChIP analysis:

IL8 κB site: Fw5’AAACTTTCGTCATACTCCGTATTTGA

Rev 5’ TCTCTTGGCAGCCTTCCTGA

IL6 kB site: Fw5’ AGCCTCAATGACGACCTAAGCT

Rev 5’ CGTCCTTTAGCATGGCAAGAC

## Acknowledgements

This study was supported in part by grant 293438 (RUBICAN) from the European Research Council (MO), the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (MO), the Israel Cancer Research Fund (I.B.) and a Jacob and Lena Joels Memorial Foundation Senior Lectureship for Excellence in the Life and Medical Sciences (I.B.). M.O. is incumbent of the Andre Lwoff chair in molecular biology.

## Conflict of interests

The authors declare no conflict of interest.

## Author contributions

O.T. and R.G. designed and performed the experiments, and wrote the manuscript; I.P. and V.G. performed and analyzed histological studies; H.M. aided with the cell culture experiments; E.P. and B.M. conducted pathological analysis and provided suggestions; Y.Z. and Y.Y. performed the analysis in Fig. S2D. M.O. and I.B. supervised the work and wrote the manuscript.

## References


Aceto, N., S. Duss, et al. (2012). "Co-expression of HER2 and HER3 receptor tyrosine kinases enhances invasion of breast cells via stimulation of interleukin-8 autocrine secretion." Breast Cancer Res **14**(5): R131.

Barbie, T. U., G. Alexe, et al. (2014). "Targeting an IKBKE cytokine network impairs triple-negative breast cancer growth." J Clin Invest **124**(12): 5411-5423.

Bedi, U., A. H. Scheel, et al. (2014). "SUPT6H controls estrogen receptor activity and cellular differentiation by multiple epigenomic mechanisms." Oncogene.

Blank, M., Y. Tang, et al. (2012). "A tumor suppressor function of Smurf2 associated with controlling chromatin landscape and genome stability through RNF20." Nat Med **18**(2): 227-234.

Campos, E. I. and D. Reinberg (2009). "Histones: annotating chromatin." Annu Rev Genet **43**: 559-599.

Carroll, J. S., X. S. Liu, et al. (2005). "Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1." Cell **122**(1): 33-43.

Chernikova, S. B., J. A. Dorth, et al. (2010). "Deficiency in Bre1 impairs homologous recombination repair and cell cycle checkpoint response to radiation damage in mammalian cells." Radiat Res **174**(5): 558-565.

Chernikova, S. B., O. V. Razorenova, et al. (2012). "Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability." Cancer Res.

Duan, Y., D. Huo, et al. (2016). "Ubiquitin ligase RNF20/40 facilitates spindle assembly and promotes breast carcinogenesis through stabilizing motor protein Eg5." Nat Commun **7**: 12648.

Duan, Y., D. Huo, et al. (2016). "Ubiquitin ligase RNF20/40 facilitates spindle assembly and promotes breast carcinogenesis through stabilizing motor protein Eg5." Nat Commun **7**: 12648.

Elenbaas, B., L. Spirio, et al. (2001). "Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells." Genes Dev **15**(1): 50-65.

Ellis, M. J. and C. M. Perou (2013). "The genomic landscape of breast cancer as a therapeutic roadmap." Cancer Discov **3**(1): 27-34.

Fuchs, G. and M. Oren (2014). "Writing and reading H2B monoubiquitylation." Biochim Biophys Acta.

Fuchs, G., E. Shema, et al. (2012). "RNF20 and USP44 Regulate Stem Cell Differentiation by Modulating H2B Monoubiquitylation." Mol Cell **46**(5): 662-673.

Glinsky, G. V., A. B. Glinskii, et al. (2006). "Dual-color-coded imaging of viable circulating prostate carcinoma cells reveals genetic exchange between tumor cells in vivo, contributing to highly metastatic phenotypes." Cell Cycle **5**(2): 191-197.

Hahn, M. A., K. A. Dickson, et al. (2012). "The tumor suppressor CDC73 interacts with the ring finger proteins RNF20 and RNF40 and is required for the maintenance of histone 2B monoubiquitination." Hum Mol Genet **21**(3): 559-568.

Hartman, Z. C., G. M. Poage, et al. (2013). "Growth of triple-negative breast cancer cells relies upon coordinate autocrine expression of the proinflammatory cytokines IL-6 and IL-8." Cancer Res **73**(11): 3470-3480.

Kendellen, M. F., J. W. Bradford, et al. (2014). "Canonical and non-canonical NF-kappaB signaling promotes breast cancer tumor-initiating cells." Oncogene **33**(10): 1297-1305.

Kim, G., M. Ouzounova, et al. (2015). "SOCS3-mediated regulation of inflammatory cytokines in PTEN and p53 inactivated triple negative breast cancer model." Oncogene **34**(6): 671-680.

Kim, J., M. Guermah, et al. (2009). "RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells." Cell **137**(3): 459-471.

Kim, T., S. J. Yang, et al. (2015). "A basal-like breast cancer-specific role for SRF-IL6 in YAP-induced cancer stemness." Nat Commun **6**: 10186.

Kleer, C. G., Q. Cao, et al. (2003). "EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells." Proc Natl Acad Sci U S A **100**(20): 11606-11611.

Liu, T., B. Sun, et al. (2015). "USP44+ Cancer Stem Cell Subclones Contribute to Breast Cancer Aggressiveness by Promoting Vasculogenic Mimicry." Mol Cancer Ther **14**(9): 2121-2131.

Minsky, N., E. Shema, et al. (2008). "Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells." Nat Cell Biol **10**(4): 483-488.

Moyal, L., Y. Lerenthal, et al. (2011). "Requirement of ATM-Dependent Monoubiquitylation of Histone H2B for Timely Repair of DNA Double-Strand Breaks." Molecular Cell **41**(5): 529-542.

Nakamura, K., A. Kato, et al. (2011). "Regulation of Homologous Recombination by RNF20-Dependent H2B Ubiquitination." Molecular Cell **41**(5): 515-528.

Osley, M. A. (2006). "Regulation of histone H2A and H2B ubiquitylation." Brief Funct Genomic Proteomic **5**(3): 179-189.

Parker, J. S., M. Mullins, et al. (2009). "Supervised risk predictor of breast cancer based on intrinsic subtypes." J Clin Oncol **27**(8): 1160-1167.

Pietersen, A. M., H. M. Horlings, et al. (2008). "EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer." Breast Cancer Res **10**(6): R109.

Prat, A., B. Adamo, et al. (2013). "Molecular characterization of basal-like and non-basal-like triple-negative breast cancer." Oncologist **18**(2): 123-133.

Prat, A. and C. M. Perou (2011). "Deconstructing the molecular portraits of breast cancer." Mol Oncol **5**(1): 5-23.

Prenzel, T., Y. Begus-Nahrmann, et al. (2011). "Estrogen-Dependent Gene Transcription in Human Breast Cancer Cells Relies upon Proteasome-Dependent Monoubiquitination of Histone H2B." Cancer Res **71**(17): 5739-5753.

Rakha, E. A., M. E. El-Sayed, et al. (2009). "Patho-biological aspects of basal-like breast cancer." Breast Cancer Res Treat **113**(3): 411-422.

Ramos, E. A., M. Grochoski, et al. (2011). "Epigenetic changes of CXCR4 and its ligand CXCL12 as prognostic factors for sporadic breast cancer." PLoS One **6**(12): e29461.

Shema, E., J. Kim, et al. (2011). "RNF20 Inhibits TFIIS-Facilitated Transcriptional Elongation to Suppress Pro-oncogenic Gene Expression." Molecular Cell **42**(4): 477-488.

Shema, E., I. Tirosh, et al. (2008). "The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression." Genes & Development **22**(19): 2664-2676.

Sussman, R. T., T. J. Stanek, et al. (2013). "The epigenetic modifier ubiquitin-specific protease 22 (USP22) regulates embryonic stem cell differentiation via transcriptional repression of sex-determining region Y-box 2 (SOX2)." J Biol Chem **288**(33): 24234-24246.

Tarcic, O., I. S. Pateras, et al. (2016). "RNF20 Links Histone H2B Ubiquitylation with Inflammation and Inflammation-Associated Cancer." Cell Rep **14**(6): 1462-1476.

TCGA (2012). "Comprehensive molecular portraits of human breast tumours." Nature **490**(7418): 61-70.

Thompson, L. L., B. J. Guppy, et al. (2013). "Regulation of chromatin structure via histone post-translational modification and the link to carcinogenesis." Cancer Metastasis Rev.

Urasaki, Y., L. Heath, et al. (2012). "Coupling of glucose deprivation with impaired histone H2B monoubiquitination in tumors." PLoS One **7**(5): e36775.

Wang, E., S. Kawaoka, et al. (2013). "Histone H2B ubiquitin ligase RNF20 is required for MLL-rearranged leukemia." Proc Natl Acad Sci U S A.

Wang, Z., L. Zhu, et al. (2015). "Decreased H2B monoubiquitination and overexpression of ubiquitin-specific protease enzyme 22 in malignant colon carcinoma." Hum Pathol **46**(7): 1006-1014.

Wang, Z. J., J. L. Yang, et al. (2013). "Decreased histone H2B monoubiquitination in malignant gastric carcinoma." World J Gastroenterol **19**(44): 8099-8107.

Yamamoto, M., Y. Taguchi, et al. (2013). "NF-kappaB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype." Nat Commun **4**: 2299.

Yasmin, R., S. Siraj, et al. (2015). "Epigenetic regulation of inflammatory cytokines and associated genes in human malignancies." Mediators Inflamm **2015**: 201703.

Zhang, X. Y., H. K. Pfeiffer, et al. (2008). "USP22, an hSAGA subunit and potential cancer stem cell marker, reverses the polycomb-catalyzed ubiquitylation of histone H2A." Cell Cycle **7**(11): 1522-1524.

Zwang, Y., A. Sas-Chen, et al. (2011). "Two phases of mitogenic signaling unveil roles for p53 and EGR1 in elimination of inconsistent growth signals." Mol Cell **42**(4): 524-535.

## Figure legends

**Figure 1. Levels of H2Bub1 and its regulators differ between breast cancer subtypes**

A. mRNA expression levels of the H2Bub1 pathway genes *RNF20*, *RNF40* (positive regulators) and *SMURF2, USP44* (negative regulators) in 837 breast tumors included in the TCGA database and stratified according to tumor subtype. Values are shown relative to mean across all samples (0.0). Error bars=S.E. ***, $P < 0.001$

B. Mean ratios of intensity of H2Bub1 staining relative to total H2B staining in a breast cancer tissue microarray (TMA), segregated into triple negative (TN) and estrogen receptor-positive (ER+) tumors. Error bars=S.E. *, $P < 0.05$.

C. Examples of tumor specimens from the TMA in (B), stained for total H2B or H2Bub1. Scale bar= 500μm.

D. Cancer-related survival rates of the patients carrying the tumors analyzed in (B), comparing patients with a high H2Bub1/H2B ratio (red) to those with a low ratio (green), across TN or ER+ tumors. *P* values were calculated using Gehan-Breslow-Wilcoxon test.

E. Survival of patients with luminal A (left) or basal-like (right) tumors in the TCGA dataset, predicted to have either high (blue) or low (red) H2Bub1 scores. Low H2Bub1 score was considered when either RNF20 or RNF40 display low expression (Z-score < –1) and when SMURF2 was highly expressed (Z-score > 1) within a given sample; high H2Bub1 pathway was considered when the opposite criteria were met.

**Figure 2. RNF20 silencing augments proliferation and migration of basal-like breast cancer cells but decreases these abilities in luminal breast cancer cells, while USP44 silencing causes opposite effects**

A. qRT-PCR analysis of *RNF20* mRNA in HCC1937, MDA-MB-468, T47D, MCF7, hTERT-HME and 184A1 cells, stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon). Values were normalized to *GAPDH* mRNA. Error bars= S.E.

B. Sections of tumor xenografts, formed by orthotopic injection of MCF7-RAS, MDA-MB-468 or HCC1937 cells, stained for H2Bub1 and H2B. Scale bar = 100μm. Lower panel= quantification of staining. *, $P <0.05$. Error bars= S.E.

C. Relative numbers of HCC1937, MDA-MB-468, MCF7, T47D and hTERT-HME cells stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon), 4 days after plating equal numbers of cells. Values were calculated by comparing numbers of shRNF20 cells to those of the corresponding shCon cells. *, $P$ <0.05; **, $P$ <0.01, ***,$P$ <0.001. Error bars= S.E.

D. Percentage of gap closure in cultures of MCF7, T47D, HCC1937 and hTERT-HME cells stably expressing RNF20 shRNA (shRNF20) or a control shRNA (shCon) in a "scratch" experiment endpoint, calculated by analysis of images exemplified in E. Values indicate means of closure ± S.E. in 3 assay replicates. *, $P < 0.05$; ***, $P < 0.001$.

E. Representative images of MCF7, T47D and HCC1937 cells stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon) at the start and endpoints of a scratch assay; endpoints varied according to the relative migratory capacity of each cell line.

F. Relative numbers of HCC1937 and MCF7 cells transfected with USP44 siRNA or siLacZ as control, or with RNF20 and RNF40 expression plasmids, measured as described in C. *, $P$ <0.05; **, $P$ <0.01. Error bars= S.E.

G. Percentage gap closure of HCC1937 cells transfected with either USP44 siRNA or siLacZ as control, or with a combination of RNF20 and RNF40 expression plasmids. Values indicate means of closure ± S.E. in 3 assay replicates. *, $P < 0.05$.

**Figure 3. RNF20 silencing augments the tumorigenicity and metastatic seeding of basal-like breast cancer cells but decreases them in luminal breast cancer cells**

A. Growth curves of tumors formed by MDA-MB-468 cells stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon), injected into the mammary glands of NSG mice. Values indicate average tumor volume ±S.E. n = number of tumors included in each group.

B. Final weights of tumors shown in (A) after excision. Dots indicate individual tumors. ***, $P$ <0.001. Error bars= S.E.

C. Growth curves of tumors formed by MCF7-RAS cells stably expressing RNF20 shRNA or control shRNA, measured as in (A).

D. Final weights of excised tumors in (C). *, $P$ <0.05. Error bars= S.E.

E. Lung disseminated cells in mice injected with either MDA-MB-468 or MCF7-RAS cells, quantified by image analysis of GFP fluorescence in lungs (A.U.= arbitrary units). *, $P$ <0.05. Error bars= S.E.

F. Representative images of metastatic cells (green) in lungs of tumor bearing mice, analyzed in (E).

**Figure 4. Regulation of cytokine genes and estrogen receptor target genes by RNF20 in basal-like and luminal breast cancer cells**

A. *IL8* (top) and *IL6* (bottom) mRNA levels in HCC1937, MDA-MB-468, MCF7 and T47D cells, stably expressing RNF20 shRNA or control shRNA, measured by qRT-PCR. Values were normalized to *GAPDH* mRNA. *, $P$ <0.05. Error bars= SE.

B. Pearson correlation between expression levels of H2Bub1-related genes (*RNF20*, *RNF40* and *SMURF2*) and pro-inflammatory cytokine genes across 837 tumor samples (TCGA database). Numbers indicate correlation R-values, *P<0.05* in all panels.

C. Relative *PGR* (top), *FOXA1* (bottom) and *CXCL12* (right) mRNA levels in the indicated cell lines, determined as in (A). *, $P$ <0.05; **, $P$ <0.01. Error bars= SE.

D. Correlation between expression levels of indicated H2Bub1-regulating genes and ER pathway genes in the TCGA samples described in (B).

**Figure 5. RNF20 promotes ER pathway activity in luminal breast cancer cells and restricts NF-kB activity in basal-like cells**

A. Relative expression of *IL6* mRNA assessed by qRT-PCR in HCC1937 cells stably expressing *RNF20* shRNA (shRNF20) or control shRNA (shCon) and transiently transfected with siRNA oligos targeting either *p65*, *ESR1* or LacZ as control. Expression was normalized to *GAPDH* mRNA. Error bars= S.E.

B. Relative numbers of HCC1937 cells transfected as in (A), counted 4 days post seeding and normalized to the shCon + siLacZ culture. **, $P$ <0.01, ***; $P$ <0.001. Error bars= S.E.

C. Relative expression of *PGR* mRNA, assessed by qRT-PCR in MCF7 cells stably expressing *RNF20* shRNA (shRNF20) or control shRNA (shCon), transfected as in (A). Expression was normalized to *GAPDH* mRNA. Error bars= S.E.

D. Relative numbers of MCF7 cells transfected as in (A), counted 4 days post seeding and normalized to the shCon + siLacZ culture. *, $P$ <0.05; **, $P$ <0.01. N.S. = non-significant. Error bars= S.E.

**Supplemental Figure Legends:**

**Figure S1. Correlation of H2Bub1 score with survival in basal-like versus luminal breast tumors in the METABRIC dataset**

Survival of patients with luminal A or basal-like tumors in the METABRIC dataset, displaying either a high (blue) or low (red) H2Bub1 score. Low H2Bub1 score was considered when either *RNF20* or *RNF40* mRNA display low expression (Z-score < –1) and when *SMRUF2* mRNA is highly expressed (Z-score > 1) within a given sample (suggesting that the H2B ubiquitylation machinery is downregulated); high H2Bub1 score was considered when the opposite criteria were met.

**Figure S2. RNF20 and H2Bub1 levels are lower in basal-like breast cancer cell lines than in luminal lines**

A. qRT-PCR analysis of *RNF20* and *RNF40* mRNA in a panel of breast cancer cell lines. Values were normalized to *GAPDH* mRNA. Error bars= S.E.

B. Western blot analysis of RNF20 and H2Bub1 in basal-like (MDA-MB-468, HCC1937) and luminal (T47D, MCF7) cell lines stably expressing *RNF20* shRNA (shRNF20) or control shRNA (shCon). GAPDH served as loading control.

C. Western blot analysis of H2Bub1 in HCC1937 and MCF7 cells following USP44 KD. H2B served as loading control.

D. BrdU incorporation after exit from quiescence in 184A1 human mammary epithelial cells. Cells were growth factor starved for 16 hours, and then either treated with EGF ("1E", red) for 1 hour, or mock treated ("1S", green). Thereafter, cells were washed and incubated in starvation medium for 7 hours, as indicated, either followed by a second one hour pulse of EGF, or not. Cells were then washed and incubated for 3 hours with BrdU in starvation medium. Thereafter, the cells were fixed, stained with BrdU antibody, and counted under a fluorescent microscope. BrdU incorporation into DNA was measured by determining the ratio of BrdU to DAPI-stained nuclei. Bars represent standard errors calculated from at least 15 non-overlapping photomicrograph fields (>500

nuclei). Significant *P*-values of two-tailed student's *t*-test are indicated. The experiment was repeated twice, with similar results.

**Figure S3. Quantification of H2Bub1 in xenografts of MDA-MB-468 cells**

A. Mean ratios of intensity of H2Bub1 staining relative to total H2B staining in tumors formed by MDA-MB-468 cells stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon), injected into the mammary glands of NSG mice. Error bars=S.E. *, $P < 0.05$.

B. Examples of specimens from the tumors in (A), stained for total H2B or for H2Bub1. Scale bar= 100μm.

**Figure S4. Basal-like tumors display increased expression of cytokine genes, while luminal tumors have increased expression of ER target genes**

A. Boxplot representation of relative expression levels of the indicated cytokine mRNAs across 837 breast tumors in the TCGA database, stratified by tumor subtype. Each box represents the second and third quartiles within each group; center line marks the median and circles represent outliers.

B. HCC1937 cells stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon) were harvested and subjected to chromatin immunoprecipitation (ChIP) with H3 and H3K9me3 antibodies. Precipitated DNA was subjected to qRT-PCR analysis with primers spanning the NF-κB sites (κB) of the *IL8* and *IL6* genes. H3K9me3 values were normalized to H3 values. Error bars = S.D.

**Figure S5. Changes in gene expression upon RNF20 knockdown in luminal and basal-like cell lines**

A. MCF7 cells stably expressing RNF20 shRNA (shRNF20) or control shRNA (shCon) were infected with a lentivirus expressing mutant H-RAS. 48 hours later,

expression of the indicated genes was quantified by qRT-PCR. Values were normalized to *GAPDH* mRNA in the same sample. Error bars= S.E.

B. *EZH2* mRNA levels in HCC1937 and MCF7 cells stably expressing RNF20 shRNA or control shRNA, measured by qRT-PCR. Values were normalized to *GAPDH* mRNA. *, P <0.05. Error bars= S.E.

C. *FN1* mRNA levels in HCC1937 and MCF7 cells stably expressing RNF20 shRNA or control shRNA, measured by qRT-PCR. Values were normalized to *GAPDH* mRNA. *, P <0.05. Error bars= SE.

**Figure S6. p65 and ESR1 knockdown efficiency**

A. Relative *p65* mRNA levels, determined by qRT-PCR, in HCC1937 or MCF7 cells stably expressing RNF20 shRNA (RNF) or control shRNA (Con) and transiently transfected with siRNA oligos against either p65, ESR1 or LacZ as control (as indicated below each bar). Values were normalized to *GAPDH* mRNA in the same sample. Error bars= S.E.

B. Relative *ESR1* mRNA levels in the same cells as in A. Levels were normalized to *GAPDH* mRNA in the same sample. Error bars= S.E.

**Supplemental Table 1**

List of samples included in the TCGA and METABRIC analysis along with their molecular subtype.

Figure 1

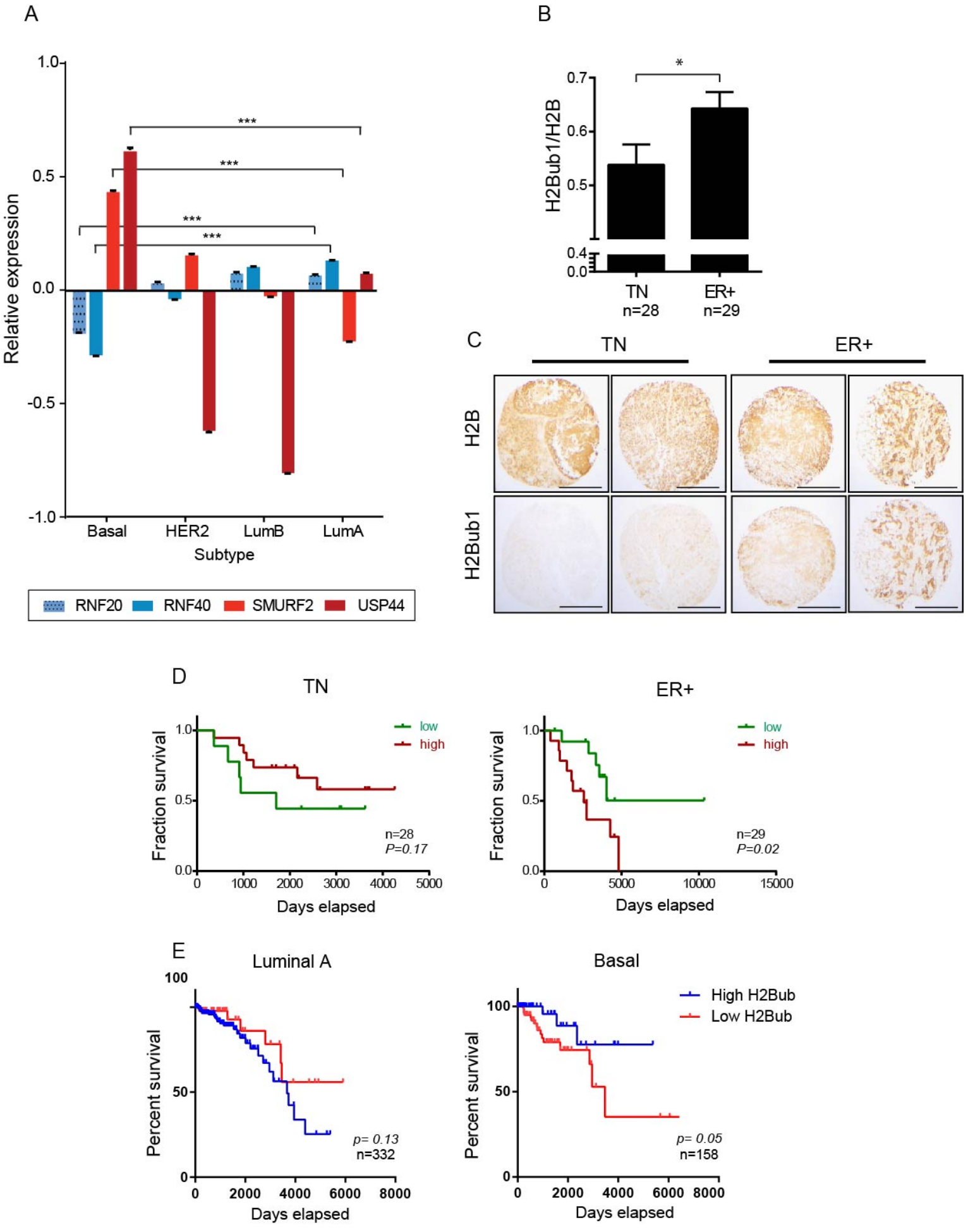

A
Relative expression
1.0
0.5
0.0
-0.5
-1.0
***
***
***
***
Basal
HER2
LumB
LumA
Subtype
RNF20
RNF40
SMURF2
USP44
B
H2Bub1/H2B
0.7
0.6
0.5
0.4
0.0
*
TN
n=28
ER+
n=29
C
TN
ER+
H2B
H2Bub1
D
TN
low
high
Fraction survival
1.0
0.5
0.0
n=28
P=0.17
0
1000
2000
3000
4000
5000
Days elapsed
ER+
low
high
Fraction survival
1.0
0.5
0.0
n=29
P=0.02
0
5000
10000
15000
Days elapsed
E
Luminal A
Percent survival
100
50
0
p= 0.13
n=332
0
2000
4000
6000
8000
Days elapsed
Basal
High H2Bub
Low H2Bub
Percent survival
100
50
0
p= 0.05
n=158
0
2000
4000
6000
8000
Days elapsed

Figure 2

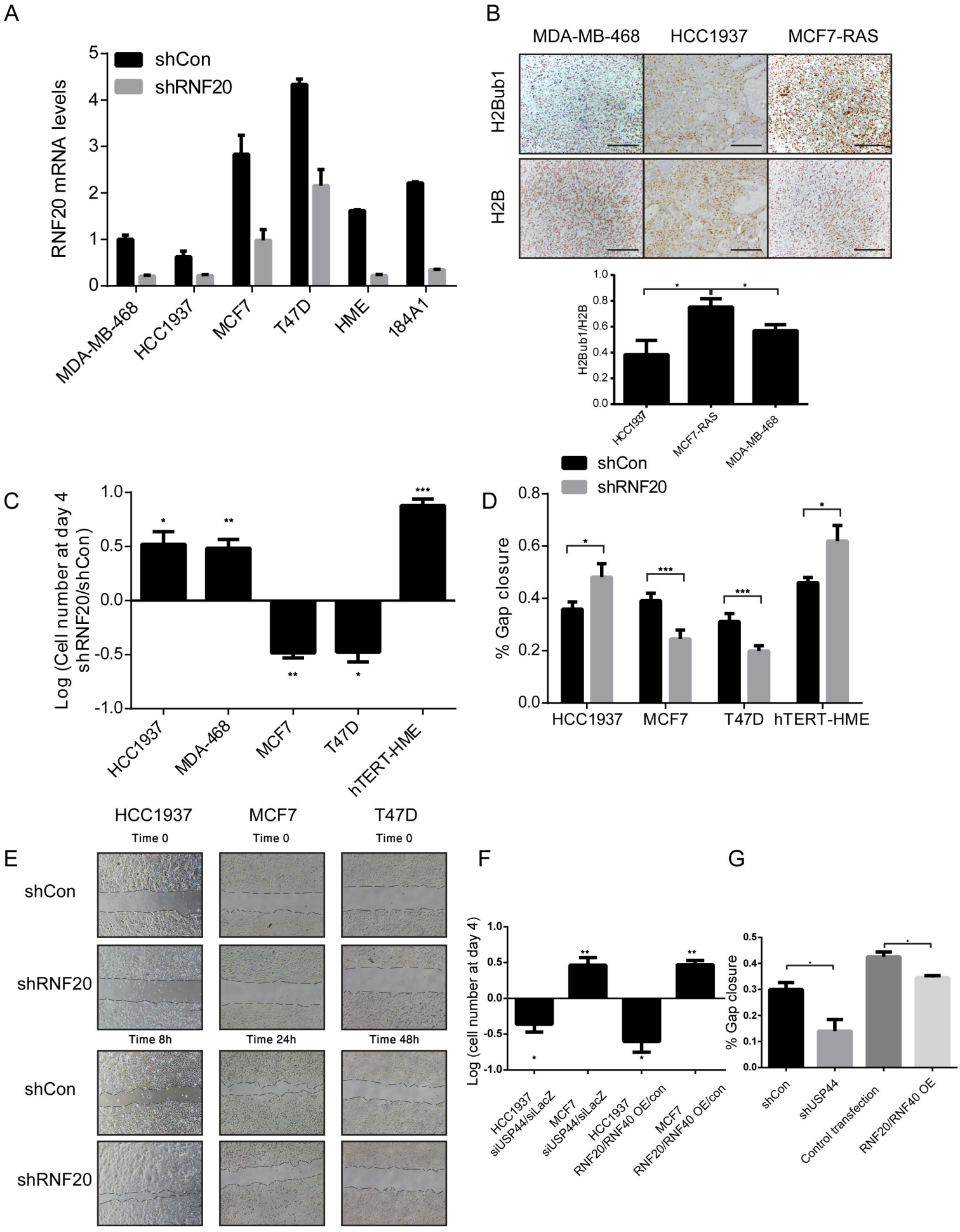

A
RNF20 mRNA levels
shCon
shRNF20
MDA-MB-468
HCC1937
MCF7
T47D
HME
184A1
B
MDA-MB-468
HCC1937
MCF7-RAS
H2Bub1
H2B
H2Bub1/H2B
HCC1937
MCF7-RAS
MDA-MB-468
C
Log (Cell number at day 4 shRNF20/shCon)
HCC1937
MDA-468
MCF7
T47D
hTERT-HME
D
shCon
shRNF20
% Gap closure
HCC1937
MCF7
T47D
hTERT-HME
E
HCC1937
MCF7
T47D
Time 0
shCon
shRNF20
Time 8h
Time 24h
Time 48h
shCon
shRNF20
F
Log (cell number at day 4)
HCC1937 siUSP44/siLacZ
MCF7 siUSP44/siLacZ
HCC1937 RNF20/RNF40 OE/con
MCF7 RNF20/RNF40 OE/con
G
% Gap closure
shCon
shUSP44
Control transfection
RNF20/RNF40 OE

Figure 3

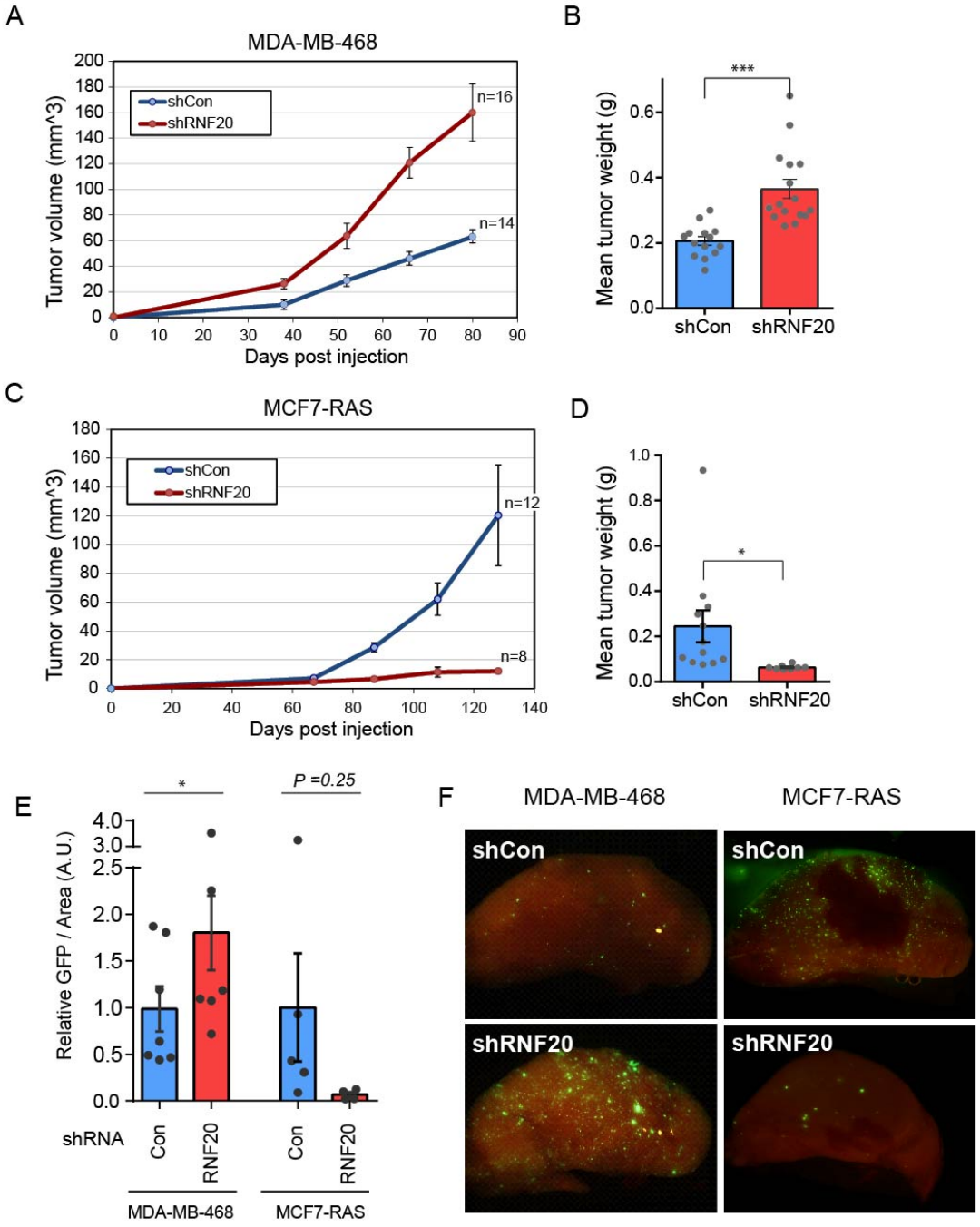

A
MDA-MB-468
Tumor volume (mm^3)
Days post injection
shCon
shRNF20
n=16
n=14
B
Mean tumor weight (g)
***
shCon
shRNF20
C
MCF7-RAS
Tumor volume (mm^3)
Days post injection
shCon
shRNF20
n=12
n=8
D
Mean tumor weight (g)
*
shCon
shRNF20
E
Relative GFP / Area (A.U.)
*
P =0.25
shRNA
Con
RNF20
Con
RNF20
MDA-MB-468
MCF7-RAS
F
MDA-MB-468
MCF7-RAS
shCon
shCon
shRNF20
shRNF20

Figure 4

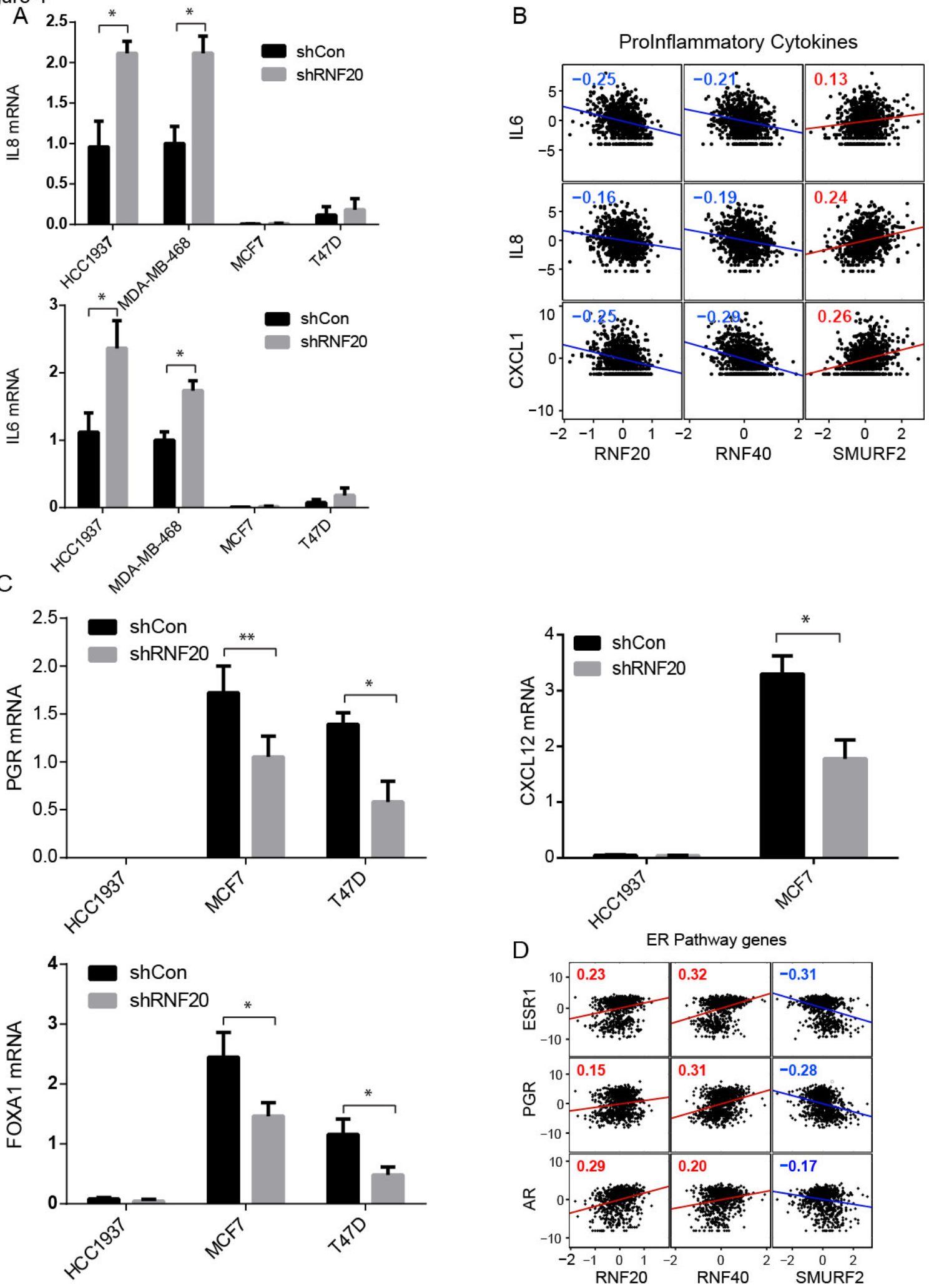

A
IL8 mRNA
shCon
shRNF20
HCC1937
MDA-MB-468
MCF7
T47D
IL6 mRNA
B
ProInflammatory Cytokines
IL6
IL8
CXCL1
−0.25
−0.21
0.13
−0.16
−0.19
0.24
−0.25
−0.29
0.26
RNF20
RNF40
SMURF2
C
PGR mRNA
CXCL12 mRNA
FOXA1 mRNA
D
ER Pathway genes
ESR1
PGR
AR
0.23
0.32
−0.31
0.15
0.31
−0.28
0.29
0.20
−0.17

Figure 5

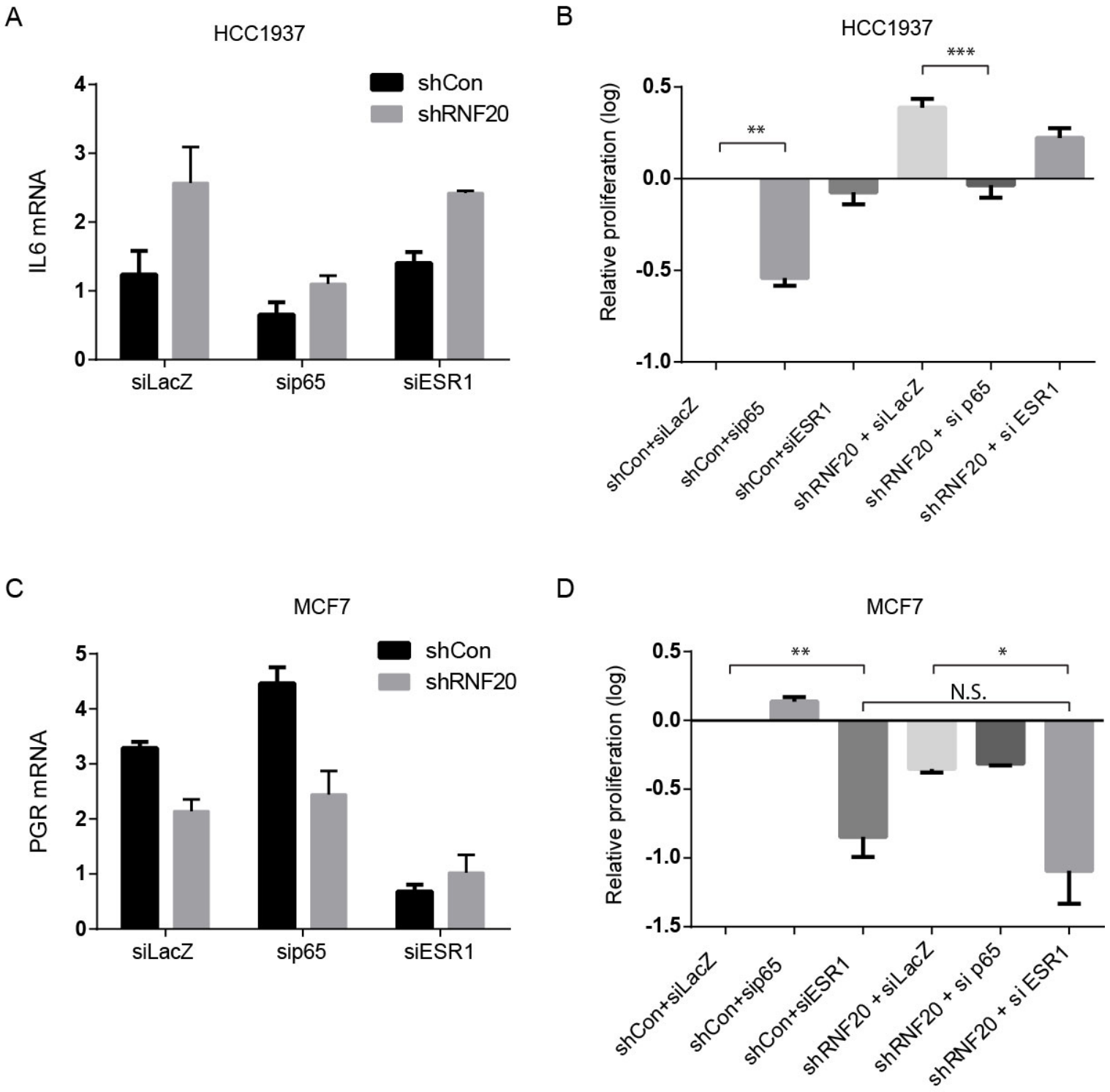

A
HCC1937
shCon
shRNF20
IL6 mRNA
4
3
2
1
0
siLacZ
sip65
siESR1
B
HCC1937
***
**
Relative proliferation (log)
0.5
0.0
-0.5
-1.0
shCon+siLacZ
shCon+sip65
shCon+siESR1
shRNF20 + siLacZ
shRNF20 + si p65
shRNF20 + si ESR1
C
MCF7
shCon
shRNF20
PGR mRNA
5
4
3
2
1
0
siLacZ
sip65
siESR1
D
MCF7
**
*
N.S.
Relative proliferation (log)
0.5
0.0
-0.5
-1.0
-1.5
shCon+siLacZ
shCon+sip65
shCon+siESR1
shRNF20 + siLacZ
shRNF20 + si p65
shRNF20 + si ESR1

Figure S1

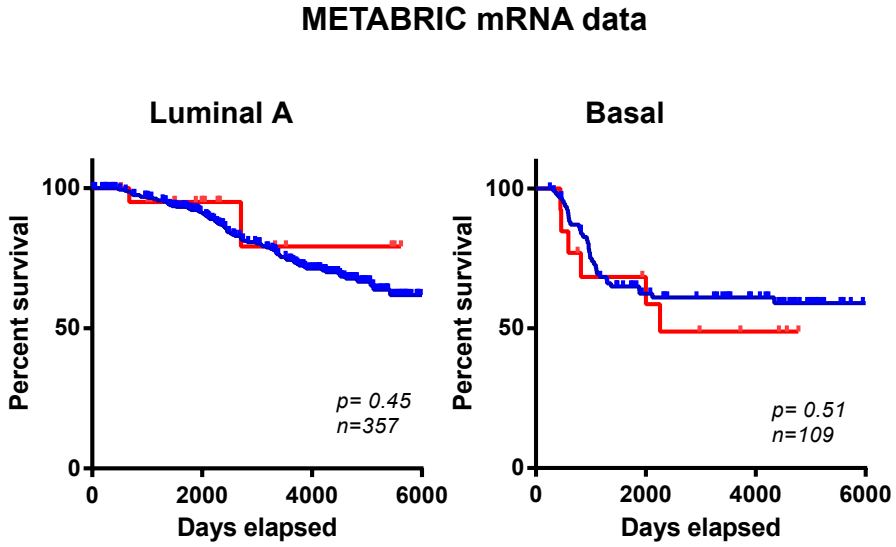
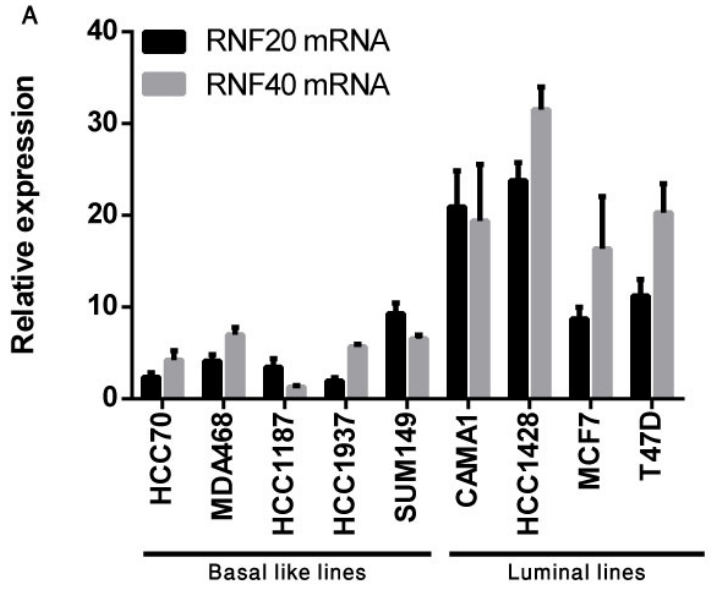
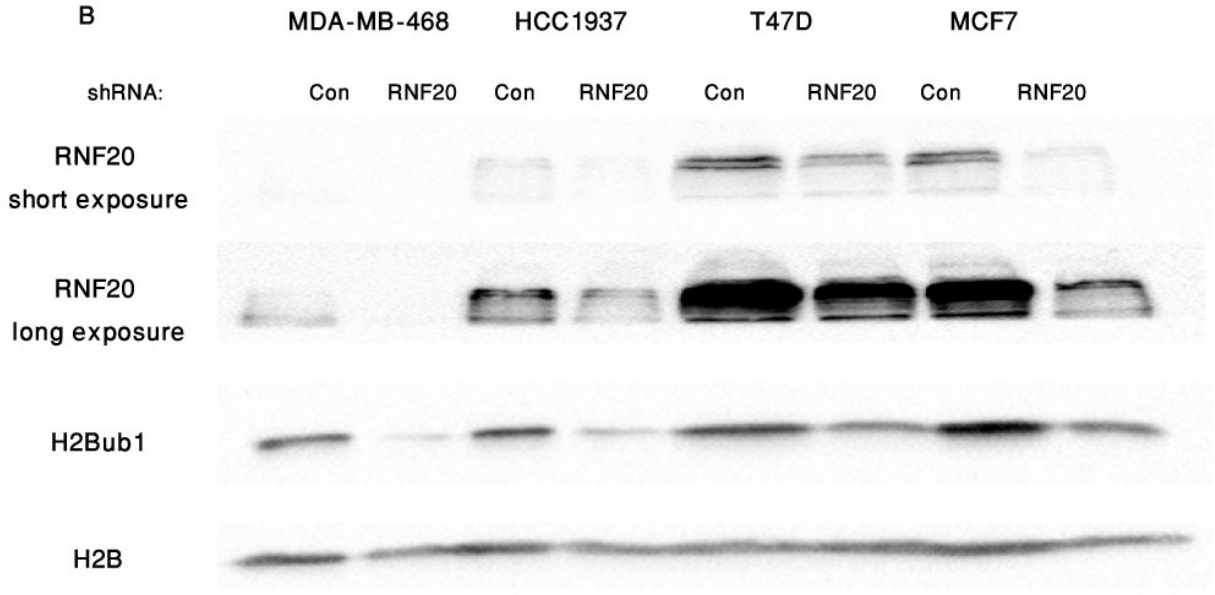
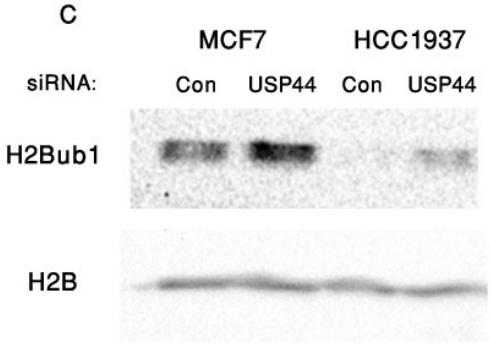

METABRIC mRNA data
Luminal A
Basal
Percent survival
100
50
0
0
2000
4000
6000
Days elapsed
p= 0.45
n=357
p= 0.51
n=109

Figure S2


A
RNF20 mRNA
RNF40 mRNA
Relative expression
40
30
20
10
0
HCC70
MDA468
HCC1187
HCC1937
SUM149
CAMA1
HCC1428
MCF7
T47D
Basal like lines
Luminal lines


B
MDA-MB-468
HCC1937
T47D
MCF7
shRNA:
Con
RNF20
Con
RNF20
Con
RNF20
Con
RNF20
RNF20
short exposure
RNF20
long exposure
H2Bub1
H2B


C
MCF7
HCC1937
siRNA:
Con
USP44
Con
USP44
H2Bub1
H2B

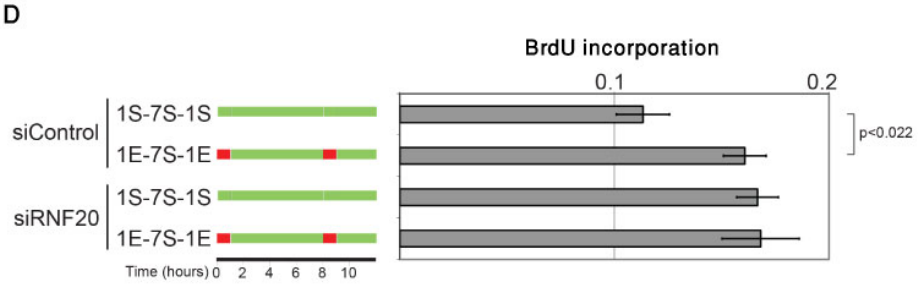

D
BrdU incorporation
0.1
0.2
siControl
1S-7S-1S
1E-7S-1E
siRNF20
1S-7S-1S
1E-7S-1E
Time (hours) 0 2 4 6 8 10
p<0.022

Figure S3

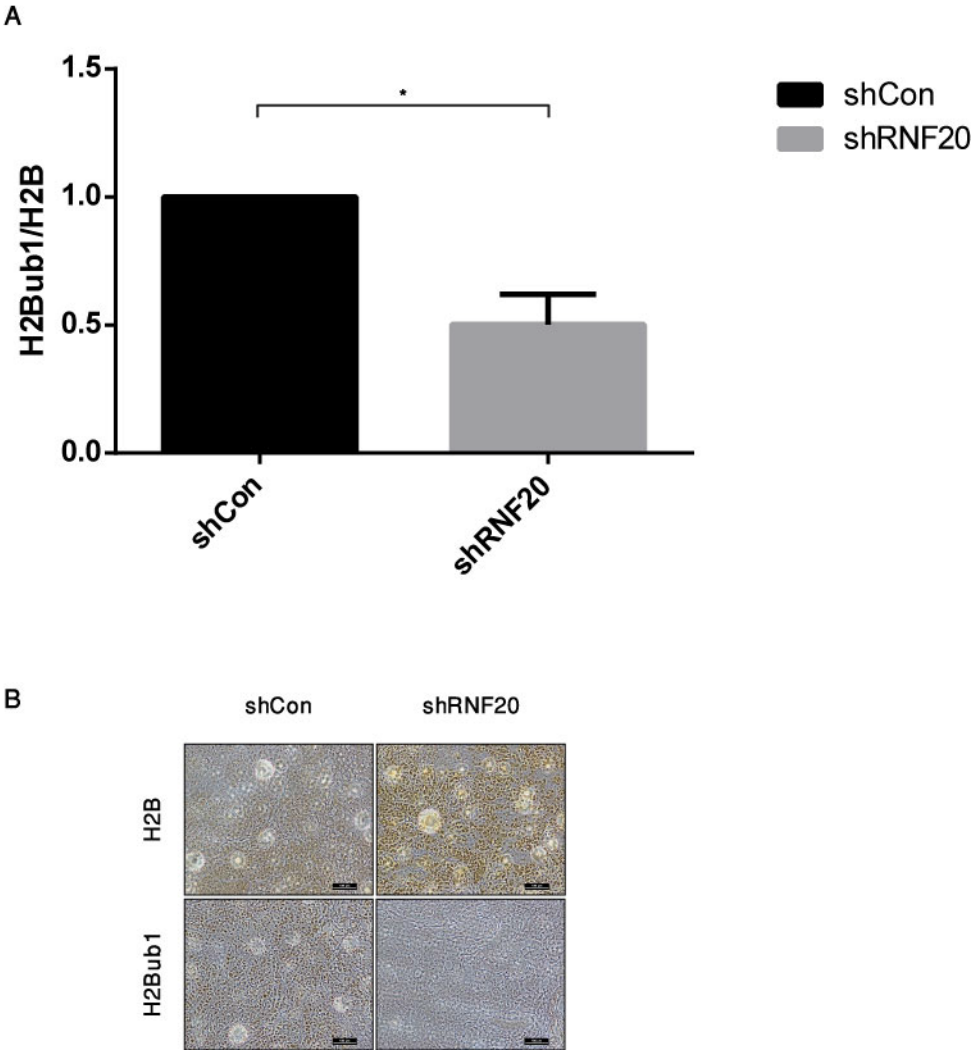
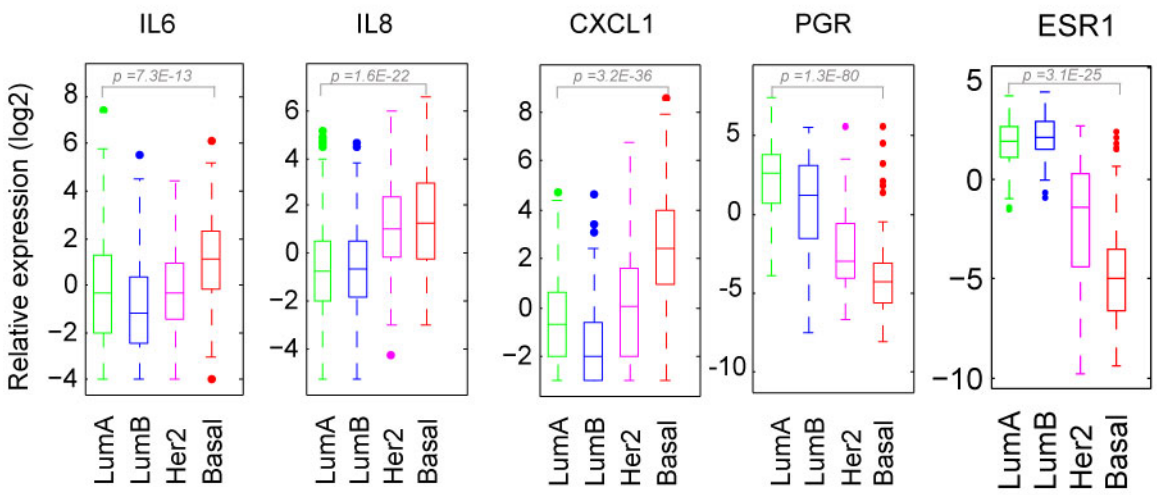

A
1.5
1.0
0.5
0.0
H2Bub1/H2B
*
shCon
shRNF20
shCon
shRNF20
B
shCon
shRNF20
H2B
H2Bub1

Figure S4

A




| Subtype | Samples(n) |
|---|---|
| LumA | 332 |
| LumB | 217 |
| HER2 | 119 |
| Basal | 169 |

B

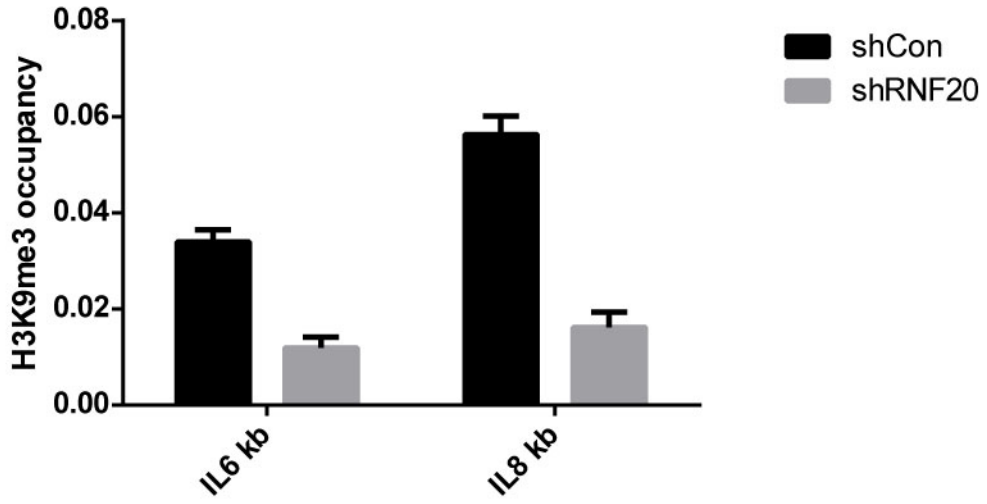


Figure S5

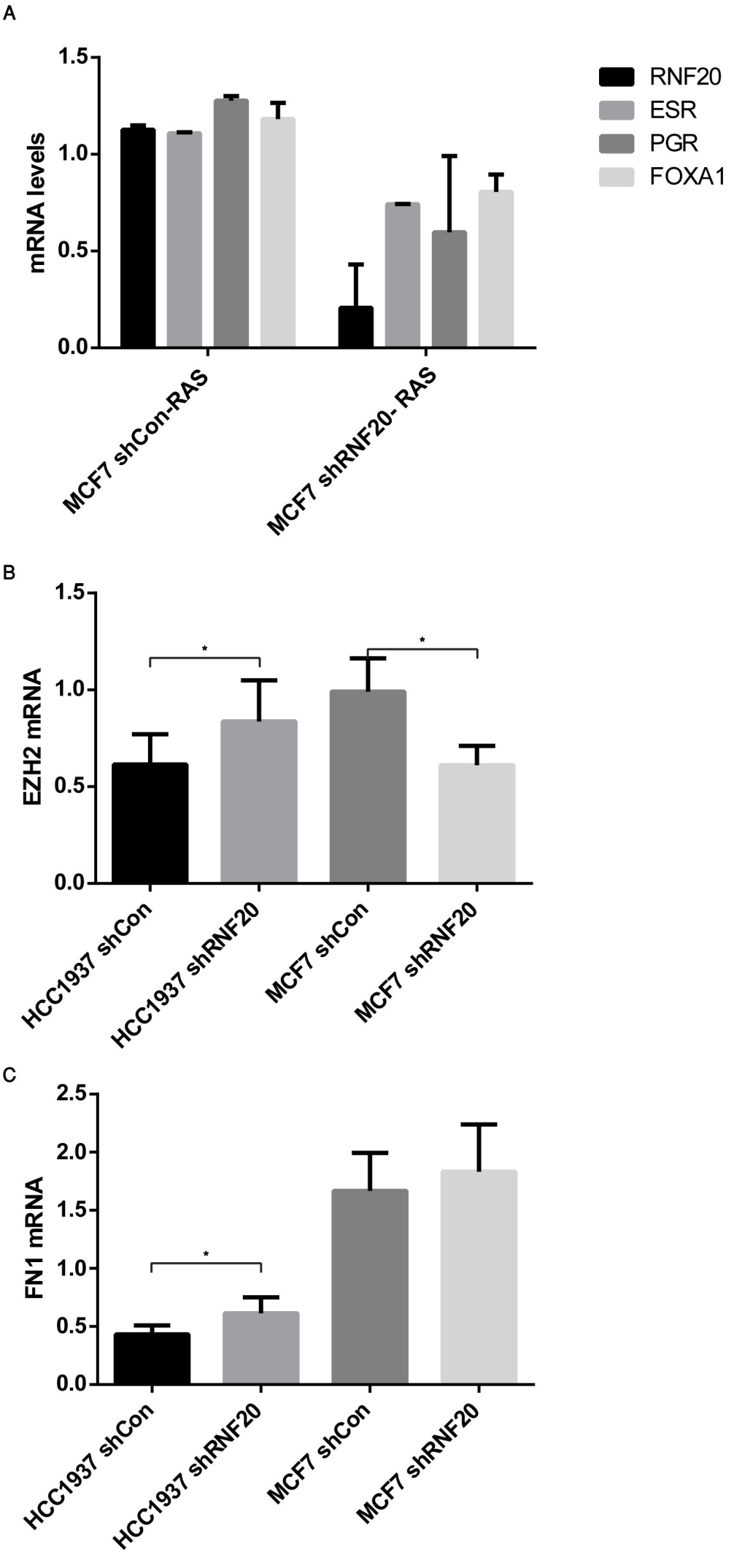
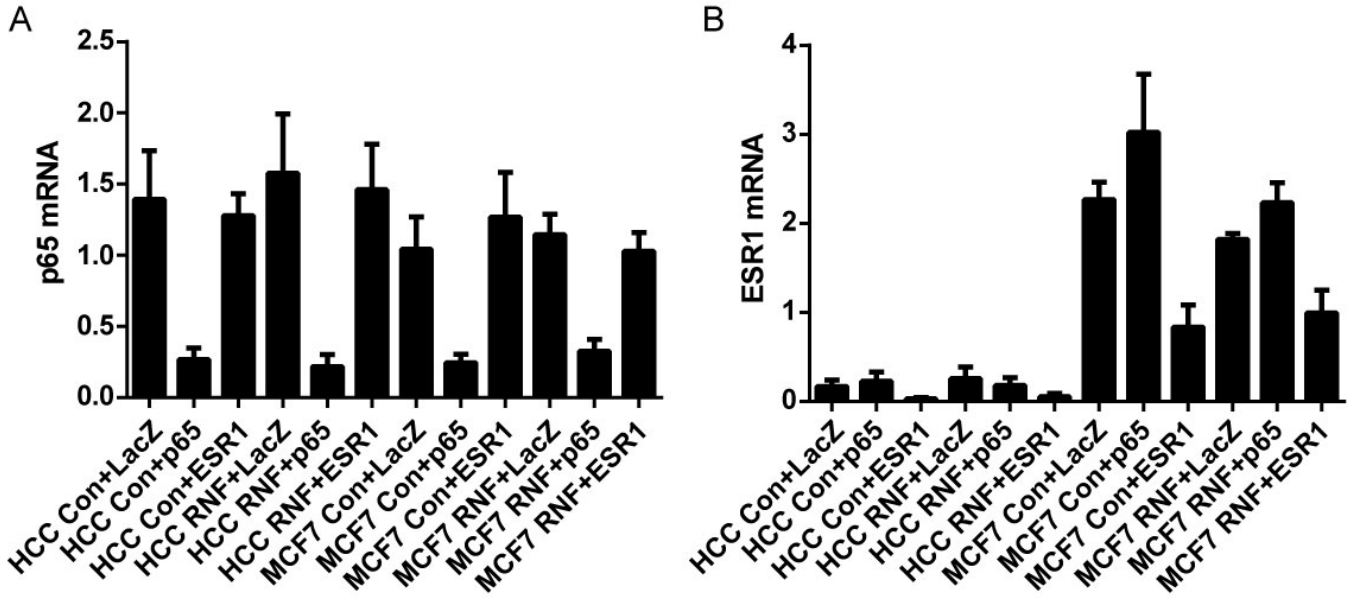

A
1.5
1.0
0.5
0.0
mRNA levels
RNF20
ESR
PGR
FOXA1
MCF7 shCon-RAS
MCF7 shRNF20- RAS
B
1.5
1.0
0.5
0.0
EZH2 mRNA
*
*
HCC1937 shCon
HCC1937 shRNF20
MCF7 shCon
MCF7 shRNF20
C
2.5
2.0
1.5
1.0
0.5
0.0
FN1 mRNA
*
HCC1937 shCon
HCC1937 shRNF20
MCF7 shCon
MCF7 shRNF20

Figure S6

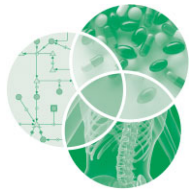

A
p65 mRNA
2.5
2.0
1.5
1.0
0.5
0.0
HCC Con+LacZ
HCC Con+p65
HCC Con+ESR1
HCC RNF+LacZ
HCC RNF+p65
HCC RNF+ESR1
MCF7 Con+LacZ
MCF7 Con+p65
MCF7 Con+ESR1
MCF7 RNF+LacZ
MCF7 RNF+p65
MCF7 RNF+ESR1
B
ESR1 mRNA
4
3
2
1
0
HCC Con+LacZ
HCC Con+p65
HCC Con+ESR1
HCC RNF+LacZ
HCC RNF+p65
HCC RNF+ESR1
MCF7 Con+LacZ
MCF7 Con+p65
MCF7 Con+ESR1
MCF7 RNF+LacZ
MCF7 RNF+p65
MCF7 RNF+ESR1

# Chapter 4 (supplement): Axes of differentiation in breast cancer: untangling stemness, lineage identity, and the epithelial to mesenchymal transition

Status: published in WIREs Syst Biol Med (2014)

R.Z. Granit, M Slyper, I Ben-Porath

Advanced Review



# Axes of differentiation in breast cancer: untangling stemness, lineage identity, and the epithelial to mesenchymal transition

Roy Z. Granit, Michal Slyper and Ittai Ben-Porath*

Differentiation-associated regulatory programs are central in determining tumor phenotype, and contribute to heterogeneity between tumors and between individual cells within them. The transcriptional programs that control luminal and basal lineage identity in the normal mammary epithelium, as well as progenitor and stem cell function, are active in breast cancers, and show distinct associations with different disease subtypes. Also active in some tumors is the epithelial to mesenchymal transition (EMT) program, which endows carcinoma cells with mesenchymal as well as stem cell traits. The differentiation state of breast cancer cells is thus dictated by the complex combination of regulatory programs, and these can dramatically affect tumor growth and metastatic capacity. Breast cancer differentiation is often viewed along an axis between a basal–mesenchymal–stem cell state and a luminal–epithelial–differentiated state. Here we consider the links, as well as the distinctions, between the three components of this axis: basal versus luminal, mesenchymal versus epithelial, and stem cell versus differentiated identity. Analysis on a multidimensional scale, in which each of these axes is assessed separately, may offer increased resolution of tumor differentiation state. Cancer cells possessing a high degree of stemness would display increased capacity to shift between positions on such a multidimensional scale, and to acquire intermediate phenotypes on its different axes. Further molecular analysis of breast cancer cells with a focus on single-cell profiling, and the development of improved tools for dissection of the circuits controlling gene activity, are essential for the elucidation of the programs dictating breast cancer differentiation state. 

**How to cite this article:**
*WIREs Syst Biol Med* 2014, 6:93–106. doi: 10.1002/wsbm.1252

## INTRODUCTION

Coordinated gene expression programs of differentiation and development are central in dictating the biological identity of cancer cells.[1–5] A cancer cell inherits the regulatory circuits active in its cell of origin; subsequently, additional differentiation programs may be activated or silenced in the developing tumor, driven by oncogenic pathways and environmental cues. These events lead to an evolution of the differentiation state of the cancer cells, which can dramatically influence the biological and clinical traits of the tumor. Stem or progenitor-like phenotypes, or the acquisition of a mesenchymal state, can strongly influence tumor growth rates, invasiveness, and metastatic capacity.[6,7] Furthermore, variation in activation of differentiation programs can contribute greatly to tumor heterogeneity, and affect tumor response to therapy.[3]

*Correspondence to: ittaibp@mail.huji.ac.il

Department of Developmental Biology and Cancer Research, Institute for Medical Research – Israel-Canada, Hadassah School of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

Conflict of interest: The authors have declared no conflicts of interest for this article.

Elucidation of the differentiation programs active within cancer cells and the master regulators controlling them is therefore of great importance, and the recent advances in molecular profiling of tumors have provided detailed information regarding gene activity in many cancer types. At the same time, identification of co-regulated genes within a given tumor and of the factors that reside at the top of the regulatory hierarchy, and the elucidation of the manner by which various co-active programs integrate to generate a particular differentiation state, remain challenging analytical tasks.[8] Furthermore, since differentiation state can be highly plastic and fate choices are often made by individual cells, gene expression profiling of single tumor cells is required for an accurate view of cancer cell state, and this technology is only currently emerging.[9]

The importance of differentiation programs is evident in the case of breast cancer. This disease is highly diverse and includes several clearly observable subtypes: the estrogen receptor (ER)-expressing luminal A and B subtypes, the HER2-overexpressing subtype, and the aggressive triple-negative/basal-like subtype.[5] These major tumor groups, as well as additional more minor subtypes, differ in their clinical behavior, mutation profile, and other characteristics.[5,10–13] Importantly, however, they are also distinct in their differentiation state: key lineage markers and regulators of differentiation are differentially expressed across subtypes.[4,5] In fact, the differential expression of such markers served as the basis for identification of novel subtypes (e.g., Basal-like and Claudin-low), while key lineage regulators, such as GATA3, were discovered through their expression in particular tumor types.[5,14–16]

Understanding of cancer cell differentiation state requires detailed elucidation of the cell lineages present in the normal tissue, and of their molecular regulation. While knowledge of the cell types present in the normal mammary epithelium and their functions is still limited, important progress has been made in this field in recent years.[4] Major cell types in the luminal and basal lineages, including stem and progenitor cell populations, have been isolated and profiled in both mouse and human tissue.[17–23] Importantly, this work allowed the comparison of tumor gene expression profiles to those of normal mammary cells, analyses that have provided major insights into the biological nature of breast cancers of the different subtypes and their links to normal stem and progenitor cells.[4,5]

Further insights into breast cancer differentiation state arrived from the study of the epithelial to mesenchymal transition (EMT). EMT is a coordinated program by which epithelial cells undergo trans-differentiation to acquire mesenchymal traits, and which is important in various developmental processes in embryogenesis, such as gastrulation.[24,25] In breast cancer, EMT has been implicated in the generation of cells possessing mesenchymal traits within tumors, and such cells display increased invasiveness and metastatic capacity.[26] Surprisingly, EMT and its regulators have also been shown to promote stem cell identity and function.[27–29] EMT and its regulators are thus central players in the determination of breast cancer cell differentiation state and cancer stem cell function, as well as in the regulation of normal mammary stem cells.

The various differentiation programs active in the breast are therefore linked with one another: stemness is linked to basal lineage identity, while mesenchymal traits induced by EMT go hand in hand with the promotion of basal identity and stem cell function. These links have given rise to a central model often used to describe the differentiation state of breast cancers[4,5]: tumors can be assessed along a linear differentiation axis, on the one end of which are those enriched for cells displaying stem cell-like characteristics, basal-lineage markers, and mesenchymal-like traits, while on the other end are tumors comprised of differentiated luminal epithelial cells (Figure 1(a)).

This linear model is powerful, and allows a link between cancer subtypes and normal cell types. At the same time, in its reduced form, it may represent an oversimplification of the complexity of cancer differentiation state and its control. Stemness, EMT, and basal identity are obviously not equal terms, and their interchangeable use for characterization of breast cancer state can in some contexts lead to confusion. Furthermore, recent experimental findings have provided novel views of plasticity in the normal breast and in cancer cell populations, departing from the hierarchical model of differentiation, and these findings could have important implications on this model.[7,22]

More complex models for the dissection of breast cancer cell differentiation state could potentially offer increased resolution of the contribution of distinct programs to the identity of tumors or individual tumor cells. Differentiation of tumors could be assessed, for example, along several distinct axes, including basal versus luminal identity, mesenchymal versus epithelial identity, and stem cell versus differentiated identity (Figure 1(b)). Such multidimensional models could incorporate the distinctions between these programs, as well as their

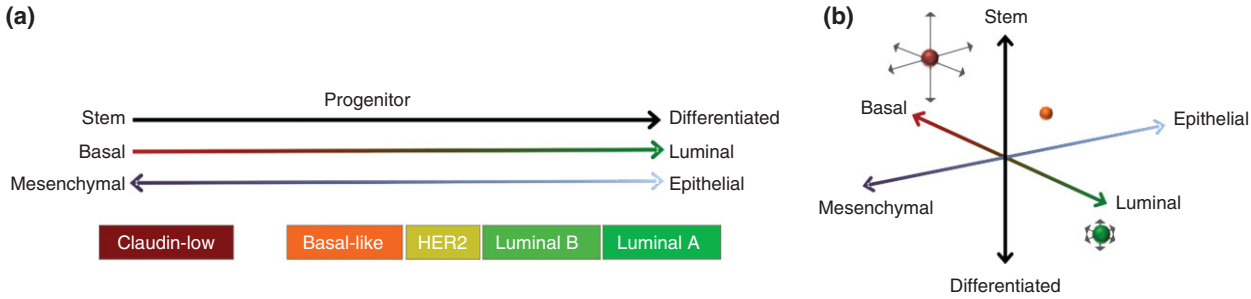



**FIGURE 1** | Axes of differentiation in the normal and cancerous breast. (a) Alignment of breast cancer subtypes (bottom) on a combined linear axis representing linked differentiation states: stem cell to differentiated, basal to luminal, and mesenchymal to epithelial. (b) The same differentiation axes, now positioned in a three dimensional scale, with spheres illustrating potential positions of breast cancer subtypes. The red sphere represents tumors containing cells with high plasticity, which are able to transition (arrows) between differentiation states (e.g., Claudin-low tumors). The green sphere represents tumors with low plasticity, such as Luminal tumors.

interrelatedness, and allow description of intermediate and mixed differentiation states.

In this review we discuss the different views of the differentiation state of breast cancers, assessing the insights obtained from gene expression signature analyses, experimental findings, and patient data. We discuss the interrelationships between basal and luminal identity, EMT, normal stem and progenitor cells and breast cancer stem cells, and the biological implications of the links between these states.

## BREAST EPITHELIAL LINEAGES AND THEIR GENE EXPRESSION SIGNATURES: THE LUMINAL, THE BASAL, AND THE STEM CELL

The mammary gland is organized as a tree-like structure of epithelial ducts embedded in stroma.[4] Units containing lobular structures are interconnected by these ducts, which are composed of two main distinct cell layers: an inner luminal layer containing cuboid cells, and an outer layer of basal, myoepithelial cells, which allow duct contraction (Figure 2(a)). The luminal and basal layers express different markers, including distinct cytokeratins. During pregnancy, luminal cells differentiate into milk-secreting alveolar cells.

In recent years there has been a concerted effort to dissect the lineage hierarchy of the mammary epithelium and identify stem and progenitor cell populations within it. This resulted in the landmark finding that cells within the basal/myoepithelial lineage can function as multipotent mammary stem cells (MaSCs), which are capable of generating multilineage functional mammary epithelia *in vivo*.[17,18,20,21] These cells have thus been viewed as representing the top of the hierarchy of the mammary epithelium, giving rise to more restricted lineage-specific progenitors, and being responsible for the continuous generation of all mammary epithelial lineages in the adult[4] (Figure 2(b)). It currently remains unclear whether cells possessing mammary stem cell potential in the basal layer are distinct from differentiated, non-stem cell basal cells, or whether stem cell potential is a general characteristic of cells in this layer (Figure 2(c)).

The luminal compartment has been shown to contain mature differentiated luminal cells, which are unable to generate structures, as well as a subpopulation of luminal progenitor cells capable of giving rise to structures containing luminal cells[17,18,20,21] (Figure 2(a) and (b)). Both ER-expressing and ER-negative cells are present in the luminal layer, the former comprising 30–50% of cells[30]; the overlap between these two subpopulations and the functional division to mature luminal and luminal progenitor cells remains somewhat poorly defined. It is clear, however, fewer of the progenitor cells express the ER[21] and that the ER-negative fraction is more proliferative[30]; there is evidence that both ER-negative and ER-positive progenitor cell populations exist[23] (Figure 2(b)).

Lineage tracing experiments in the mouse have offered a modified view of the differentiation hierarchy in the breast: while basal cells do indeed possess the inherent potential to regenerate the full mammary epithelium and its two layers, in the adult gland they do not appear to contribute to the luminal layer, but, rather, cells in each of the two layers replenish themselves.[22] Thus, a potential modification of the hierarchical model is that only during embryogenesis MaSCs of the primordial mammary epithelium contribute to both lineages, while in the adult, gland growth and maintenance is performed by lineage restricted progenitors, yet some or all basal cells retain bipotent MaSC potential.[31] Interestingly, when transplanted in the presence of

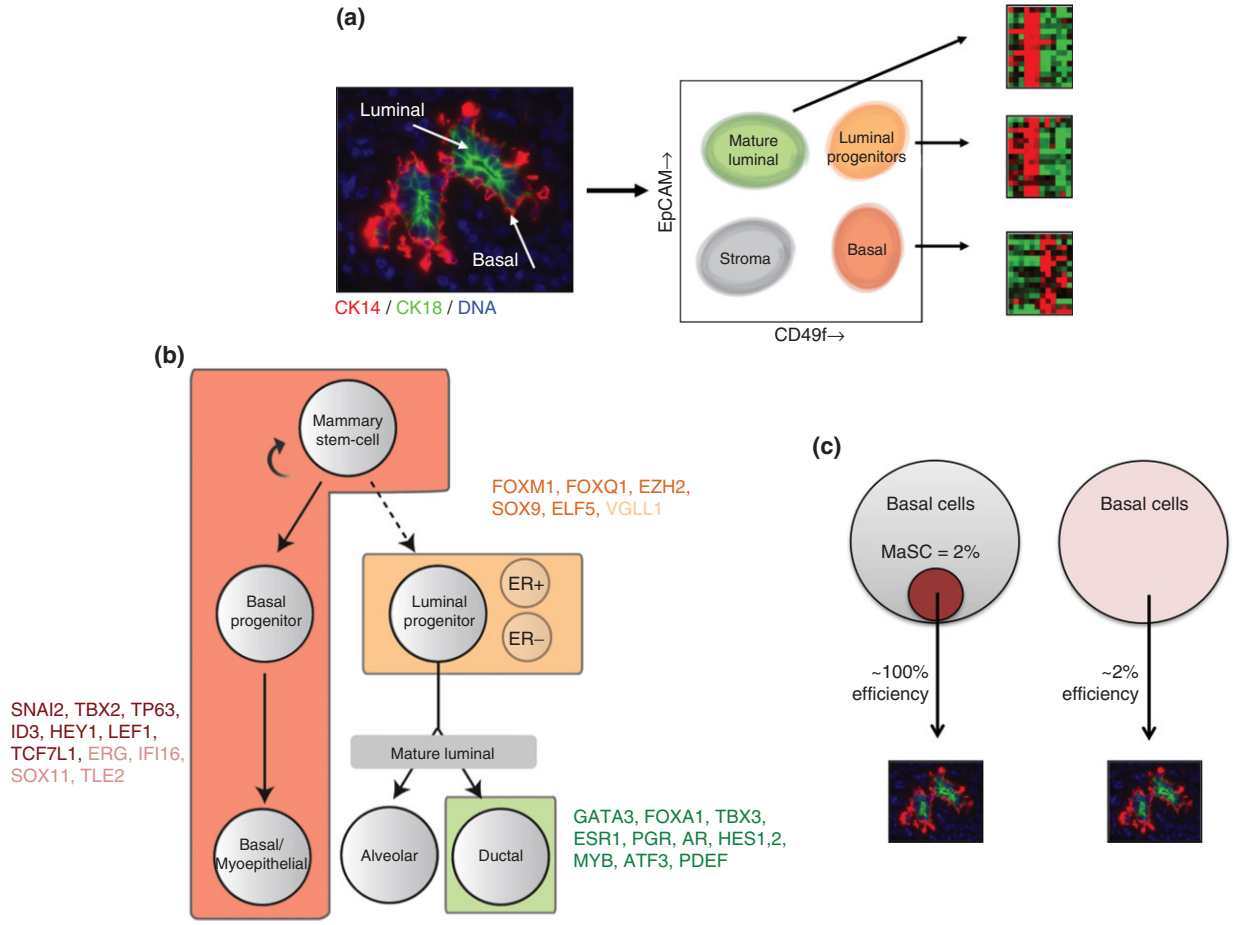



**FIGURE 2 |** Cell lineages in the normal mammary epithelium. (a) Methodology of mammary epithelial lineage separation and profiling. Left: image of normal mammary epithelial ducts, the outer basal/myoepithelial layer is marked by CK14 (red) and the inner luminal layer is marked by CK18 (green). Right: schematic FACS plot representing the separation and isolation of the indicated cell populations by EpCAM and CD49f staining, followed by messenger RNA profiling.[21] (b) A model for the differentiation hierarchy in the normal mammary epithelium. Cell types potentially located in the basal compartment, including bipotent mammary stem cells (MaSCs), are colored red, luminal progenitors are colored orange, and mature luminal cells are colored green. The uncertainty of the contribution of basal MaSCs to the luminal lineage in the adult is marked by a dashed arrow. Potential subtypes of luminal progenitors, e.g., ER+ and ER−, are indicated. Some of the regulatory transcription factors included in the lineage-specific signatures[21] are noted adjacent to each cell type, with dark font indicating genes for which experimental evidence for function in the gland exists. (c) Two possible scenarios representing the stem cell capacity of basal cells: left—a small subpopulation of MaSCs in the basal compartment possesses high gland reconstitution capacity; right—all basal cells possess limited reconstitution capacity.

matrigel, luminal cells are also able, with very low efficiency, to give rise to ductal structures containing both lineages,[32] indicating that more limited MaSC potential may be present in this lineage as well.

While additional potential types of mammary stem and progenitors have been described,[23,33–36] the most widely utilized methods currently allow the isolation, through tissue dissociation and surface marker staining and sorting, of three main populations[37]: mature luminal cells, luminal progenitors, and basal cells, the latter, as noted, containing most of the mammary stem cell potential (Figure 2(a)). The Visvader and Lindeman group isolated these fractions from human individuals, profiled their gene expression, and defined signatures containing genes over- or underexpressed in each population relative to the two others, with each gene receiving a score for its discriminatory power reflecting its differential expression.[21] The Eaves group identified similar signatures relying on slightly different methods of cell isolation and signature definition.[20,38] There is statistically significant overlap in the genes contained in the signatures compiled in these studies, suggesting they indeed represent a sampling of common cell entities and expression programs. However, only 10–20% of genes are shared between signatures representing the

same cells, illustrating the difficulty in generating general high-fidelity signatures from such expression data. Comparison with signatures derived from parallel cell populations isolated from the mouse gland has allowed further compilation of conserved genes.[19,39]

## PATHWAYS AND REGULATORS ACTIVE IN NORMAL MAMMARY EPITHELIAL LINEAGES

Analysis of the genes enriched in each of these cell populations would be expected to reveal some of the regulatory and functional pathways acting in them. Despite the recent development of various enrichment and network analysis tools, it remains quite challenging to derive such insights from signature data. Several important aspects of the molecular characteristics of these different cell types are, however, evident.[4,19,39]

Mature luminal cells express known central regulators of luminal maturation and function, as well as hormonal receptors involved in these processes. These include the estrogen, progesterone, and androgen receptors (ER, PR, and AR, respectively) and the master transcription factors GATA3, FOXA1, and TBX3[40–42] (Figure 2(b)). In addition, components of the Notch pathway, which has been shown to promote luminal differentiation, are active in these cells, including the HES1,2 transcription factors.[4,43]

Characterization of luminal progenitor cells has been highly important both functionally and molecularly. These cells appear to express, in addition to luminal lineage markers, some basal lineage markers, such as keratins 5/6,[21,39] suggesting they maintain some aspects of bi-lineage identity. They also express specific surface markers, including c-KIT and CD133 (in the human gland), and high levels of EGFR. Various transcription factors are associated with these cells, some of which, such as SOX9, FOXM1, EZH2, and ELF5, have been experimentally shown to regulate the progenitor state[29,44–49] (Figure 2(b)).

The basal/myoepithelial signature has most often been termed the MaSC signature, in light of the MaSC potential of cells in the basal compartment. It is important to note, however, that this signature was developed on what appears to be total basal cells rather than a subpopulation distinct from differentiated cells in this layer.[21] How many of these are bona fide stem cells? Mammary epithelial reconstitution by isolated mouse basal cells was achieved with an efficiency on the order of 2%,[17,18] while human cells did so with much lower efficiency.[20,21] It is obvious that reconstitution efficiency is at least partly limited by the technical parameters of these transplantation experiments.[32] It is therefore difficult to determine whether this partial efficiency represents the existence of a small MaSC fraction possessing very high capacity for regeneration, or whether all basal cells possess equal, but limited, stem cell potential (Figure 2(c)). The definition of this signature as that of MaSCs may thus overstate the true stemness represented by it.

The basal (MaSC) expression signature clearly includes genes associated with basal/myoepithelial function,[21,39] such as cell-matrix adhesion proteins (integrins and other focal adhesion components) and extracellular matrix proteins, reflecting the interaction of the basal layer with the basal lamina, as well as proteins associated with the contractile ability of these cells, such as smooth muscle actin. The p63 protein, expressed in many basal epithelia, is prominently expressed in this layer, as well as components of the Wnt pathway such as LEF1 and TCF3 (TCF7L1),[50] consistent with the known role of Wnts in MaSC function[17,51,52] (Figure 2(b)). Importantly, the SLUG (SNAIL2) transcription factor, which is one of the central and powerful activators of EMT, is preferentially expressed in basal cells,[29] illuminating the connection between EMT and basal cells, further discussed below.

## BASAL–LUMINAL IDENTITY AND STEMNESS: A SINGLE AXIS?

The model of differentiation most often applied to the findings above is that of the traditional hierarchal structure, leading from stem to progenitor to differentiated cell (Figure 2(b)). If, however, basal cells are equated with MaSCs, and are thought to continuously contribute to all lineages in the breast, then the following, simpler linear model is derived: MaSC → luminal progenitor → mature luminal (Figure 3(a)). Indeed, in light of the absence of a distinction between differentiated basal cells and MaSCs, this linear flow model is often used, and serves as the basis for assessment of breast cancer cell differentiation.[16]

Basal/myoepithelial versus luminal lineage characteristics can also be viewed as an axis distinct from that of stem versus differentiated: the B↔L axis versus the S↔D axis, the latter representing the degree of plasticity and self-renewal capacity (Figure 3(b)). This type of model allows the consideration of various relationships between the two axes. It may be that the highest degree of stemness indeed resides in cells with 'maximal' basal lineage characteristics (Figure 3(b) left). However, MaSCs might possess

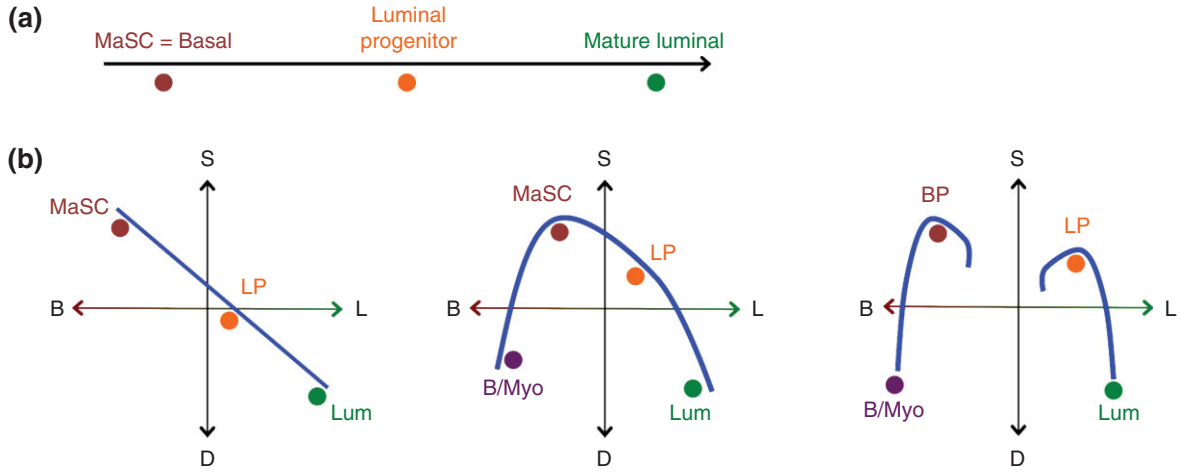



**FIGURE 3** | Relationships between stemness and basal–luminal lineage identity. (a) A differentiation axis showing flow from mammary stem cells (possessing basal identity), to luminal progenitors, to mature luminal cells. (b) Illustration on a two-axis scale of potential relationships between stemness, represented on the S↔D axis, and basal versus luminal lineage identity, represented on the B↔L axis. Left: a linear relationship, identical to (a); center: mammary stem cells (MaSCs) are distinct from differentiated basal/myoepithelial cells and possess an intermediate basal–luminal phenotype; right: the luminal and basal lineages are each maintained by distinct progenitors. B/Myo, basal/myoepithelial; LP, luminal progenitor; BP, basal progenitor.

lineage characteristics that are at an intermediate point on the basal–luminal axis (Figure 3(b), middle). Fetal mammary stem cells (fMaSCs), which give rise to the entire mammary gland, do in fact appear to express both luminal and basal markers, representing an intermediate position on this axis, and their gene expression signature is closest to that of adult luminal progenitors, rather than to adult basal cells.[22,53] Other potential versions of this relationship could take into account the data suggesting that basal cells do not contribute to the luminal lineage, and that some luminal cells also possess bipotent potential (Figure 3(b) right). The assessment of stemness on a separate axis from the axis between full basal or luminal identity may therefore offer additional detail and insight.

## CLASSIFICATION OF BREAST CANCERS AND LINKS TO NORMAL MAMMARY LINEAGES

The routine clinical classification of breast adenocarcinomas is performed mainly by immunohistochemical staining, in which the expression of the ER and the HER2 oncoprotein is examined (the latter is often genomically amplified). This allows for classification of hormone responsive, ER-positive tumors (comprising 70% of cases), and HER2-overexpressing tumors (10–20% of cases), most of which are ER-negative. The remaining 15% of cases are classified as 'triple negative', due to lack of expression of HER2, ER, and the progesterone receptor (usually co-expressed with ER).[4,5]

The advent of gene expression profiling allowed finer resolution and detail in this classification. Clustering of expression profiles of hundreds of tumors revealed four major 'intrinsic subtypes'[5,54,55] (Figure 1(a)): Luminal A and Luminal B subtypes largely correspond to ER-positive tumors, with type B tumors being more proliferative and aggressive; a HER2-like subtype corresponding to HER2 amplified ER-negative tumors, and a Basal-like subtype, representing a majority of tumors clinically defined as triple negative. Additional minor subtypes have been defined over the years, notable among which are Claudin-low tumors, that are enriched for EMT markers, and discussed further below.[16] Recent genomic analyses have also highlighted the existence of a Luminal, ER-positive HER2-positive subtype.[10]

The Perou lab has developed original methods for expression profile-based subtype classification.[5,55] These rely on the definition of genes showing variable expression levels in different subtypes, which serve as 'centroids' for classification of new samples. This method has undergone various modifications, and in its current form has been simplified to a 50-gene classifier (PAM50), which can also be applied in the clinical setting.[56]

What can the analysis of breast cancer gene expression profiles reveal about the biology of the different subtypes, their differentiation state and their links to the normal lineages? The isolation and profiling of the normal mammary cell lineages[21]

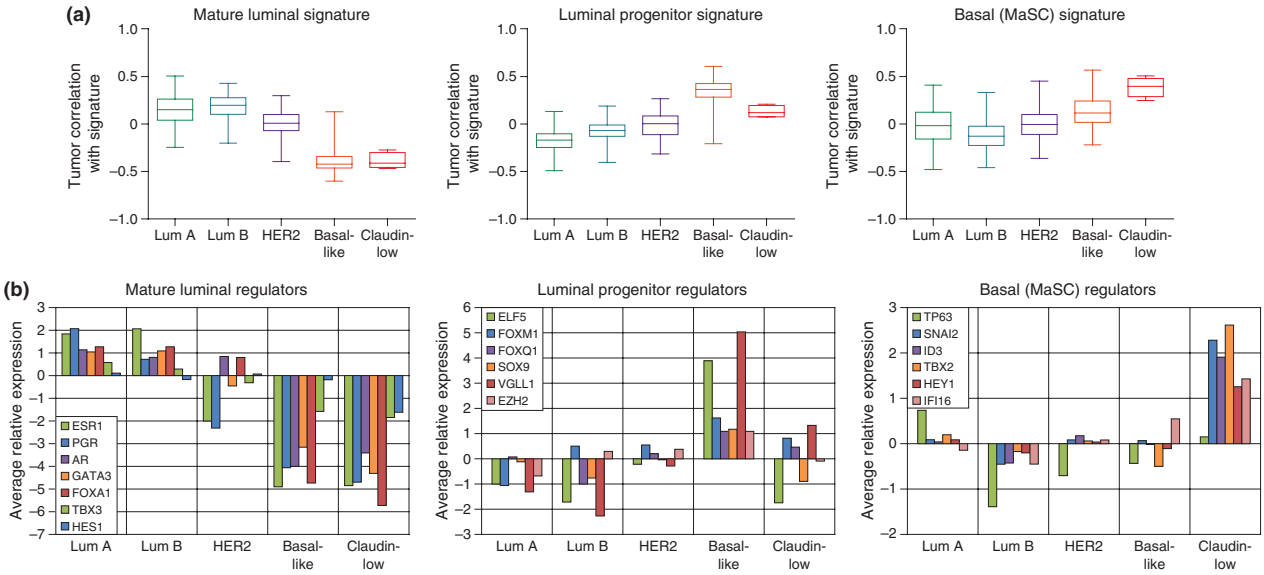



**FIGURE 4** | Breast cancer subtypes and normal mammary lineage signatures. (a) Correlation scores of each of the normal lineage signatures in breast cancers of different subtypes. Similar analyses were originally published in Lim et al.[21] Expression profiles of 918 tumors were obtained from The Cancer Genome Atlas[10] and classified using the PAM50 classifier (https://genome.unc.edu/pubsup/breastGEO/Guide%20to%20Intrinsic%20Subtyping%209-6-10.pdf). A correlation score was calculated for each signature[21] in each tumor, as described in Creighton et al.[59] and box plots represent distribution of the scores of tumors in each subtype. Higher scores indicate a better correlation between the expression levels of signature genes in the tumors and in the normal lineage. (b) Average expression levels in the same breast cancer groups of individual transcription regulators included in each of the lineage signatures. Values are presented in $Log_2$ and normalized to the mean expression across all samples (=0).

opened the door to the assessment of the expression of their specific signatures in tumors of different subtypes. Various methods for testing the expression of gene signatures in tumor samples have been used, highlighting the need for uniformly applied tools for signature comparisons. Simple methods assess the enrichment of signature genes (e.g., 'highly expressed in luminal cells') among genes specifically upregulated in each subtype, or test the average expression of the signature genes in tumors. More detailed methods take into account both up- and downregulated components of a gene signature.[16,21,57,58] An algorithm providing a differentiation score on the MaSC → Luminal progenitor (LP) → Mature luminal scale (L) has also been developed, allowing assessment of individual tumor profiles.[16]

The results provided by these different methods are qualitatively similar. Luminal tumors display the highest enrichment for the expression signature of normal mature luminal cells[5,21] (Figure 4(a)). Luminal subtypes A and B show similar expression levels of this signature, indicating that they both maintain the mature luminal differentiation state. Luminal lineage regulators, such as ESR1 (encoding the ER), GATA3, FOXA1, and TBX3 are highly expressed in these tumors (Figure 4(b)). GATA3 and FOXA1, both central regulators of differentiation of mature luminal cells, are often mutated in luminal tumors (14% in the case of GATA3),[10] suggesting that the pro-differentiation effects of these genes contribute to their tumor suppressive powers.

It is interesting that Luminal B tumors do not show substantially reduced luminal differentiation, despite the increased proliferation rates in them, often driven by mutations in p53, CyclinD1, Myc, and FOXM1. In this case, increased proliferation does not appear to come at the expense of differentiation.

HER2-like tumors also express the mature luminal cell signature, despite their general lack of ER and PR expression[5,21] (Figure 4(a) and (b)). However, these tumors do display a lower score for this signature than Luminal tumors, as well as lower expression levels of the luminal differentiation regulators (Figure 4(b)), indicating that cells within these tumors possess a less differentiated state, perhaps promoted by HER2 itself.

## BASAL-LIKE BREAST CANCERS ARE LINKED TO NORMAL LUMINAL PROGENITORS

One of the most striking observations provided by these analyses regards the biological nature of the basal-like subtype. The aggressive nature of these tumors, the lack of targeted therapy available for their

treatment, and the inherent inaccuracy of the 'triple-negative' definition, all give rise to particular interest in this subtype. Basal-like tumors were originally shown to express markers of both the luminal and basal lineages.[14,60] This provided the basis for the term 'basal-like', and also for the thinking that these tumors may possess mammary stem cell traits. Signature analysis clearly revealed, however, that these tumors score highest for the luminal progenitor gene signature rather than for the basal signature; they also show a very low score for the mature luminal signature[21] (Figure 4(a)). The link to luminal progenitors is also evident by the shared expression of regulatory transcription factors, such as SOX9, ELF5, and others (Figure 4(b)).

Thus, basal-like tumors represent a highly distinct biological identity from that of luminal and HER2 tumors, even though they appear to arise from the luminal layer as well. Experimental studies support the notion that these tumors originate from transformed luminal progenitor cells.[61,62] It is also possible that mutations in p53 and BRCA1, common in these tumors, or increased activation of factors such as EZH2 and FOXM1 during tumorigenesis, contribute to the acquisition of the progenitor-like state of these cancers.[47,63–65] Interestingly, recent analyses have suggested that these tumors share molecular features with serous ovarian tumors[10]; the common underlying forces generating this similarity are highly intriguing.

These signature-based analyses have thus provided fundamental insights into the biological nature of different breast cancers, their links to normal lineages, shared regulatory pathways, and tools to assess these parameters.

## EMT IN BREAST CANCERS: THE M↔E AXIS

EMT, which occurs in various instances during normal embryonic development, has been traditionally viewed as a trans-differentiation program.[24,25] In cancer, this process has been tied to the acquisition of motile and invasive traits by carcinoma cells. While the idea that carcinomas contain nonepithelial tumor cells was initially controversial, various studies have provided evidence that cells with mesenchymal traits do exist within breast cancers, and such cells are most often found in triple-negative tumors.[5,59,66] The fraction of mesenchymal cells is higher among circulating tumor cells isolated from patient blood than among cells in the primary tumor, suggesting that indeed EMT provides an advantage in intravasation.[66] Recent studies have also provided evidence that reversal of EMT (i.e., MET) is required for effective metastatic colonization,[67] pointing to the importance of plasticity in activation and de-activation of this program.

What are the signals that induce epithelial carcinoma cells to undergo this dramatic phenotypic change, which apparently does not occur in the normal mammary gland? EMT could be induced by cell-autonomous activation of its master regulators, such as SLUG, TWIST, and ZEB1/2; this could be driven by oncogenes such as EGFR and MET, by silencing of E-cadherin, or by extrinsic signals such as TGF$\beta$ or WNT, as well as by environmental conditions such as hypoxia.[6] It is likely, however, that some tumor cells, and some cancer subtypes, are more responsive to such signals than others.

The Claudin-low tumor subtype represents a manifestation of the importance of EMT in breast cancer. These tumors, first defined in 2007, are quite rare and are included within the triple-negative category.[10,16,68] They are distinguished by their high levels of expression of mesenchymal markers such as Vimentin and N-Cadherin, and low levels of junction proteins such as E-cadherin and Claudins. EMT-associated regulators are highly expressed in these tumors, including ZEB1,2, TWIST, and SLUG (SNAIL2); a core expression signature of EMT, compiled from mammary epithelial cells induced to undergo this transition by various regulators, is highly expressed in these tumors[16,69] (Figure 5(a) and (b)). Claudin-low tumors therefore appear to contain higher numbers of cells that have undergone a mesenchymal transition through EMT than other subtypes.

The EMT program can thus be viewed as a third axis of differentiation, the M↔E axis (Figure 1), and tumor cells appear to possess different levels of propensity to transition between mesenchymal and epithelial traits.

## THE CONVERGENCE OF STEMNESS, BASAL IDENTITY, AND EMT IN CLAUDIN-LOW TUMORS

The examination of the expression of normal lineage signatures in breast cancers provided the important insight that the tumor subtype scoring highest for the basal/myoepithelial signature (=the MaSC signature) was not, as expected, the basal-like subtype, but, rather, the Claudin-low tumor subtype[21] (Figure 4(a)).

What does the link between the basal signature and Claudin-low tumors signify? One possible interpretation is that these tumors represent a MaSC phenotype. Through this lens, Claudin-low tumors are the least differentiated tumors, and are located on the

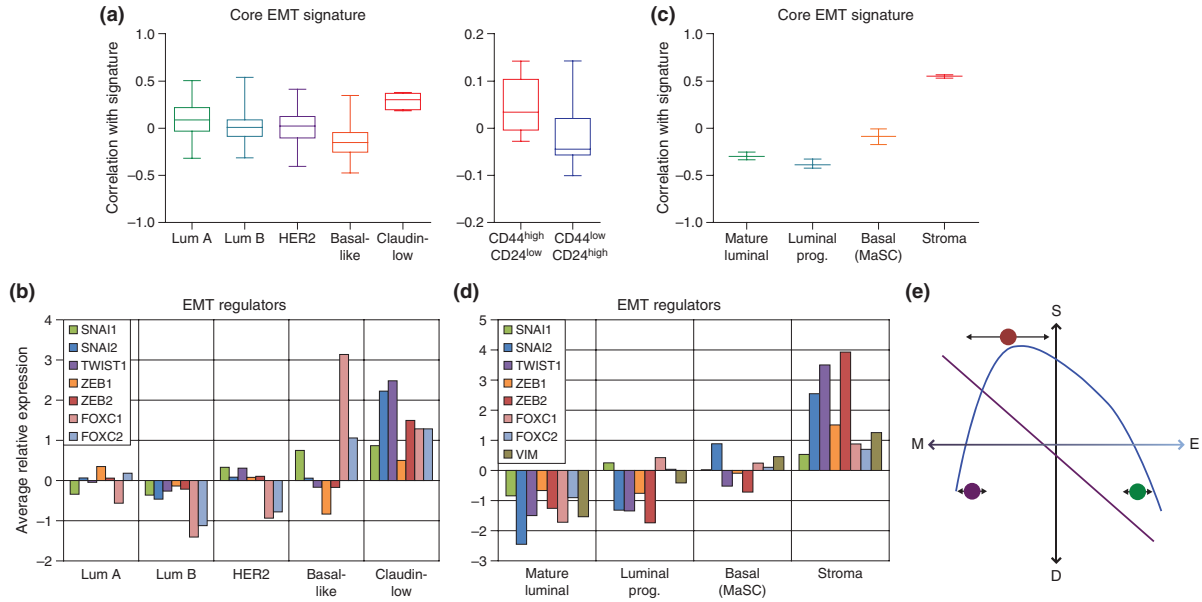



**FIGURE 5** | The epithelial to mesenchymal transition (EMT) signature in the normal and cancerous breast. (a) Correlation scores of the core EMT signature in breast cancers of different subtypes (left), as in Figure 4(a). A similar analysis was originally published in Taube et al.[69] The signature includes genes consistently up or downregulated (twofold or more) in human mammary epithelial cells induced to undergo EMT by different factors.[69] Graph on right shows same analysis on cancer stem cells (CSCs; $CD44^{high}/CD24^{low}$) versus non-CSCs ($CD44^{low}/CD24^{high}$), isolated from breast cancer patients and expression profiled.[70] (b) Average expression levels in the same breast cancer groups of individual known EMT transcriptional regulators. SNAI2=SLUG, SNAI1=SNAIL. (c) Correlation scores of the Core EMT signature in normal breast lineages, expression data from Lim et al.[21] (d) Average expression levels of EMT transcriptional regulators in normal lineages, as well as of the mesenchymal marker Vimentin (VIM). (e) Scheme of two possible relationships between EMT and stemness. The purple line indicates a linear correlation with mesenchymal cells showing the highest level of stemness; the blue line shows the highest degree of stemness at an intermediate position on the M↔E axis, with low plasticity of fully mesenchymal or epithelial cells.

left side of the MaSC → LP → L axis (Figure 3(a)). It could also be inferred that these tumors arise from cells in the basal layer: either a 'common' basal/myoepithelial cell, or a rare MaSC in this layer. Once again the question arises as to whether basal cells are equated with MaSCs or whether MaSCs are a minute subpopulation of the basal compartment, since, as stated above, a small fraction of MaSCs is unlikely to contribute significantly to the basal signature. There seems to be much evidence, however, to indicate that breast adenocarcinomas rarely if ever arise from basal cells: experimentally transformed mouse or human basal cells do not give rise to adenocarcinomas, but, rather, to squamous cell carcinomas, metaplasias or adenomyoepithliomas.[61,62] Human myoepitheliomas, (and their malignant counterparts adenomyoepithliomas) maintain myoepithelial characteristics and are likely to arise from basal cells.[71] Nevertheless, Claudin-low tumors express some regulators of basal lineage identity, such as ID3 and TBX2 (Figure 4(b)), suggesting that at least some cells in these tumors are capable of activating the basal differentiation program, even if they do not arise from this layer.

Another obvious component linking basal cells and Claudin-low tumors is EMT. Basal cells, by virtue of their myoepithelial nature, share some traits with mesenchymal cells: they possess a more spindle-like morphology than their luminal counterparts, are contractile, secrete ECM factors, and express Vimentin to a moderate degree.[21,71] Accordingly, basal cells also express a higher level of the core EMT gene signature than the other mammary cell fractions, as well as the EMT regulator SLUG (Figure 5(c) and (d)).

Importantly, however, basal cells are not fibroblasts, and are distinct from epithelial cells that have undergone a full EMT. They maintain central aspects of epithelial identity, forming adherens junctions (relying on P-cadherin, rather than E-cadherin) and expressing cytokeratins specific to this layer (such as CK5 and 14). While SLUG is prominently expressed in basal cells, ZEB1/2, and TWIST are apparently not significantly expressed in these cells, in contrast to EMT'd cells (Figure 5(d)). In comparison, stromal cells isolated from normal mammary gland, which include mammary fibroblasts, show a much higher score for the 'Core EMT' signature (Figure 5(c) and (d)).

The expression in basal cells of components of the EMT program therefore no doubt contributes to their apparent association with the Claudin-low tumors, yet this shared feature of expression may not be a consequence of their being the cells of origin for this subtype.

## LINKS BETWEEN EMT AND STEM CELL FUNCTION IN THE NORMAL AND CANCEROUS BREAST

The view of EMT primarily as a course for cancer cells to acquire mesenchymal traits assisting metastasis has changed drastically in light of compelling demonstrations that EMT can promote stem cell function. These findings highlight the relationship between the M↔E axis and the S↔D and B↔L axes, and justify their co-alignment.

Recent work has shown that SLUG is not merely a marker of basal cells, but is, in fact, essential for the functioning of these cells as mammary stem cells: in its absence mouse basal cells are unable to reconstitute the mammary epithelium.[29] Furthermore, Slug overexpression, when performed together with Sox9, converts differentiated luminal cells into basal cells, and at the same time endows them with MaSC function, namely gland reconstitution capacity.[29] It is unclear which components of the EMT program are activated by Slug in basal cells; however, these findings indicate that there is overlap between its activity as a regulator of EMT, of basal identity, and of stem cell function. Furthermore, they suggest that the partially mesenchymal state of the myoepithelial cells may contribute to their stem cell capacity.

Cancer stem cells (CSCs) are defined as a subpopulation of tumor cells that possess increased ability to initiate tumor formation, and that can self-renew as well as give rise to non-tumor-initiating cells.[72] Breast cancers were the first solid tumors in which CSCs were identified: a subpopulation of cells carrying the $CD44^{high}/CD24^{low}$ marker profile were shown to possess increased tumor initiation and differentiation capacity.[73] Strikingly, it became clear that the carcinoma cells bearing this profile possess some mesenchymal-like traits, and that activation of EMT regulators in human mammary epithelial cells, including SLUG, TWIST, and others, induces them to adopt the $CD44^{high}/CD24^{low}$ marker profile.[27] This conversion also endows the cells with sphere-formation capacity, increased tumor initiation and metastatic potential, and increased drug resistance, all traits of CSCs.[27,74] Expression profiling of breast CSCs[70] indicates their enrichment for the EMT signature (Figure 5(a)), and Claudin-low tumors are particularly enriched for the $CD44^{high}/CD24^{low}$ derived signature.[5,57,59]

Whether breast cancer stem cells and normal MaSCs are directly linked is unclear, and it is possible that the $CD44^{high}/CD24^{low}$ profile is primarily a reflection of a mesenchymal state achieved by 'plastic' cancer cells. Nevertheless, the apparent ability of EMT and its regulators to control definitive stem cell functions is highly compelling.

## PLASTICITY AND ITS CONTROL

On the basis of the observations above, mesenchymal cells (on the M side of the M↔E axis) score high on the stemness S↔D scale, while cells that are fully epithelial have reduced plasticity and self-renewal. Does the highest level of stemness—self-renewal capacity and plasticity—indeed lie at the extreme M end of the M↔E axis? It is possible that a fully mesenchymal identity in fact comes at the expense of stemness, and that it is an intermediate phenotype that is associated with a high degree of plasticity (Figure 5(e)). Some widely used cultured breast cancer cell lines, such as MDA-MB-231 cells (of the group defined as Basal B[75]), display a fully EMT'd phenotype, with no cytokeratin or E-cadherin expression, and virtually all cells in these populations display a $CD44^{high}/CD24^{low}$ profile.[76] While MDA-MB-231 cells and other similar lines are often highly metastatic and possess high tumor initiating potential, they strictly form sarcoma-like tumors, and, generally, do not spontaneously give rise in culture to epithelial cells. These cells therefore do not possess plasticity on the M↔E axis, and seem to be locked in the mesenchymal state. It has been shown that this is enforced by chromatin structure regulation, e.g., by factors such as the H3K9 methylase G9a and the H3K27 methylase EZH2, which both contribute to maintaining E-cadherin silencing.[77,78] A fully mesenchymal state therefore does not necessarily offer increased plasticity.

In contrast, cell lines which maintain similarity to basal-like breast cancers in their marker and gene expression profiles (of the group termed Basal A[75]) display a higher degree of diversity and plasticity. These lines largely maintain epithelial identity (i.e., their carcinoma-like state), express E-cadherin alongside basal, and luminal cytokeratins, as well as markers of luminal progenitors. Within these culture populations, however, one can find cells co-expressing Vimentin and cytokeratins, cells co-expressing luminal, and basal cytokeratins, as well as varying numbers of $CD44^{high}/CD24^{low}$ cells.[5,64,66,76,79] It has recently been shown that cells in these lines are able to transition between

differentiation states, and that this plasticity relies on the stemness/EMT transcription factor ZEB1,[80] whose regulatory elements are bound by chromatin in a bivalent state, allowing it to be transiently expressed or repressed. In contrast, ZEB1 is locked in a silenced chromatin state in luminal epithelial differentiated lines, and therefore such cells are unable to transition toward the CSC state. Chromatin regulation is thus central in controlling plasticity, allowing transitions on both the B↔L and the M↔E axes. As noted above, plasticity appears to be an essential feature of EMT in cancer, and this plasticity could be controlled by additional forces, such as microRNAs, the inflammatory microenvironment, and others.[67,81,82]

## CONCLUSION: ONE AXIS OR THREE?

Untangling stemness, lineage identity, and EMT in breast cancer, as well as in the normal breast, is a complex task. The discussion above illustrates the links between the B↔L, M↔E and S↔D axes, providing the basis for their alignment as an informative tool for the assessment of tumor differentiation state (Figure 1(a)). The distinctions between these programs, also highlighted above, are no less important, and a higher resolution of differentiation state analysis requires their individual assessment (Figure 1(b)).

A multidimensional scale could provide a tool for assessing differentiation state transitions. Tumor cells with a high degree of plasticity could transition on the M↔E axis as well as on the B↔L, acquiring a variety of mixed and intermediate phenotypes. Indeed, a hallmark of Claudin-low tumors and Basal-like tumors is such cellular heterogeneity, and the presence of cells with intermediate differentiation states.[5,64,66] Conceivably, such tumor cells could also transition on the S↔D axis, and when reaching the D side of this axis they might lock into a particular position on the differentiation scale (Figure 1(b)). Cells in differentiated subtypes, such as Luminal A tumors, score high on the luminal and epithelial ends, and in light of their low score on the stemness scale, these tumor cells would be expected to show little differentiation diversity (Figure 1(b)).

Much further insight into breast cancer differentiation state is likely to be gained in the near future from the analysis of single-cell transcriptomes and of the dynamics of cancer and normal cell phenotypic transitions. The revelation of additional differentiation programs active in tumor cells will require sophisticated multidimensional tools for analysis of differentiation state.

## ACKNOWLEDGMENTS

We thank Reba Condiotti for critical review of the manuscript. Research related to the topics of this review is conducted under support of project grants from the Israel Science Foundation (Grant 859/09), Israel Cancer Research Fund (Project Grant), Niedersachsen-Israel Cooperative Research (Project Grant), US-Israel Binational Science Foundation (Grant 2007459), and German-Israel Foundation (Grant 1115/2010).

## REFERENCES

1. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature* 2001, 414:105–111.
2. Visvader JE. Cells of origin in cancer. *Nature* 2011, 469:314–322.
3. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 2012, 12:323–334.
4. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev* 2009, 23:2563–2577.
5. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* 2011, 5:5–23.
6. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* 2009, 9:265–273.
7. Marjanovic ND, Weinberg RA, Chaffer CL. Cell plasticity and heterogeneity in cancer. *Clin Chem* 2013, 59:168–179.
8. Kim HD, Shay T, O'Shea EK, Regev A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* 2009, 325:429–432.
9. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 2013, 14:618–630.
10. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490:61–70.

11. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012, 486:400–404.
12. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486:346–352.
13. Ellis MJ, Perou CM. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discov* 2013, 3:27–34.
14. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. Molecular portraits of human breast tumours. *Nature* 2000, 406:747–752.
15. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, Langerod A, Karesen R, Oh DS, Dressler LG, et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004, 23:7669–7678.
16. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 2010, 12:R68.
17. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, Wu L, Lindeman GJ, Visvader JE. Generation of a functional mammary gland from a single stem cell. *Nature* 2006, 439:84–88.
18. Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, Eaves CJ. Purification and unique properties of mammary epithelial stem cells. *Nature* 2006, 439:993–997.
19. Kendrick H, Regan JL, Magnay FA, Grigoriadis A, Mitsopoulos C, Zvelebil M, Smalley MJ. Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* 2008, 9:591.
20. Eirew P, Stingl J, Raouf A, Turashvili G, Aparicio S, Emerman JT, Eaves CJ. A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. *Nat Med* 2008, 14:1384–1389.
21. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* 2009, 15:907–913.
22. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C. Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 2011, 479:189–193.
23. Shehata M, Teschendorff A, Sharp G, Novcic N, Russell A, Avril S, Prater M, Eirew P, Caldas C, Watson CJ, et al. Phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res* 2012, 14:R134.
24. Yang J, Weinberg RA. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* 2008, 14:818–829.
25. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell* 2009, 139:871–890.
26. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science* 2011, 331:1559–1564.
27. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008, 133:704–715.
28. Brabletz S, Brabletz T. The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer? *EMBO Rep* 2010, 11:670–677.
29. Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, Itzkovitz S, Noske A, Zurrer-Hardi U, Bell G, et al. Slug and sox9 cooperatively determine the mammary stem cell state. *Cell* 2012, 148:1015–1028.
30. Brisken C. Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. *Nat Rev Cancer* 2013, 13:385–396.
31. Keller PJ, Arendt LM, Kuperwasser C. Stem cell maintenance of the mammary gland: it takes two. *Cell Stem Cell* 2011, 9:496–497.
32. Vaillant F, Lindeman GJ, Visvader JE. Jekyll or Hyde: does Matrigel provide a more or less physiological environment in mammary repopulating assays? *Breast Cancer Res* 2011, 13:108.
33. Pece S, Tosoni D, Confalonieri S, Mazzarol G, Vecchi M, Ronzoni S, Bernard L, Viale G, Pelicci PG, Di Fiore PP. Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* 2010, 140:62–73.
34. Bai L, Rohrschneider LR. s-SHIP promoter expression marks activated stem cells in developing mouse mammary tissue. *Genes Dev* 2010, 24:1882–1892.
35. Roy S, Gascard P, Dumont N, Zhao J, Pan D, Petrie S, Margeta M, Tlsty TD. Rare somatic cells from human breast tissue exhibit extensive lineage plasticity. *Proc Natl Acad Sci USA* 2013, 110:4598–4603.
36. Sale S, Lafkas D, Artavanis-Tsakonas S. Notch2 genetic fate mapping reveals two previously unrecognized mammary epithelial lineages. *Nat Cell Biol* 2013, 15:451–460.
37. Smalley MJ, Kendrick H, Sheridan JM, Regan JL, Prater MD, Lindeman GJ, Watson CJ, Visvader JE, Stingl J. Isolation of mouse mammary epithelial subpopulations: a comparison of leading methods. *J Mammary Gland Biol Neoplasia* 2012, 17:91–97.
38. Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, Iscove N, Jones S, McKinney S, Emerman J, et al.

Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* 2008, 3:109–118.

39. Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK, Visvader JE. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res* 2010, 12:R21.
40. 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.
41. Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* 2007, 9:201–209.
42. Douglas NC, Papaioannou VE. The T-box transcription factors TBX2 and TBX3 in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia* 2013, 18:143–147.
43. Bouras T, Pal B, Vaillant F, Harburg G, Asselin-Labat ML, Oakes SR, Lindeman GJ, Visvader JE. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. *Cell Stem Cell* 2008, 3:429–441.
44. Oakes SR, Naylor MJ, Asselin-Labat ML, Blazek KD, Gardiner-Garden M, Hilton HN, Kazlauskas M, Pritchard MA, Chodosh LA, Pfeffer PL, et al. The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev* 2008, 22:581–586.
45. Chakrabarti R, Wei Y, Romano RA, DeCoste C, Kang Y, Sinha S. Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. *Stem Cells* 2012, 30:1496–1508.
46. Chakrabarti R, Hwang J, Andres Blanco M, Wei Y, Lukacisin M, Romano RA, Smalley K, Liu S, Yang Q, Ibrahim T, et al. Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nat Cell Biol* 2012, 14:1212–1222.
47. Carr JR, Kiefer MM, Park HJ, Li J, Wang Z, Fontanarosa J, DeWaal D, Kopanja D, Benevolenskaya EV, Guzman G, et al. FoxM1 regulates mammary luminal cell fate. *Cell Rep* 2012, 1:715–729.
48. Michalak EM, Nacerddine K, Pietersen A, Beuger V, Pawlitzky I, Cornelissen-Steijger P, Wientjens E, Tanger E, Seibler J, van Lohuizen M, et al. The Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal progenitor pool. *Stem Cells* 2013.
49. Pal B, Bouras T, Shi W, Vaillant F, Sheridan JM, Fu N, Breslin K, Jiang K, Ritchie ME, Young M, et al. Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. *Cell Rep* 2013, 3:411–426.
50. Slyper M, Shahar A, Bar-Ziv A, Granit RZ, Hamburger T, Maly B, Peretz T, Ben-Porath I. Control of breast cancer growth and initiation by the stem cell-associated transcription factor TCF3. *Cancer Res* 2012, 72:5613–5624.
51. Brennan KR, Brown AM. Wnt proteins in mammary development and cancer. *J Mammary Gland Biol Neoplasia* 2004, 9:119–131.
52. Zeng YA, Nusse R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* 2010, 6:568–577.
53. 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.
54. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001, 98:10869–10874.
55. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006, 7:96.
56. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol* 2012, 9:48–57.
57. Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee JS, Fridlyand J, Sahin A, Agarwal R, Joy C, et al. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res* 2009, 69:4116–4124.
58. Wu D, Lim E, Vaillant F, Asselin-Labat ML, Visvader JE, Smyth GK. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* 2010, 26:2176–2182.
59. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, Rimm DL, Wong H, Rodriguez A, Herschkowitz JI, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci USA* 2009, 106:13820–13825.
60. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006, 19:264–271.
61. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, Mackay A, Grigoriadis A, Tutt A, Ashworth A, 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.

62. Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, Smith AE, Prat A, Perou CM, Gilmore H, et al. Defining the cellular precursors to human breast cancer. *Proc Natl Acad Sci USA* 2011, 109:2772–2777.

63. Liu S, Ginestier C, Charafe-Jauffret E, Foco H, Kleer CG, Merajver SD, Dontu G, Wicha MS. BRCA1 regulates human mammary stem/progenitor cell fate. *Proc Natl Acad Sci USA* 2008, 105:1680–1685.

64. Granit RZ, Gabai Y, Hadar T, Karamansha Y, Liberman L, Waldhorn I, Gat-Viks I, Regev A, Maly B, Darash-Yahana M, et al. EZH2 promotes a bi-lineage identity in basal-like breast cancer cells. *Oncogene* 2013, 32:3886–3895.

65. Knight JF, Lesurf R, Zhao H, Pinnaduwage D, Davis RR, Saleh SM, Zuo D, Naujokas MA, Chughtai N, Herschkowitz JI, et al. Met synergizes with p53 loss to induce mammary tumors that possess features of claudin-low breast cancer. *Proc Natl Acad Sci USA* 2013, 110:E1301–E1310.

66. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* 2013, 339:580–584.

67. Tsai JH, Donaher JL, Murphy DA, Chau S, Yang J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* 2012, 22:725–736.

68. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 2007, 8:R76.

69. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, et al. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci USA* 2010, 107:15449–15454.

70. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, Sherlock G, Lewicki J, Shedden K, Clarke MF. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007, 356:217–226.

71. Moumen M, Chiche A, Cagnet S, Petit V, Raymond K, Faraldo MM, Deugnier MA, Glukhova MA. The mammary myoepithelial cell. *Int J Dev Biol* 2011, 55:763–771.

72. Magee JA, Piskounova E, Morrison SJ. Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell* 2012, 21:283–296.

73. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA* 2003, 100:3983–3988.

74. Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA, Lander ES. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 2009, 138:645–659.

75. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 2006, 10:515–527.

76. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res* 2008, 10:R25.

77. Dong C, Wu Y, Yao J, Wang Y, Yu Y, Rychahou PG, Evers BM, Zhou BP. G9a interacts with Snail and is critical for Snail-mediated E-cadherin repression in human breast cancer. *J Clin Invest* 2012, 122:1469–1486.

78. Tiwari N, Tiwari VK, Waldmeier L, Balwierz PJ, Arnold P, Pachkov M, Meyer-Schaller N, Schubeler D, van Nimwegen E, Christofori G. Sox4 is a master regulator of epithelial-mesenchymal transition by controlling ezh2 expression and epigenetic reprogramming. *Cancer Cell* 2013, 23:768–783.

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

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

81. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, Reichert M, Beatty GL, Rustgi AK, Vonderheide RH, et al. EMT and dissemination precede pancreatic tumor formation. *Cell* 2012, 148:349–361.

82. Chang CJ, Chao CH, Xia W, Yang JY, Xiong Y, Li CW, Yu WH, Rehman SK, Hsu JL, Lee HH, et al. p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nat Cell Biol* 2011, 13:317–323.

# 3. Discussion

The composition of the cancerous tissue is dynamic, and cell subpopulations can acquire distinct phenotypes, and, potentially, diverse functional roles that allow synergism. This heterogeneity greatly contributes disease progression and resistance to treatment (Cheung et al., 2013; Meacham and Morrison, 2013; Wood, 2016). Thus, uncovering the phenotypes of distinct cell populations in tumors, and the elucidation of the molecular mechanisms regulating tumor heterogeneity will have valuable implications for cancer diagnosis and treatment. However, the underlying regulatory mechanisms controlling cell identity require much further research.

I began my work with an attempt to identify the regulators of tumor subtype identity, identifying candidates based on tumor gene expression profiles. My experimental work identified EZH2 as defining factor of basal-like identity and driver of progenitor-like traits, mediated by the repression the luminal regulators GATA3 and FOXA1. Further study revealed that EZH2 regulates the composition of the tumor cell population, and promotes the expansion of the $K18^{+}K14^{+}$, progenitor-like, population. These findings shifted my focus to the study of intratumoral heterogeneity. I uncovered novel insights about the functional implications of intratumoral heterogeneity, illustrated through the study of the $K18^{+}K14^{+}$ population which I found to poses enhanced tumor and metastasis seeding ability. Subsequently, I completed the functional screening of the candidates proposed initially and uncovered additional pathways that control tumor composition. These include Notch, NFIB and additional genes. I also studied how shifts in differentiation of cancer cells occur, and suggested regulation of asymmetric division rates as an underlying mechanism that could mediate population changes in basal-like breast tumors. Along these studies we have developed and implemented the 'axes of differentiation' approach for observing composition/state, which could contribute to future analyses. Finally, my work revealed the distribution and function of the H2Bub1 post-translational epigenetic modification in breast cancer. I found that H2Bub1 and its regulators are enriched in luminal tumors, where they play a pro-tumorigenic role, while they are downregulated in basal-like tumors where they perform a tumor-suppressive function. These findings thus unveil another aspect of intertumoral heterogeneity and emphasize the need for consideration of cellular context while inspecting the activity of particular pathways. Together, my studies have advanced the

understanding of breast cancer biology and have yielded important practical and conceptual insights that could be utilized to advance the treatment of this disease and additional cancer types.

## 3.1 Lineage differentiation as a measure of heterogeneity in breast tumors

Tumors cells are heterogeneous in various ways, including genetic composition, and epigenetic modifications that influence gene expression and affect their biological function (Marusyk et al., 2012; Meacham and Morrison, 2013). Heterogeneity can be found between different tumors as well as between individual cells that compose them, whose combined phenotypes determine the identity of the bulk tumor. As cancer is predominantly a mutation-driven disease, the effect and prevalence of genetic changes have gained considerable attention and have been extensively studied (Cancer Genome Atlas Network, 2012; Eirew et al., 2015; Nik-Zainal et al., 2016). The latest advancements in genome sequencing have led to the unprecedented characterization of germline and somatic mutations, which were correlated with clinical and biological traits of tumors yielding valuable insights about the prevalence and role of these alterations (Cancer Genome Atlas Network, 2012; Nik-Zainal et al., 2016).

Other studies have explored functional heterogeneity based on the phenotypic and biological features presented by tumor cells. Particular attention has been placed on the assessment of stem cell-like features in tumors, since the traits presented by stem cells, such as self-renewal and differentiation capacity are potentially advantageous in the cancerous setting, giving rise to the CSC model (Al-Hajj et al., 2003; Reya et al., 2001; Visvader and Lindeman, 2008). Moreover, it was noted that tumors with poor clinical outcome, which often display poor histologic differentiation, preferentially activate adult and embryonic SC programs, providing a mechanistic explanation to their poorly differentiated state (Ben-Porath et al., 2008). Several studies have stratified subpopulations of breast cancer cells using a combination of markers such as $CD44^{high}$ $CD24^{low}$, $ALDH1^{+}$ and others, to enrich for populations with CSC traits, and these were found to preferentially activate SC gene signatures and often also activated the EMT program (Al-Hajj et al., 2003; Dontu et al.,

2003; Ginestier et al., 2007). Yet these approaches are not without limitations, as it is difficult to find general markers to identify CSCs across different tumors. Moreover, these methods normally rely on markers that are detached from normal tissue differentiation, and they do not take into account states of intermediate differentiation.

We have taken a somewhat different approach and evaluated heterogeneity by inspecting markers and gene expression signatures derived from normal cell lineages. Since these individual states have been shown to correlate with certain breast cancer subtypes and driven by defined cell-autonomous regulators, we speculated that their study could lead to the discovery of novel insights and mechanisms controlling tumor heterogeneity. We have stained tumor samples of different subtypes against basal, luminal and EMT lineage-specific intermediate filaments. These stains demonstrated that, as previously noted (Livasy et al., 2006; Santagata et al., 2014), basal-like tumors display substantial intratumoral heterogeneity, harboring cell subpopulations expressing different combinations of markers. These include a luminal-like subpopulation, a partially EMT'd subpopulation that co-expresses luminal and mesenchymal markers, and a subpopulation of cells co-expressing a mixture of basal and luminal markers ($K18^+K14^+$). Basal-like tumors, by definition, are the only tumor subtype in which co-expression of basal and luminal markers is observed (Livasy et al., 2006; Rakha et al., 2009; Yu et al., 2013). Our approach has refined previous observations and pointed to the fact that such heterogeneity could be detected at the single cell level and that different phenotypes could be found in proximity. Interestingly, we found that this heterogeneity is preserved, at least in part, in cell lines of basal-like identity, in which these three main subpopulations are maintained in stable equilibrium.

Our work has placed particular attention to cells with dual-lineage marker expression in single cells, focused on $K18^+K14^+$ cells. This pattern of $K18^+K14^+$ expression in single cells is frequently indicative of poor differentiation and plasticity, which often characterizes progenitors and embryonic mammary stem cells as discussed below (Lim et al., 2009; Spike et al., 2012). Such an approach, which employs normal lineage markers, has not been widely used owing to technical challenges it poses, and the limited understanding of hierarchy cell types within the normal breast tissue. A recent study that has used a high-throughput means to screen thousands of normal and cancerous breast sections stained for normal-lineage markers has supported the association of myoepithelial markers with TN/basal-like tumors and highlighted the heterogeneity in lineage markers expression (Santagata et al., 2014). This report also noted the existence of sub-cluster of basal-like tumors and cell lines that

display a double positive (K8/18$^+$ K14/5$^+$) phenotype similar to our findings, yet the study remained observational only.

Besides the observation of individual markers, we have also taken a wider approach to determine the differentiation state or tumor cells using cell type specific gene signatures. These gene signatures are were derived from studies that have isolated and profiled normal cell populations and also illustrated the a functional traits associated with each population (Lim et al., 2009; Spike et al., 2012). Our results are in line with these findings, as we show that subpopulations that are enriched for certain signatures also correspond with the biological behavior that is attributed to the original cell type. During our work we have noticed the need for standards for selecting genes for inclusion is such lists, and for analytic methods to evaluate their enrichment. Therefore, in our analyses we have chosen to use notable and robust gene signatures that have found to faithfully represent their cells of origin. Thus, we believe that these signatures and their co-expression could reliably indicate the differentiation state of individual tumor cell subpopulations and possibly single cells.

## 3.2 Biological properties of mixed lineage cells

Among the mentioned ways to measure heterogeneity, we have given in our studies particular attention to the K18$^+$K14$^+$ mixed luminal/basal lineage phenotype. Mixed-lineage phenotype and specifically K18$^+$K14$^+$ have been described in several studies, frequently associated with a stem or progenitor state. During embryonic development of the murine mammary gland, the fetal MaSCs co-express basal and luminal markers, and these later differentiate into the two lineages (Spike et al., 2012; Van Keymeulen et al., 2011). In the adult tissue, dual-lineage cells are rare, but appear to overlap with a progenitor state (Lim et al., 2009) and are expanded in tissues with aberrations in BRCA1 and p53 (Proia et al., 2011; Zhang et al., 2008). These indications from the healthy tissue hint that the K18$^+$K14$^+$ profile might represent a progenitor-like identity. We were therefore intrigued to study these cells in depth, and profile their expression signature so we could elucidate which pathways are active in them and assess their functional tumorigenic potential.

To gain insights into the transcriptional landscape active in these cells we have obtained the expression signature of $K18^{+}K14^{+}$ cells isolated from cell lines and from patient derived xenografts (PDX) of human tumors in mice (Figure 5). We then assessed the degree of correlation of the expression profiles of this and other subpopulations to published gene signatures. Our analysis revealed that $K18^{+}K14^{+}$ preferentially activate the luminal progenitor expression signature, as well as that of fetal MaSCs (Figure 5). The progenitor signatures contain central TF-encoding genes such as SOX9, VGLL1, FOXQ1 (Granit et al., 2013).

Our analysis also revealed that additional cell populations that co-exist alongside $K18^{+}K14^{+}$ cells, also demonstrated significant correlation with that of certain cell types in the normal breast. The $K18^{+}$ population correlated to mature-luminal cells and $K18^{+}Vim^{+}$ with the MaSCs and an EMT state. The fact that cell populations that are found in close proximity within the tumor, and in culture, can activate such distinct transcriptional programs is itself intriguing, and stress the biologic and clinical complexity presented by intratumoral heterogeneity.

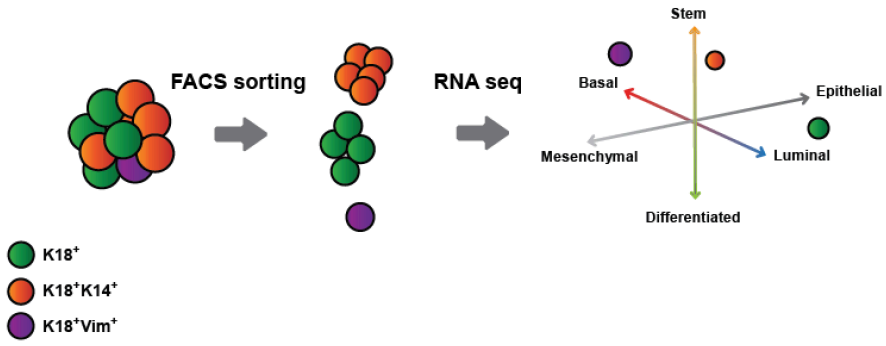



**Figure 5: Inspecting subpopulation gene signature enrichment.** To obtain gene expression signatures of each of the subpopulations, tumor cells were stained for the indicated markers and FACS-sorted, and their mRNA was then profiled. Next, the profile of each subpopulation was correlated with gene signatures representing different cell types in the normal tissue.

Despite the observed correlation at the signature level, the functional traits of the $K18^{+}K14^{+}$ population, and its role at different stages of tumor growth warranted further investigation. Yet, due to the fact that one cannot label kertains in live cells our analysis

relied mainly on retroactive observations done at different time points, and the use of the K14-GFP reporter construct that allowed us to enrich for $K18^+K14^+$ cells. We assessed the seeding and self-renewal capacity of $K18^+K14^+$ cells using a spheroid formation assay and observed the outcome at different time points as the spheroids were formed. This revealed that the $K18^+K14^+$ phenotype is expanded in spheres along their growth, suggesting that this that it might be advantageous in this setting. To further validate this hypothesis, we isolated of $K18^+K14^+$ enriched cells using the K14-GFP reporter, and noted that they demonstrated preferential sphere forming ability. Together, these findings supported an increased functional ability for $K18^+K14^+$ cells under encourage in-depended growth conditions. Several other studies of normal and malignant mammary cells have also yielded similar results (Guo et al., 2012; Pece et al., 2010; Proia et al., 2011), and support the notion that subpopulation expressing mixed basal-luminal markers are found in a progenitor-like state. The fact that $K18^+K14^+$ cells preferentially give rise to spheroid structures, that were found to have a mixture of $K18^+K14^+$ and $K18^+$ only cells, suggested that $K18^+K14^+$ cells might be able to differentiate and give rise to $K18^+$. Yet our limited ability to isolate and follow $K18^+K14^+$ cells in real time has prevented us from providing direct proof for this.

To support our *in vitro* findings, we used the K14-GFP reporter to isolate cells enriched for the $K18^+K14^+$ profile and injected them into mice. This revealed that these cells possess increased tumor initiating capacity *in vivo* relative to the other cell populations. Looking at tumor sections at different time-points we noticed that the $K18^+K14^+$ fraction is increased in parallel to tumor growth and that it is further enriched in sections of metastatic cells originating from these tumors. By injecting cells to the tail-vein of mice and examining their seeding in the lungs, we found that $K18^+K14^+$ fraction is enriched in the process of metastatic colonization, supporting their preferable metastatic seeding capacity. Our findings are in agreement with a previous study linking K14 expression with increased tumorigenic traits such as invasion, collective migration and spheroid formation (Cheung et al., 2013). Together, our results shed light on the functional properties associated with the $K18^+K14^+$ phenotype, relative to other cell subpopulations, and suggest it as promising candidate treatment that is unique to basal-like tumors. It also unveils the tight connection between normal progenitor identity and the $K18^+K14^+$ state which provides a mechanistic explanation their functional abilities. Experiments that ablate $K18^+K14^+$ cells during tumor or metastasis seeding could further substantiate our findings, yet their conduction requires constructing a suitable experimental system. In addition, our findings also expose the unique

complexity found in basal-like tumors versus luminal tumors, at least using our choice of differentiation markers.

There are likely additional variables by which heterogeneity could be examined in basal-like tumors. One example, which has been extensity studied, is the $CD44^{high}$ $CD24^{low}$ profile used to label CSCs and explore the stem <-> differentiated differentiation axis (Mani et al., 2008; Polyak and Weinberg, 2009). Many studies have found that this axis is highly correlated with EMT in breast cancer (Mani et al., 2008; Morel et al., 2008; Polyak and Weinberg, 2009). In the basal-like cell lines we have used we could not detect a distinct $CD44^{high}$ $CD24^{low}$ population, and we believe this phenotype is more prevalent in the more mesenchymal 'Basal B' class of cell lines (Fillmore and Kuperwasser, 2008). Though not the major focus of our work, we used vimentin to explore the mesenchymal <-> epithelial differentiation axis. We believe that this state does not represent full EMT, since we did not observe distinct morphologic differences or activation of 'classic' EMT regulators in these cells. Nevertheless, we find that this population is relatively enriched with the EMT and MaSC signatures. We also find that the $K18^{+}Vim^{+}$ population has very little overlap with the $K18^{+}K14^{+}$ phenotype, suggesting that basal differentiation and EMT could be distinct in certain cancers despite their correlation in the normal tissue (Granit et al., 2013). Further studies could try and functionally compare these two populations and to determine if indeed the $K18^{+}K14^{+}$ phenotype is functionally advantageous over the EMT-like cells marked by vimentin in this context.

As for the basal <-> luminal axis, the literature demonstrates that additional genes, such as CD49f, EpCAM and ALDH1 could be also employed to dissect normal cell lineages in the breast (Lim et al., 2009; Shehata et al., 2012), and possibly also tumor heterogeneity. ALDH1 has been found to label normal luminal progenitors, as well as cells with CSC traits in breast cancer (Ginestier et al., 2007; Ricardo et al., 2011). The expression pattern of CD49f and EpCAM is used to discern between mature luminal, luminal progenitors and MaSCs in the normal breast, yet these markers have not been widely used to identify cell populations in cancer. Preselection of markers for analysis thus dictates the degree of heterogeneity observed, and perhaps more unbiased approaches such as single cell gene expression profiling could reveal in more detail array of populations present in the cancerous tissue.

## 3.3 Regulators of differenitation and hetrogeneity

### 3.3.1 Control of tumor heterogeneity by EZH2

I was driven to explore the regulatory mechanisms that dictate the differentiation state of the malignant cells, and generating heterogeneity. We decided to focus our efforts on the screening of 177 candidate regulators that were identified computationally or known to play a role in dictating various differentiation programs in the breast.

I initially focused on EZH2 and found that upon its silencing basal-like cells show an overall reduced expression of signatures associated with their identity, specifically the signature observed in basal-like tumors and in normal luminal progenitors. I also noted that EZH2 silencing reduces some of the tumorigenic/progenitor traits of tumor cells and these cells showed decreased spheroid and tumor formation capacity. Strikingly, I also found that EZH2 acts as a regulator of tumor composition: in its absence the numbers of the progenitor-like $K18^{+}K14^{+}$ cells were reduced from ~25% of the population to about half of that. I also found that similar results are achieved by pharmacologic inhibition of EZH2; conversely its overexpression increased the numbers of these cells. This indicated that EZH2 may not act simply by affecting the degree of differentiation of all cells, but instead by regulating transitions between distinct identities. Few factors have been previously shown to regulate tumor population composition and the equilibrium between subpopulations. Thus, my study further strengthened the association between EZH2 and the basal-like subtype and its pro-tumorigenic functions, a connection that has also been noted in other studies (Fujii et al., 2011; Gonzalez et al., 2009, 2014). However, my findings provided a novel view of how overall composition may be regulated, and furthermore, how it could be targeted by pharmacologic means.

We have made considerable efforts in choice of a proper cell line systems to study heterogeneity. Unlike many labs, we have made the distinction between ‘basal A’ cell lines that are epithelial and closely resemble actual basal-like tumors, and ‘basal B’ cell lines that are more EMT’d and actually resemble claudin-low tumors (Prat et al., 2010). The proportion of $K18^{+}K14^{+}$ cells ranged between 20-50% in most basal a cells lines, resembling the 1-60% range found in stained primary tumor samples. This, together with their

enrichment for the basal-like gene signature, suggests that these lines are suitable models to study this aspect of the disease. Yet despite this, it appears that in culture, cells appear to lose some of the diversity that is found in primary tumors, such as a $K14^+$ only population that seems to be lacking in most lines.

### 3.3.2 Identification of additional factors that control heterogeneity

The fact that tumor cells maintain the relative proportion of populations within them even after passaging hints that these are regulated by defined pathways that are able to balance them. My findings regarding EZH2 demonstrated the plasticity of the cancerous population and the ability of major regulatory genes to drastically and stably affect the equilibrium between subpopulations. Encouraged by my findings concerning EZH2 I proceeded with the high-throughput screening of the rest of the candidate regulators I defined earlier. These genes were selected due to their potential regulatory function and their association with basal-like tumors, and my hope was to identify which of these indeed govern tumor composition, and to expand our understanding of the regulatory network underlying tumor heterogeneity. The screening process involved silencing each of candidate genes using several shRNA sequences and identification of hits using FACS reading of the population composition with a $K18^+K14^+$ co-stain (Figure 4).

This yielded several regulator genes that alter the population composition upon silencing. Some of these gene are known regulators of normal lineage differentiation such as the luminal regulator FOXA1, the progenitor regulator SOX9, alongside genes that have not been described in this context such as KLF5, FOXQ1, SFRP1, CDK8 and HIF1A. A detailed report of the current knowledge about the function of these genes, and description of my own findings appears below and in the chapters that follow.

We next wished to examine how each of these regulators affect the gene expression program of the breast cancer cells. We profiled cells in which each of the 15 genes was silenced, with two separate shRNA sequences, as well as cells expressing control shRNAs (Figure 6). To accomplish this, we employed a method for uniquely barcoding the mRNA of each sample and next pooling the samples for sequencing – the 'Cel-Seq' method (Hashimshony et al., 2012). I checked changes in the expression of normal lineage signatures

in each of these profiles averaging the profiles from the two hairpins. I found that the regulators clustered into five distinct groups based on their effect on normal lineage gene signatures. One cluster of regulators that act to increase K18+K14+ cell numbers, including EZH2, EED and FOXQ1, caused cells to adopt a more luminal state when silenced. Other TFs that enhance K18+K14+ cell numbers, including SOX9 and KLF5, did not bring out an increase in luminal identity when silenced, but instead an increase in the basal/MaSC signature. One possible explanation is that these factors are important for proper differentiation into a luminal-progenitor state from a less differentiated one such as MaSC identity. Regulators in other clusters had additional distinct effects on the lineage signatures.

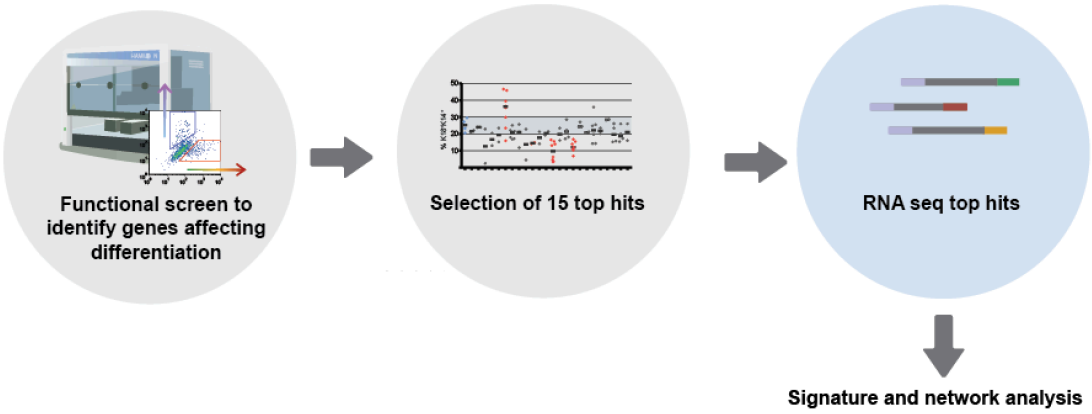



**Figure 6: Strategy for profiling top screen hits.** The functional screen was used to identify genes that regulate the differentiation state, this yielded several hits, of which the top 15 genes were selected and profiled. Using the expression profile of each gene silencing we then performed signature analysis and attempted to create an interaction network.

The screen also pointed to the Notch pathway as a central component in the circuitry responsible for maintaining the progenitor-like identity. Silencing of the downstream effector TF, RBPJ, appeared amongst the top hits reducing the $K18^{+}K14^{+}$ population. Subsequent experiments have validated the importance of Notch as a governor of $K18^{+}K14^{+}$ and progenitor identity as I discuss below. Previous studies have indicated that aberrant Notch pathway activity expands the luminal progenitor population in the normal tissue and leads to tumorigenesis (Bouras et al., 2008). It was also shown that the activation of Notch

in basal-like tumors can occur by ligands expressed in activated stroma surrounding the tumor (Yamamoto et al., 2013). My work thus linked previous studies that have tied Notch to basal/progenitor identity in the normal gland to events within tumors. It also provided a novel functional view of the its regulation of population balance by controlling division choices that have not been described in detail in breast cancer.

Similarly to Notch, NFIB silencing also decreased the $K18^{+}K14^{+}$ cell numbers and reduced progenitor identity. NFIB is a TF found to be upregulated in TN breast tumors, translocated in MYB-NFIB adenoid cystic breast carcinomas (Fusco et al., 2016; Moon et al., 2011) and has been recently shown to promote metastasis in lung cancer and osteosarcoma (Denny et al., 2016; Mirabello et al., 2015; Semenova et al., 2016). The literature also supports a broad regulatory role for NFIB: a recent study explored the function of NFIB in the normal skin epithelium revealed that it acts to maintain the pool of hair follicle stem cells, and that its perturbation leads to the expansion of melanocyte stem cells (Chang et al., 2013). Therefore, NFIB appears as a master regulator that can influence SC population balance in the skin, suggesting it may perform a similar function in the breast epithelia. Yet the exact function of NFIB and its target genes in the normal breast and in cancer are currently poorly characterized and warrant further investigation. I found that NFIB silencing dramatically reduces Notch activity, indicating it acts upstream to it, through a mechanism requiring elucidation.

Other regulators exhibited an opposite effect, and their silencing brought about an increase in the $K18^{+}K14^{+}$ fraction concordant with an increase in the luminal progenitor signature, indicating that they act to repress this state. These included the luminal master regulators GATA3 and FOXA1, that are known to enforce luminal identity in the normal tissue, and are highly active in luminal tumors (Asselin-Labat et al., 2007; Nakshatri and Badve, 2009; Tlsty, 2007). My findings are thus in line with the role of these regulators in the normal breast tissue. Despite our expectation that silencing of FOXA1, which increases the $K18^{+}K14^{+}$ fraction, would enhance the tumorigenic capacity of cells, I have found that this in fact reduces tumor growth (not shown). This suggested that there is either a need for basal expression levels of FOXA1 to allow tumor proliferation, or that excessive $K18^{+}K14^{+}$ numbers actually hinder the ability of the tumor to proliferate. In contrast, I found that GATA3 overexpression causes significant decrease in the capacity of cells to proliferate. My results also suggest that despite the obvious association of GATA3 and FOXA1 with luminal tumors, these factors are also active, at lower levels, in basal-like tumors, and perhaps

constitute a balancing factor in the population. Together, this suggests that within the cancerous tissue, different transcriptional programs compete to enforce their dominance, and perhaps it is selective pressures that determine which ones prevail.

In some cases, our screen output led to somewhat conflicting results. Silencing of genes such as HIF1alpha and SFRP1 demonstrated an opposite correlation between their effect on the $K18^{+}K14^{+}$ fraction and the influence on expression signatures. Yet, these cases were the minority, and in general the expression signatures and the marker-based phenotype were in agreement. Nonetheless, these observations could be explained by a significant role for the perturbed gene in controlling other transcriptional programs, off-target effects or direct regulation of the K14 gene that is disconnected from the luminal progenitor program. These findings highlighted the importance of obtaining the expression profile of these samples, to expand the interpretation of the initial results beyond changes only in K14. Together, these expression profiles have provided valuable information about the function of the different factors and allowed to detect which of these act in a similar manner.

### 3.3.3 Network of interactions between regulatory factors

While all of these factors can be studied individually, the regulatory interactions between them are of importance, as they likely represent a regulatory network. I attempted to construct such a network through evaluating the correlations between the downstream genes that are affected in each experiment, trying to identify regulators that control similar genes. However, these efforts have not yielded robust enough output, and so in my work I have been able to shed light only on some of these interactions that we deduced by more biased study.

Using combinatorial inhibition I established that, consistent with previous studies (Gonzalez et al., 2014), EZH2 acts upstream of the Notch pathway, and identified that NFIB performs a similar function independently of EZH2. Yet I have not established a direct interaction between NFIB and Notch, which could be examined using chromatin IP experiments. Moreover, since co-silencing of EZH2 and NFIB did not completely mimic the direct inhibition of Notch, we suspect there are additional factors that might act to activate Notch in these cells (Figure 7).

Another interaction that was revealed by my work is the repression of GATA3 and FOXA1 by EZH2 (Figure 7). These two TFs, together with the estrogen receptor, comprise the core elements of luminal differentiation and act collaboratively (Theodorou et al., 2013). Repression of lineage-restricted differentiation programs is a well-characterized function of EZH2, which has been noted to bind GATA3 and FOXA1 gene loci in embryonic SC and possibly repress them (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). A recently published study has also demonstrated that FOXA1 is repressed by EZH2 in breast cancer cells in a manner that is dependent on BRCA1 mutation (Gong et al., 2015). Yet, surprisingly, a study that performed overexpression of EZH2 in mouse mammary glands found an increase in the expression levels of GATA3 (Li et al., 2009). Moreover, an additional study that explored the function of EZH2 in the maintenance of the luminal progenitor pool, found that it does so independently of GATA3 (Michalak et al., 2013). These findings suggest that the regulation of GATA3 by EZH2 in breast cancer might be context-dependent and involve additional factors such as BRCA1, similarly to FOXA1. Another possibility is that there is a regulatory difference between mouse-human cells. These finding suggest again that these genes play certain regulatory function in basal-like tumors and possibly prevent them from becoming 'purely basal', possibly competing with the activity of EZH2.

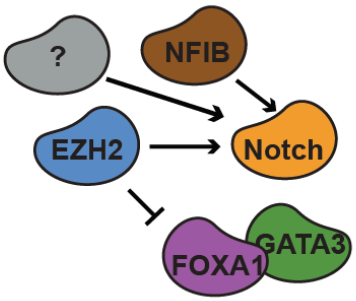



**Figure 7: Gene network of key regulators.** Interaction map demonstrating the regulatory relationship among central basal and luminal regulators.

## 3.4 Cell division switches as a mechanisem for regulation of tumor composition

Following the identification of regulators that could alter tumor composition I wished to conduct a deeper analysis of the mechanisms these genes employ in the process. Unlike the unidirectional differentiation hierarchy thought to occur in normal tissue development, recent studies suggest that tumor cells can display plastic differentiation transitions and may transition into and out of poorly differentiated states (Chaffer et al., 2011, 2013; Gupta et al., 2009). Additionally, it was noted that cultured breast cancer cells can transition between differentiation states, until the population reaches an equilibrium of specific percentages of cell types (Gupta et al., 2011). In our work we also observe that basal-like cell lines, despite growing in culture, maintain specific stable equilibria. Yet under certain conditions, such as during the formation of spheroids *in vitro* or along tumor progression *in vivo*, we did observe enrichment of the $K18^{+}K14^{+}$ cells, suggesting that certain conditions can alter this equilibrium. Despite this, little is known about the mechanistic and regulatory mechanisms that bring about such shifts in differentiation.

Observing differentiation transitions directly as they occur is technically challenging, and only few studies have provided a detailed description of such processes. Our findings suggest a mechanism for differentiation transitions via asymmetric cell divisions. By looking at cell divisions as they occurred we noted asymmetric cell divisions giving rise to one $K18^{+}K14^{+}$ and one $K18^{+}$ daughter cells. The remainder of divisions were symmetric, and gave rise to either two $K18^{+}K14^{+}$ or two $K18^{+}$ cells. If more such asymmetric division occur at the expense of symmetric divisions giving rise to two $K18^{+}K14^{+}$ cells, this would lead to a decrease in the $K18^{+}K14^{+}$ fraction and vice versa (Figure 8).

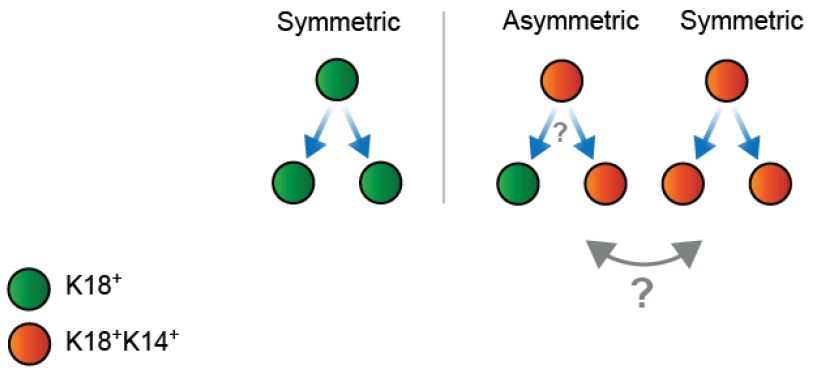



**Figure 8: Division type alters population balance.** We identified three types of replications: symmetric divisions giving rise to two $K18^+$ cells, symmetric divisions giving rise to two $K18^+K14^+$ cells, and asymmetric divisions giving rise to one $K18^+$ and one $K18^+K14^+$ cell. Increased rates of asymmetric divisions can alter the population composition.

Looking at the distribution of vimentin, I could not detect asymmetric divisions, suggesting that a different mechanism regulates changes in mesenchymal differentiation, at least in our system. Previous studies have indicated that during cell division cytokeratin filaments disintegrate and can be distributed between daughter cells before their reassembly (Lane et al., 1982) and these have been shown to differentially divide during cell division in colon cancer (Bu et al., 2013). Other studies conducted in breast cancer cells have uncovered that the activity of p53 and AKT regulates the proportion of asymmetric divisions of CSCs (Cicalese et al., 2009; Dey-Guha et al., 2011). Thus the theoretical, and evidential information supports the possibility of asymmetric division of keratins. We hypothesized that the differentiation shifts observed in my studies, such as those seen upon EZH2 silencing, might be mediated, at least in part, by altering the ratio between symmetric and asymmetric divisions.

We indeed observed that EZH2 and NFIB regulate Notch and act thought it to promote symmetric divisions of $K18^+K14^+$ cells to increase their proportion. In contrast, GATA3 and FOXA1 act to counter this and induce more asymmetric divisions that reduce the proportion of these cells (Figure 9). Notch is a central pathway that is known to maintain cell fate choices, and its activation is known to maintain cells in a non-differentiated state and has been shown to promote their symmetric cell division in cancer (Bu et al., 2013; Chiba, 2006;

Kopan and Ilagan, 2009). Notch therefore provides a mechanistic mediator that could bring about the observed changes in division type, correlating with its role in normal tissue homeostasis and in other cancer types. While other studies have suggested its involvement altering population composition in breast cancer (Dontu et al., 2004; Yamamoto et al., 2013), our study provides a novel direct observation of role of Notch in regulation of division choices in breast cancer. We could not detect a correlation between GATA3 and FOXA1 activity and expression of Notch pathway genes, suggesting that these factors employ other means to influence divisions.

One limitation of these analyses of divisions is that this data is obtained in retrospect, due to the inability to follow keratin segregation in live cells. Thus, the identity of the 'mother' cell giving rise to the distinct divisions is unknown, and I was limited in analyzing the propensity of different cell types to undergo symmetric and asymmetric divisions. Live tracing of cell divisions, alongside a deeper analysis of Notch activity during these processes warrant further investigation.

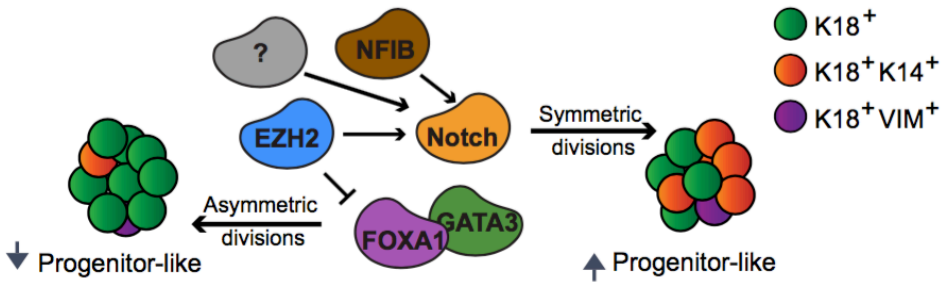



**Figure 9: Proposed model for regulation of basal-like tumor composition.** Integrating our findings, we demonstrate that EZH2, NFIB and potentially additional factors active Notch to promote symmetric cell divisions and increase the $K18^{+}K14^{+}$, progenitor-like, population. EZH2 also inhibits FOXA1 & GATA3 that in turn enforce asymmetric divisions to decrease the $K18^{+}K14^{+}$ fraction.

My findings provide a mechanistic explanation of how differentiation transitions and tumor heterogeneity come about, across the basal <-> luminal differentiation axis, and reveal novel roles for the involved regulators. Yet, we cannot overrule that other mechanisms, such as changes in differentiation and expression of keratins that occur in non-dividing cells,

might also take part in bringing about these changes. My findings also raise several intriguing questions that will require further investigation, such as involvement of additional regulators in this process, the interaction between luminal factors and the Notch pathway, and the roles of these factors in divisions in the normal breast. Also, the involvement of other known regulators of division choices such as components of the Notch pathway require further dissection and may shed light on the mechanistic processes that govern differentiation shifts.

## 3.5 Opposing effects of H2Bub1 in breast cancer subtypes

Chromatin regulators are often dysregulated in cancer, yet much about their mechanistic function, influence on phenotype and their ability to dictate distinct identities in the malignant tissue remains to be studied (Jones and Baylin, 2007; Sauvageau and Sauvageau, 2010). I found that EZH2, a modifier of H3K27 methylation controls the expression of key regulators, affects differentiation programs and promotes symmetric divisions and activity of Notch pathway. Another modification that was implicated in the context of cancer is monoubiquitylation of histone H2B, yet much less is known about it, and this raised the question of whether it and its regulators contribute to cancer (Cao and Yan, 2012; Fuchs and Oren, 2014). This modification is believed to play important regulatory roles, amongst these transcription elongation and DNA repair, which could affect cancer (Cao and Yan, 2012; Fuchs and Oren, 2014; Fuchs et al., 2014).

When I began to investigate the function of H2Bub1 in breast cancer the literature held several studies demonstrating contradicting findings of the role of this modification. One study that explored the function of the H2B ubiquitin ligase RNF20 in normal-immortalized cells found that it performs a tumor suppressive function and demonstrated that the RNF20 promoter is silenced by methylation in breast cancer (Shema et al., 2008). Another study claimed that in breast cancer the RNF20 repressor SMURF2 is itself tumor-suppressor and that RNF20 expression promotes tumorigenesis and DNA damage (Blank et al., 2012). Yet another study suggested that RNF20 is required for MLL-rearranged leukemia by promoting the expression of oncogenic target genes in collaboration with the H3K79 methyltransferase DOT1L (Wang et al., 2013a). We therefore collaborated with the lab of Prof. Moshe Oren,

that has been extensively studying this modification, to try and resolve this issue. Our efforts were driven by the hypothesis that differential regulation and effects of H2Bub1 in different tumor types might be the reason for the conflicting observations.

We found that regulators of the H2Bub1 modification, such as RNF20 and RNF40 are highly expressed in luminal tumors, but expressed in lower levels in tumors of the Basal-like subtype. In contrast, negative regulators of H2Bub1, such as Smurf2 and USP44, display the opposite pattern. Next, we examined the association of H2bub1 with patient outcome and noticed that in patients with luminal tumors high H2Bub1 levels were associated with poor survival; in contrast, over-activation of this pathway was associated with better survival in patients carrying basal-like tumors. These findings suggested differential activation and activity amongst different tumor types, and led us to investigate the underlining mechanisms that could allow the same pathway to bring about different outcomes in tumors of the same tissue of origin.

Using cell line models, we demonstrated that in luminal cells H2Bub1 and its positive regulators are drivers of tumor growth, while they perform the opposite function in basal-like cells. We discovered that the basis for these differences could be found, at least in part, in the pathways that characterize and drive the two different breast cancer subtypes. We found that H2Bub1 is needed for expression of estrogen receptor targets which drive luminal tumor growth. This is consistent with previous and recent works showing that H2Bub1 is important for the expression of ER target genes (Oh et al., 2006; Perou et al., 2000) and that its supports tumorigenesis in $ER^+$/Luminal tumors (Duan et al., 2016; Prenzel et al., 2011).

In contrast, basal-like tumors do not express the estrogen receptor and rely on different pathways for their proliferation; and so we explored the influence of H2Bub1 on such genes. Studies performed in the intestine have suggested that H2Bub1 leads to repression NF-κB activity and its downstream pro-inflammatory cytokines (Tarcic et al., 2016), which have been noted to drive tumorigenic features in basal-like tumors (Barbie et al., 2014; Kim et al., 2015). We found that presence of H2Bub1 represses IL6 and IL8 expression in breast cancer of all subtypes, yet since the expression of these genes is only substantial in basal-like tumors, this function imposed anti-tumorigenic behavior only in these malignancies. Hence we established that H2Bub1 promotes oncogenesis in luminal tumors by promoting expression of ER targets, while in basal-like tumors its repression of IL6 and IL8 yields tumor suppression. In our studies we used the manipulation of its regulator RNF20 to probe the

significance of this modification; we therefore could not exclude that RNF20 itself might be involved in tumorigenic processes outside the context of H2Bub1. However, we obtained opposite results when manipulating USP44, an inhibitor of H2Bub1, providing further support that it is indeed H2Bub1 that mediates the observed effects.

Together, our findings shed new light on the prevalence and role of H2Bub1 in breast cancer and provide an explanation for the seeming contradiction between previously published results. Moreover, they emphasize the critical importance of considering intertumoral heterogeneity while examining genes and phenomena that relate to cancer and not to treat it as a single disease. It also highlights another epigenetic modification, that, similarly to H3K27me3, present heterogeneous regulation, distribution and functional outcome in breast cancer. Our study points to several paths that could be pursued, including the dissection of the role of H2Bub1 in the normal versus cancerous breast tissue and its distribution within individual tumors. Furthermore, it would be interesting to explore the involvement of this modification if normal breast development.

## 3.6 Pharmacological regulation of tumor composition

Driven by findings described above I was determined to explore means that could modify tumor composition and could be translated into actual cancer treatments. Unlike genetic mutations, transcriptional and chromatin modulators could be potentially influenced thanks to recent advancements in the development of selective pharmacological agents, which have emerged as promising treatments for cancer (Hoelder et al., 2012). Some inhibitors, such as Imatinib and Nilotinib, show increased selectivity against specific cancer altered or fusion proteins (Manley et al., 2010; O'Brien et al., 2003). Other inhibitors target key receptor oncogenes such as EGFR or transcription factors such as in the case of inhibition of the Myc-Max interaction (Hoelder et al., 2012; Huang et al., 2006). Interestingly, it was recently demonstrated that Monastrol, an inhibitor of Eg5, a kinesin protein that is involved in mitotic spindle formation, and one of the factors found downstream to the H2Bub1 pathway, can specifically target luminal tumors suggesting a new targeted treatment for this disease (Duan et al., 2016).

In my experiments I used pharmacological inhibitors, which target EZH2 or the γ-secretase enzyme which is necessary for Notch activation. I found that inhibition of both pathways leads to a dose-depended decrease in the $K18^+K14^+$ cell fraction concurrent with reduction in progenitor-like identity of cancer cells. Additionally, I have noted that Notch inhibition reduced the symmetric $K18^+K14^+$ replications to an even greater extent than EZH2 inhibition. Notch inhibition could be accomplished using GSI molecules or monoclonal antibodies that target the receptors or ligands of this pathway (Yuan et al., 2015). GSI agents are currently being used for the treatment of Alzheimer's disease, and the growing evidence of Notch involvement in cancer have led to attempts to employ these drugs in these contexts (Barten et al., 2006; Imbimbo, 2008). Notch inhibitors are currently found in phase I/II clinical trials, assessing their treatment potential in several cancer types, including breast, pancreas and colorectal malignancies (Imbimbo, 2008; Yuan et al., 2015). The emergence of EZH2 as an oncogene overexpressed in several cancers and mutated in lymphomas has led to the recent development of small molecule inhibitors against it, yet these are only in preclinical or early state clinical trials (Kim and Roberts, 2016; McCabe et al., 2012). While inhibitors of EZH2 have not been extensively studied in breast cancers, a recent study found that FOXA1 is up-regulated upon EZH2 inhibition using GSK-126 in BRCA1 mutated breast cancers providing support to some of my findings (Gong et al., 2015). Notch inhibition has been used in many breast cancer studies, including TN models, and led to a decrease in tumorigenic traits, yet its effect on tumor composition has not been characterized (D'Angelo et al., 2015; Robinson et al., 2011; Stoeck et al., 2014).

Together, my findings suggest a possible pharmacological 'differentiation-treatment' that can be used to shift basal-like tumor composition away from its initial progenitor-like state, perhaps towards a more druggable, less aggressive state. This concept is supported by our finding that EZH2 silencing let to a decrease in the $K18^+K14^+$ fraction, and decreased tumor growth. Despite this, our efforts to achieve these effects on growing tumor xenografts by treating mice with these chemical inhibitors have not been successful due to technical difficulties in the delivery of these drugs in a prolonged manner. Studies have suggested that poorly differentiated cells might be more resistant to chemotherapy (Meacham and Morrison, 2013), and so we hypothesized that $K18^+K14^+$ cells might demonstrate similar resistance. In fact, preliminary experiments I conducted suggested the opposite – that these cells are more sensitive to drug treatment, and perhaps due to the fact that these cells proliferate faster than the $K18^+$ cells (not shown). Further understanding of tumor composition could yield new

combinatorial treatments and synthetic lethality approaches to treat cancer. It remains to be found if our findings could be translated to *in vivo* studies and eventually into the clinic.

## 3.7 Conclusion

Together, my findings have unveiled important new insights into the regulation of breast cancer subtype identity, and revealed novel regulatory mechanisms that govern intratumoral heterogeneity, emphasizing the role of epigenetic regulators. My work also revealed new insights about the functional significance of intratumoral heterogeneity, and specifically the $K18^{+}K14^{+}$ population, suggesting that differentiation related lineage-markers could be used to identify the degree of heterogeneity and tumorigenic potential. Furthermore, we revealed a mechanism that allows shifts in tumor composition to take place via regulation of cell division choices. We suggest that a core network governed by EZH2 and NFIB maintains the $K18^{+}K14^{+}$, progenitor-like state, by activating Notch and suppressing luminal transcriptional regulators. Importantly we show that this network could be perturbed using pharmacological treatment, paving the way to novel treatments that could tilt tumor composition towards a more benign state. These findings lay the ground to diverse future studies that could expand the understanding of regulatory networks by inspecting additional regulators. Moreover, we have chosen to focus on a cell-autonomous view, and the effect of external factors such as the microenvironment, cytokines and inflammation on tumor composition still holds great unexplored potential.

Also, together with our collaborators, we obtained new insights into the prevalence of H2Bub1 and its regulators and revealed fundamental regulatory differences underlying distinct subtypes and their clinical behavior. We discovered novel insights about the prominence of H2Bub1 in breast cancer, and find that it plays opposed roles in different tumor subtypes.

My studies emphasize the importance of inspecting tumors as complex entities that require in-depth dissection and analysis in both basic cancer research and the clinic. Increased adoption of the newly emerging technologies for molecular analysis of single tumor cells is likely to drastically improve the understanding of tumor heterogeneity. Ideally,

various areas within the tumor and different time points could be sampled and modeled to predict the optimal treatment options for each individual tumor.

# 4. References


Alford, S.H., Toy, K., Merajver, S.D., and Kleer, C.G. (2012). Increased risk for distant metastasis in patients with familial early-stage breast cancer and high EZH2 expression. Breast Cancer Res. Treat. *132*, 429–437.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. U. S. A. *100*, 3983–3988.

Allen, E., Miéville, P., Warren, C.M., Saghafinia, S., Li, L., Peng, M.-W., and Hanahan, D. (2016). Metabolic Symbiosis Enables Adaptive Resistance to Anti-angiogenic Therapy that Is Dependent on mTOR Signaling. Cell Rep. *15*, 1144–1160.

Almendro, V., Cheng, Y.-K., Randles, A., Itzkovitz, S., Marusyk, A., Ametller, E., Gonzalez-Farre, X., Muñoz, M., Russnes, H.G., Helland, A., et al. (2014). Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. Cell Rep. *6*, 514–527.

Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: cell fate control and signal integration in development. Science *284*, 770–776.

Asangani, I.A., Ateeq, B., Cao, Q., Dodson, L., Pandhi, M., Kunju, L.P., Mehra, R., Lonigro, R.J., Siddiqui, J., Palanisamy, N., et al. (2013). Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. Mol. Cell *49*, 80–93.

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

Banerjee, S., Reis-Filho, J.S., Ashley, S., Steele, D., Ashworth, A., Lakhani, S.R., and Smith, I.E. (2006). Basal-like breast carcinomas: clinical outcome and response to chemotherapy. J. Clin. Pathol. *59*, 729–735.

Barbie, T.U., Alexe, G., Aref, A.R., Li, S., Zhu, Z., Zhang, X., Imamura, Y., Thai, T.C., Huang, Y., Bowden, M., et al. (2014). Targeting an IKBKE cytokine network impairs triple-negative breast cancer growth. J. Clin. Invest. *124*, 5411–5423.

Barten, D.M., Meredith, J.E., Zaczek, R., Houston, J.G., and Albright, C.F. (2006). Gamma-secretase inhibitors for Alzheimer's disease: balancing efficacy and toxicity. Drugs RD *7*, 87–97.

Bedi, U., Scheel, A.H., Hennion, M., Begus-Nahrmann, Y., Rüschoff, J., and Johnsen, S.A. (2015). SUPT6H controls estrogen receptor activity and cellular differentiation by multiple epigenomic mechanisms. Oncogene *34*, 465–473.

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat. Genet. *40*, 499–507.

Blank, M., Tang, Y., Yamashita, M., Burkett, S.S., Cheng, S.Y., and Zhang, Y.E. (2012). A tumor suppressor function of Smurf2 associated with controlling chromatin landscape and genome stability through RNF20. Nat. Med. *18*, 227–234.

Blanpain, C. (2013). Tracing the cellular origin of cancer. Nat. Cell Biol. *15*, 126–134.

Blows, F.M., Driver, K.E., Schmidt, M.K., Broeks, A., van Leeuwen, F.E., Wesseling, J., Cheang, M.C., Gelmon, K., Nielsen, T.O., Blomqvist, C., et al. (2010). Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med. *7*, e1000279.

Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.-L., Oakes, S.R., Lindeman, G.J., and Visvader, J.E. (2008). Notch Signaling Regulates Mammary Stem Cell Function and Luminal Cell-Fate Commitment. Cell Stem Cell *3*, 429–441.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., and Helin, K. (2003). EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J. *22*, 5323–5335.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Brooks, M.D., Burness, M.L., and Wicha, M.S. (2015). Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer. Cell Stem Cell *17*, 260–271.

Bu, P., Chen, K.-Y., Chen, J.H., Wang, L., Walters, J., Shin, Y.J., Goerger, J.P., Sun, J., Witherspoon, M., Rakhilin, N., et al. (2013). A microRNA miR-34a Regulated Bimodal Switch targets Notch in Colon Cancer Stem Cells. Cell Stem Cell *12*, 602–615.

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

Cao, J., and Yan, Q. (2012). Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. Front. Oncol. *2*, 26.

Chaffer, C.L., and Weinberg, R.A. (2011). A perspective on cancer cell metastasis. Science *331*, 1559–1564.

Chaffer, C.L., Brueckmann, I., Scheel, C., Kaestli, A.J., Wiggins, P.A., Rodrigues, L.O., Brooks, M., Reinhardt, F., Su, Y., Polyak, K., et al. (2011). Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc. Natl. Acad. Sci. U. S. A. *108*, 7950–7955.

Chaffer, C.L., Marjanovic, N.D., Lee, T., Bell, G., Kleer, C.G., Reinhardt, F., D'Alessio, A.C., Young, R.A., and Weinberg, R.A. (2013). Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell *154*, 61–74.

Chang, C.-Y., Pasolli, H.A., Giannopoulou, E.G., Guasch, G., Gronostajski, R.M., Elemento, O., and Fuchs, E. (2013). NFIB is a governor of epithelial-melanocyte stem cell behaviour in a shared niche. Nature *495*, 98–102.

Chase, A., and Cross, N.C.P. (2011). Aberrations of EZH2 in Cancer. Clin. Cancer Res. *17*, 2613–2618.

Chen, S., Li, J., Wang, D.-L., and Sun, F.-L. (2012). Histone H2B lysine 120 monoubiquitination is required for embryonic stem cell differentiation. Cell Res. *22*, 1402–1405.

Cheung, K.J., Gabrielson, E., Werb, Z., and Ewald, A.J. (2013). Collective Invasion in Breast Cancer Requires a Conserved Basal Epithelial Program. Cell *155*, 1639–1651.

Chiba, S. (2006). Notch signaling in stem cell systems. Stem Cells Dayt. Ohio *24*, 2437–2447.

Cicalese, A., Bonizzi, G., Pasi, C.E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Fiore, P.P.D., and Pelicci, P.G. (2009). The Tumor Suppressor p53 Regulates Polarity of Self-Renewing Divisions in Mammary Stem Cells. Cell *138*, 1083–1095.

Colaluca, I.N., Tosoni, D., Nuciforo, P., Senic-Matuglia, F., Galimberti, V., Viale, G., Pece, S., and Di Fiore, P.P. (2008). NUMB controls p53 tumour suppressor activity. Nature *451*, 76–80.

Collett, K. (2006). Expression of Enhancer of Zeste Homologue 2 Is Significantly Associated with Increased Tumor Cell Proliferation and Is a Marker of Aggressive Breast Cancer. Clin. Cancer Res. *12*, 1168–1174.

Condiotti, R., Guo, W., and Ben-Porath, I. (2014). Evolving views of breast cancer stem cells and their differentiation States. Crit. Rev. Oncog. *19*, 337–348.

D'Angelo, R.C., Ouzounova, M., Davis, A., Choi, D., Tchuenkam, S.M., Kim, G., Luther, T., Quraishi, A.A., Senbabaoglu, Y., Conley, S.J., et al. (2015). Notch Reporter Activity in Breast Cancer Cell Lines Identifies a Subset of Cells with Stem Cell Activity. Mol. Cancer Ther. *14*, 779–787.

Denny, S.K., Yang, D., Chuang, C.-H., Brady, J.J., Lim, J.S., Grüner, B.M., Chiou, S.-H., Schep, A.N., Baral, J., Hamard, C., et al. (2016). Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. Cell *166*, 328–342.

Dey-Guha, I., Wolfer, A., Yeh, A.C., G Albeck, J., Darp, R., Leon, E., Wulfkuhle, J., Petricoin, E.F., Wittner, B.S., and Ramaswamy, S. (2011). Asymmetric cancer cell division regulated by AKT. Proc. Natl. Acad. Sci. U. S. A. *108*, 12845–12850.

Ding, L., and Kleer, C.G. (2006). Enhancer of Zeste 2 as a Marker of Preneoplastic Progression in the Breast. Cancer Res. *66*, 9352–9355.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Dontu, G., Jackson, K.W., McNicholas, E., Kawamura, M.J., Abdallah, W.M., and Wicha, M.S. (2004). Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. Breast Cancer Res. *6*, R605–R615.

Duan, Y., Huo, D., Gao, J., Wu, H., Ye, Z., Liu, Z., Zhang, K., Shan, L., Zhou, X., Wang, Y., et al. (2016). Ubiquitin ligase RNF20/40 facilitates spindle assembly and promotes breast carcinogenesis through stabilizing motor protein Eg5. Nat. Commun. *7*, 12648.

Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A., et al. (2015). Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. Nature *518*, 422–426.

Farnie, G., and Clarke, R.B. (2007). Mammary stem cells and breast cancer--role of Notch signalling. Stem Cell Rev. *3*, 169–175.

Feuer, E.J., Wun, L.-M., Boring, C.C., Flanders, W.D., Timmel, M.J., and Tong, T. (1993). The Lifetime Risk of Developing Breast Cancer. J. Natl. Cancer Inst. *85*, 892–897.

Fillmore, C.M., and Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res. BCR *10*, R25.

Fuchs, G., and Oren, M. (2014). Writing and reading H2B monoubiquitylation. Biochim. Biophys. Acta *1839*, 694–701.

Fuchs, G., Hollander, D., Voichek, Y., Ast, G., and Oren, M. (2014). Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. Genome Res. *24*, 1572–1583.

Fujii, S., Tokita, K., Wada, N., Ito, K., Yamauchi, C., Ito, Y., and Ochiai, A. (2011). MEK-ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes. Oncogene *30*, 4118–4128.

Fusco, N., Geyer, F.C., De Filippo, M.R., Martelotto, L.G., Ng, C.K.Y., Piscuoglio, S., Guerini-Rocco, E., Schultheis, A.M., Fuhrmann, L., Wang, L., et al. (2016). Genetic events in the progression of adenoid cystic carcinoma of the breast to high-grade triple-negative breast cancer. Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555–567.

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

Goldhirsch, A., Wood, W.C., Coates, A.S., Gelber, R.D., Thürlimann, B., Senn, H.-J., and Panel members (2011). Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. *22*, 1736–1747.

Gong, C., Fujino, K., Monteiro, L.J., Gomes, A.R., Drost, R., Davidson-Smith, H., Takeda, S., Khoo, U.S., Jonkers, J., Sproul, D., et al. (2015). FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer. Oncogene *34*, 5012–5024.

Gonzalez, M.E., Li, X., Toy, K., DuPrie, M., Ventura, A.C., Banerjee, M., Ljungman, M., Merajver, S.D., and Kleer, C.G. (2009). Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. Oncogene *28*, 843–853.

Gonzalez, M.E., Moore, H.M., Li, X., Toy, K.A., Huang, W., Sabel, M.S., Kidwell, K.M., and Kleer, C.G. (2014). EZH2 expands breast stem cells through activation of NOTCH1 signaling. Proc. Natl. Acad. Sci. U. S. A. *111*, 3098–3103.

Granit, R.Z., Slyper, M., and Ben-Porath, I. (2013). Axes of differentiation in breast cancer: untangling stemness, lineage identity, and the epithelial to mesenchymal transition. Wiley Interdiscip. Rev. Syst. Biol. Med. 93–106.

Gudjonsson, T., Adriance, M.C., Sternlicht, M.D., Petersen, O.W., and Bissell, M.J. (2005). Myoepithelial Cells: Their Origin and Function in Breast Morphogenesis and Neoplasia. J. Mammary Gland Biol. Neoplasia *10*, 261–272.

Gunasinghe, N.P.A.D., Wells, A., Thompson, E.W., and Hugo, H.J. (2012). Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. Cancer Metastasis Rev. *31*, 469–478.

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

Gupta, P.B., Chaffer, C.L., and Weinberg, R.A. (2009). Cancer stem cells: mirage or reality? Nat. Med. *15*, 1010–1012.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Harrison, H., Farnie, G., Howell, S.J., Rock, R.E., Stylianou, S., Brennan, K.R., Bundred, N.J., and Clarke, R.B. (2010). Regulation of Breast Cancer Stem Cell Activity by Signaling through the Notch4 Receptor. Cancer Res. *70*, 709–718.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Rep. *2*, 666–673.

Heldin, C.-H., Landström, M., and Moustakas, A. (2009). Mechanism of TGF-beta signaling to growth arrest, apoptosis, and epithelial-mesenchymal transition. Curr. Opin. Cell Biol. *21*, 166–176.

Hoelder, S., Clarke, P.A., and Workman, P. (2012). Discovery of small molecule cancer drugs: successes, challenges and opportunities. Mol. Oncol. *6*, 155–176.

Hu, C., Diévart, A., Lupien, M., Calvo, E., Tremblay, G., and Jolicoeur, P. (2006). Overexpression of activated murine Notch1 and Notch3 in transgenic mice blocks mammary gland development and induces mammary tumors. Am. J. Pathol. *168*, 973–990.

Huang, M.-J., Cheng, Y., Liu, C.-R., Lin, S., and Liu, H.E. (2006). A small-molecule c-Myc inhibitor, 10058-F4, induces cell-cycle arrest, apoptosis, and myeloid differentiation of human acute myeloid leukemia. Exp. Hematol. *34*, 1480–1489.

Imbimbo, B.P. (2008). Therapeutic potential of gamma-secretase inhibitors and modulators. Curr. Top. Med. Chem. *8*, 54–61.

Jones, P.A., and Baylin, S.B. (2007). The Epigenomics of Cancer. Cell *128*, 683–692.

Kim, K.H., and Roberts, C.W.M. (2016). Targeting EZH2 in cancer. Nat. Med. *22*, 128–134.

Kim, G., Ouzounova, M., Quraishi, A.A., Davis, A., Tawakkol, N., Clouthier, S.G., Malik, F., Paulson, A.K., D'Angelo, R.C., Korkaya, S., et al. (2015). SOCS3-mediated regulation of inflammatory cytokines in PTEN and p53 inactivated triple negative breast cancer model. Oncogene *34*, 671–680.

Kleer, C.G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S.A., Ghosh, D., Sewalt, R.G.A.B., Otte, A.P., Hayes, D.F., et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc. Natl. Acad. Sci. U. S. A. *100*, 11606–11611.

Kopan, R., and Ilagan, M.X.G. (2009). The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. Cell *137*, 216–233.

Lackner, M.R., Wilson, T.R., and Settleman, J. (2012). Mechanisms of acquired resistance to targeted cancer therapies. Future Oncol. Lond. Engl. *8*, 999–1014.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. Nat. Rev. Mol. Cell Biol. *15*, 178–196.

Lane, E.B., Goodman, S.L., and Trejdosiewicz, L.K. (1982). Disruption of the keratin filament network during epithelial cell division. EMBO J. *1*, 1365–1372.

Lee, C.W., Simin, K., Liu, Q., Plescia, J., Guha, M., Khan, A., Hsieh, C.-C., and Altieri, D.C. (2008). A functional Notch–survivin gene signature in basal breast cancer. Breast Cancer Res. *10*, R97.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Li, X., Lewis, M.T., Huang, J., Gutierrez, C., Osborne, C.K., Wu, M.-F., Hilsenbeck, S.G., Pavlick, A., Zhang, X., Chamness, G.C., et al. (2008). Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J. Natl. Cancer Inst. *100*, 672–679.

Li, X., Gonzalez, M.E., Toy, K., Filzen, T., Merajver, S.D., and Kleer, C.G. (2009). Targeted Overexpression of EZH2 in the Mammary Gland Disrupts Ductal Morphogenesis and Causes Epithelial Hyperplasia. Am. J. Pathol. *175*, 1246–1254.

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

Liu, X., Holstege, H., van der Gulden, H., Treur-Mulder, M., Zevenhoven, J., Velds, A., Kerkhoven, R.M., van Vliet, M.H., Wessels, L.F.A., Peterse, J.L., et al. (2007). Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. Proc. Natl. Acad. Sci. U. S. A. *104*, 12111–12116.

Livasy, C.A., Karaca, G., Nanda, R., Tretiakova, M.S., Olopade, O.I., Moore, D.T., and Perou, C.M. (2006). Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc *19*, 264–271.

Lobry, C., Oh, P., and Aifantis, I. (2011). Oncogenic and tumor suppressor functions of Notch in cancer: it's NOTCH what you think. J. Exp. Med. *208*, 1931–1935.

Magee, J.A., Piskounova, E., and Morrison, S.J. (2012). Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. Cancer Cell *21*, 283–296.

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Manley, P.W., Stiefl, N., Cowan-Jacob, S.W., Kaufman, S., Mestan, J., Wartmann, M., Wiesmann, M., Woodman, R., and Gallagher, N. (2010). Structural resemblances and comparisons of the relative pharmacological properties of imatinib and nilotinib. Bioorg. Med. Chem. *18*, 6977–6986.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Marjanovic, N.D., Weinberg, R.A., and Chaffer, C.L. (2013). Cell plasticity and heterogeneity in cancer. Clin. Chem. *59*, 168–179.

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

McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Aller, G.S.V., Liu, Y., Graves, A.P., Iii, A.D.P., Diaz, E., et al. (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature *492*, 108–112.

McGill, M.A., and McGlade, C.J. (2003). Mammalian numb proteins promote Notch1 receptor ubiquitination and degradation of the Notch1 intracellular domain. J. Biol. Chem. *278*, 23196–23203.

Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328–337.

Melling, N., Grimm, N., Simon, R., Stahl, P., Bokemeyer, C., Terracciano, L., Sauter, G., Izbicki, J.R., and Marx, A.H. (2016). Loss of H2Bub1 Expression is Linked to Poor Prognosis in Nodal Negative Colorectal Cancers. Pathol. Oncol. Res. POR *22*, 95–102.

Meyer, D.S., Brinkhaus, H., Müller, U., Müller, M., Cardiff, R.D., and Bentires-Alj, M. (2011). Luminal expression of PIK3CA mutant H1047R in the mammary gland induces heterogeneous tumors. Cancer Res. *71*, 4344–4351.

Michalak, E.M., Nacerddine, K., Pietersen, A., Beuger, V., Pawlitzky, I., Cornelissen-Steijger, P., Wientjens, E., Tanger, E., Seibler, J., van Lohuizen, M., et al. (2013). Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal progenitor pool. Stem Cells Dayt. Ohio *31*, 1910–1920.

Mirabello, L., Koster, R., Moriarity, B.S., Spector, L.G., Meltzer, P.S., Gary, J., Machiela, M.J., Pankratz, N., Panagiotou, O.A., Largaespada, D., et al. (2015). A Genome-Wide Scan Identifies Variants in NFIB Associated with Metastasis in Patients with Osteosarcoma. Cancer Discov. *5*, 920–931.

Molyneux, G., Geyer, F.C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell *7*, 403–417.

Moon, H.-G., Hwang, K.-T., Kim, J.-A., Kim, H.S., Lee, M.-J., Jung, E.-M., Ko, E., Han, W., and Noh, D.-Y. (2011). NFIB is a potential target for estrogen receptor-negative breast cancers. Mol. Oncol. *5*, 538–544.

Moore, N., and Lyle, S. (2011). Quiescent, slow-cycling stem cell populations in cancer: a review of the evidence and discussion of significance. J. Oncol. *2011*.

Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. PloS One *3*, e2888.

Mumm, J.S., and Kopan, R. (2000). Notch Signaling: From the Outside In. Dev. Biol. *228*, 151–165.

Nakshatri, H., and Badve, S. (2009). FOXA1 in breast cancer. Expert Rev. Mol. Med. *11*, e8.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

O'Brien, S.G., Guilhot, F., Larson, R.A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J.J., Fischer, T., Hochhaus, A., Hughes, T., et al. (2003). Imatinib Compared with Interferon and Low-Dose Cytarabine for Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia. N. Engl. J. Med. *348*, 994–1004.

Oh, D.S., Troester, M.A., Usary, J., Hu, Z., He, X., Fan, C., Wu, J., Carey, L.A., and Perou, C.M. (2006). Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *24*, 1656–1664.

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

Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P.G., and Di Fiore, P.P. (2010). Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell *140*, 62–73.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Pietersen, A.M., Horlings, H.M., Hauptmann, M., Langerød, A., Ajouaou, A., Cornelissen-Steijger, P., Wessels, L.F., Jonkers, J., van de Vijver, M.J., and van Lohuizen, M. (2008). EZH2 and BMI1 inversely correlate with prognosis and TP53 mutation in breast cancer. Breast Cancer Res. BCR *10*, R109.

Polyak, K., and Metzger Filho, O. (2012). SnapShot: breast cancer. Cancer Cell *22*, 562–562.e1.

Polyak, K., and Weinberg, R.A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat. Rev. Cancer *9*, 265–273.

Pradeep, C.-R., Köstler, W.J., Lauriola, M., Granit, R.Z., Zhang, F., Jacob-Hirsch, J., Rechavi, G., Nair, H.B., Hennessy, B.T., Gonzalez-Angulo, A.M., et al. (2012). Modeling ductal carcinoma in situ: a HER2–Notch3 collaboration enables luminal filling. Oncogene *31*, 907–917.

Prat, A., and Perou, C.M. (2009). Mammary development meets cancer genomics. Nat. Med. *15*, 842–844.

Prat, A., and Perou, C.M. (2011). Deconstructing the molecular portraits of breast cancer. Mol. Oncol. *5*, 5–23.

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

Prat, A., Ellis, M.J., and Perou, C.M. (2012). Practical implications of gene-expression-based assays for breast oncologists. Nat. Rev. Clin. Oncol. *9*, 48–57.

Prat, A., Pineda, E., Adamo, B., Galván, P., Fernández, A., Gaba, L., Díez, M., Viladot, M., Arance, A., and Muñoz, M. (2015). Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast Edinb. Scotl. *24 Suppl 2*, S26-35.

Prater, M.D., Petit, V., Russell, I.A., Giraddi, R., Shehata, M., Menon, S., Schulte, R., Kalajzic, I., Rath, N., Olson, M.F., et al. (2014). Mammary stem cells have myoepithelial cell properties. Nat. Cell Biol. *16*, 942–947.

Prenzel, T., Begus-Nahrmann, Y., Kramer, F., Hennion, M., Hsu, C., Gorsler, T., Hintermair, C., Eick, D., Kremmer, E., Simons, M., et al. (2011). Estrogen-dependent gene transcription in human breast cancer cells relies upon proteasome-dependent monoubiquitination of histone H2B. Cancer Res. *71*, 5739–5753.

Proia, T.A., Keller, P.J., Gupta, P.B., Klebba, I., Jones, A.D., Sedic, M., Gilmore, H., Tung, N., Naber, S.P., Schnitt, S., et al. (2011). Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. Cell Stem Cell *8*, 149–163.

Puppe, J., Drost, R., Liu, X., Joosse, S.A., Evers, B., Cornelissen-Steijger, P., Nederlof, P., Yu, Q., Jonkers, J., van Lohuizen, M., et al. (2009). BRCA1-deficient mammary tumor

cells are dependent on EZH2 expression and sensitive to Polycomb Repressive Complex 2-inhibitor 3-deazaneplanocin A. Breast Cancer Res. *11*, R63.

Raafat, A., Bargo, S., Anver, M.R., and Callahan, R. (2004). Mammary development and tumorigenesis in mice expressing a truncated human Notch4/Int3 intracellular domain (h-Int3sh). Oncogene *23*, 9401–9407.

Rakha, E.A., El-Sayed, M.E., Reis-Filho, J., and Ellis, I.O. (2009). Patho-biological aspects of basal-like breast cancer. Breast Cancer Res. Treat. *113*, 411–422.

Ranganathan, P., Weaver, K.L., and Capobianco, A.J. (2011). Notch signalling in solid tumours: a little bit of everything but not all the time. Nat. Rev. Cancer *11*, 338–351.

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J., et al. (2008). Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process. Cell Stem Cell *3*, 109–118.

Reed, A.E.M., Kutasovic, J.R., Lakhani, S.R., and Simpson, P.T. (2015). Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. Breast Cancer Res. BCR *17*.

Reedijk, M., Odorcic, S., Chang, L., Zhang, H., Miller, N., McCready, D.R., Lockwood, G., and Egan, S.E. (2005). High-level coexpression of JAG1 and NOTCH1 is observed in human breast cancer and is associated with poor overall survival. Cancer Res. *65*, 8530–8537.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.

Ricardo, S., Vieira, A.F., Gerhard, R., Leitão, D., Pinto, R., Cameselle-Teijeiro, J.F., Milanezi, F., Schmitt, F., and Paredes, J. (2011). Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. J. Clin. Pathol. *64*, 937–946.

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

Robinson, D.R., Kalyana-Sundaram, S., Wu, Y.-M., Shankar, S., Cao, X., Ateeq, B., Asangani, I.A., Iyer, M., Maher, C.A., Grasso, C.S., et al. (2011). Functionally recurrent

rearrangements of the MAST kinase and Notch gene families in breast cancer. Nat. Med. *17*, 1646–1651.

Sander, S., Bullinger, L., Klapproth, K., Fiedler, K., Kestler, H.A., Barth, T.F.E., Möller, P., Stilgenbauer, S., Pollack, J.R., and Wirth, T. (2008). MYC stimulates EZH2 expression by repression of its negative regulator miR-26a. Blood *112*, 4202–4212.

Santagata, S., Thakkar, A., Ergonul, A., Wang, B., Woo, T., Hu, R., Harrell, J.C., McNamara, G., Schwede, M., Culhane, A.C., et al. (2014). Taxonomy of breast cancer based on normal cell phenotype predicts outcome. J. Clin. Invest. *124*, 859–870.

Sauvageau, M., and Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell *7*, 299–313.

Semenova, E.A., Kwon, M.-C., Monkhorst, K., Song, J.-Y., Bhaskaran, R., Krijgsman, O., Kuilman, T., Peters, D., Buikhuisen, W.A., Smit, E.F., et al. (2016). Transcription Factor NFIB Is a Driver of Small Cell Lung Cancer Progression in Mice and Marks Metastatic Disease in Patients. Cell Rep. *16*, 631–643.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.-L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

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

Shema, E., Tirosh, I., Aylon, Y., Huang, J., Ye, C., Moskovits, N., Raver-Shapira, N., Minsky, N., Pirngruber, J., Tarcic, G., et al. (2008). The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. Genes Dev. *22*, 2664–2676.

Shih, A.H., Abdel-Wahab, O., Patel, J.P., and Levine, R.L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. Nat. Rev. Cancer *12*, 599–612.

Skibinski, A., and Kuperwasser, C. (2015). The origin of breast tumor heterogeneity. Oncogene *34*, 5309–5316.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat. Rev. Cancer *6*, 846–856.

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

Stingl, J., and Caldas, C. (2007). Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. Nat. Rev. Cancer *7*, 791–799.

Stoeck, A., Lejnine, S., Truong, A., Pan, L., Wang, H., Zang, C., Yuan, J., Ware, C., MacLean, J., Garrett-Engele, P.W., et al. (2014). Discovery of Biomarkers Predictive of GSI Response in Triple-Negative Breast Cancer and Adenoid Cystic Carcinoma. Cancer Discov. *4*, 1154–1167.

Tao, L., van Bragt, M.P.A., Laudadio, E., and Li, Z. (2014). Lineage Tracing of Mammary Epithelial Cells Using Cell-Type-Specific Cre-Expressing Adenoviruses. Stem Cell Rep. *2*, 770–779.

Tarcic, O., Pateras, I.S., Cooks, T., Shema, E., Kanterman, J., Ashkenazi, H., Boocholez, H., Hubert, A., Rotkopf, R., Baniyash, M., et al. (2016). RNF20 Links Histone H2B Ubiquitylation with Inflammation and Inflammation-Associated Cancer. Cell Rep. *14*, 1462–1476.

Theodorou, V., Stark, R., Menon, S., and Carroll, J.S. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. Genome Res. *23*, 12–22.

Thiery, J.P., Acloque, H., Huang, R.Y.J., and Nieto, M.A. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. Cell *139*, 871–890.

Tlsty, T.D. (2007). Luminal cells GATA have it. Nat. Cell Biol. *9*, 135–136.

Tsai, J.H., and Yang, J. (2013). Epithelial-mesenchymal plasticity in carcinoma metastasis. Genes Dev. *27*, 2192–2206.

Tsai, J.H., Donaher, J.L., Murphy, D.A., Chau, S., and Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. Cancer Cell *22*, 725–736.

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

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G.A.B., Otte, A.P., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624–629.

Vethantham, V., Yang, Y., Bowman, C., Asp, P., Lee, J.-H., Skalnik, D.G., and Dynlacht, B.D. (2012). Dynamic loss of H2B ubiquitylation without corresponding changes in H3K4 trimethylation during myogenic differentiation. Mol. Cell. Biol. *32*, 1044–1055.

Visvader, J.E. (2009). Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev. *23*, 2563–2577.

Visvader, J.E., and Lindeman, G.J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat. Rev. Cancer *8*, 755–768.

Wagenblast, E., Soto, M., Gutiérrez-Ángel, S., Hartl, C.A., Gable, A.L., Maceli, A.R., Erard, N., Williams, A.M., Kim, S.Y., Dickopf, S., et al. (2015). A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. Nature *520*, 358–362.

Wang, E., Kawaoka, S., Yu, M., Shi, J., Ni, T., Yang, W., Zhu, J., Roeder, R.G., and Vakoc, C.R. (2013a). Histone H2B ubiquitin ligase RNF20 is required for MLL-rearranged leukemia. Proc. Natl. Acad. Sci. U. S. A. *110*, 3901–3906.

Wang, L., Zeng, X., Chen, S., Ding, L., Zhong, J., Zhao, J.C., Wang, L., Sarver, A., Koller, A., Zhi, J., et al. (2013b). BRCA1 is a negative modulator of the PRC2 complex. EMBO J. *32*, 1584–1597.

Wang, Z.-J., Yang, J.-L., Wang, Y.-P., Lou, J.-Y., Chen, J., Liu, C., and Guo, L.-D. (2013c). Decreased histone H2B monoubiquitination in malignant gastric carcinoma. World J. Gastroenterol. *19*, 8099–8107.

Watson, C.J., and Khaled, W.T. (2008). Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development *135*, 995–1003.

Wood, K.C. (2016). Collaborating tumor cells overcome multitargeted antiangiogenic therapies. Sci. Transl. Med. *8*, 338ec77-338ec77.

Xu, K., Wu, Z.J., Groner, A.C., He, H.H., Cai, C., Lis, R.T., Wu, X., Stack, E.C., Loda, M., Liu, T., et al. (2012a). EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science *338*, 1465–1469.

Xu, K., Usary, J., Kousis, P.C., Prat, A., Wang, D.-Y., Adams, J.R., Wang, W., Loch, A.J., Deng, T., Zhao, W., et al. (2012b). Lunatic fringe deficiency cooperates with the Met/Caveolin gene amplicon to induce basal-like breast cancer. Cancer Cell *21*, 626–641.

Yamaguchi, H., and Hung, M.-C. (2014). Regulation and Role of EZH2 in Cancer. Cancer Res. Treat. Off. J. Korean Cancer Assoc. *46*, 209–222.

Yamamoto, M., Taguchi, Y., Ito-Kureha, T., Semba, K., Yamaguchi, N., and Inoue, J. (2013). NF-κB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. Nat. Commun. *4*, 2299.

Ye, X., Tam, W.L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E., and Weinberg, R.A. (2015). Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. Nature *525*, 256–260.

Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M., et al. (2013). Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. Science *339*, 580–584.

Yuan, X., Wu, H., Xu, H., Xiong, H., Chu, Q., Yu, S., Wu, G.S., and Wu, K. (2015). Notch signaling: An emerging therapeutic target for cancer treatment. Cancer Lett. *369*, 20–27.

Zhang, M., Behbod, F., Atkinson, R.L., Landis, M.D., Kittrell, F., Edwards, D., Medina, D., Tsimelzon, A., Hilsenbeck, S., Green, J.E., et al. (2008). Identification of Tumor-Initiating Cells in a p53-Null Mouse Model of Breast Cancer. Cancer Res. *68*, 4674–4682.