Mechanistic Insight and Translational Possibilities of Obesity-Breast Cancer Connection

Xiaozheng Liu

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

Mechanistic Insight and Translational Possibilities of Obesity-Breast Cancer Connection

Xiaozheng Liu

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 07.12.2021

Year: 2021

Title: Mechanistic Insight and Translational Possibilities of Obesity-Breast Cancer Connection

Name: Xiaozheng Liu

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

The work in this thesis was performed at the Department of Biomedicine, University of Bergen, during 2017-2021 under the supervision of Assoc. Prof. Nils Halberg, Prof. Stian Knappskog, and Prof. James B. Lorens. Part of the work was performed in collaboration with Prof. Philipp Scherer at the Department of Internal Medicine and Cell Biology, University of Texas Southwestern Medical Center, USA.

Acknowledgements

Undertaking this PhD has been the best decision I have ever made. My sincere gratitude goes to all the people who made my PhD study an excellent experience.

I want to thank my supervisor, Nils Halberg, for giving me the opportunity to make the work possible. He is an incredibly intelligent and insightful scientist. His deep passion for science has extremely inspired me since my master project, and eventually led me to pursue this PhD. I appreciate all the discussions with him, which have sharped my thinking and brought my work to a higher level.

A huge thank to my co-supervisor, Professor Stian Knappskog, who always kindly shares invaluable data and brilliant ideas. His feedbacks remind me to see my work from a clinical perspective. Thank you for being so generous with your time to work on my last-minute requests.

I have been lucky enough to get to work with a number of wonderful colleagues during these years. Line, an inspiring colleague and close friend, witnessed my entire PhD journey. We laughed together, cried together, and celebrated together every achievement with our "Friday beer". I am truly grateful to your continuous encouragement. Nastia, you are a fearless scientist with numerous bold ideas. I sincerely wish you all the best in your next chapter. Johanna and Sina, it has been a great pleasure working with you. Thank you for your significant contributions to our paper. I would also like to thank the colleagues who joined the team later, Kelly, Adam, Marcus, Alex. Although we did not get chance to work together in the lab, your everyday's greetings, smiles and talks made the tough writing period smoother.

I am grateful for all the support I get from CellNet group. Stacey and Sturla, thank you for spending long time figuring out the best way to present our data. Karl Johan, Ina and Ning, thank you for sharing your invaluable experiences and patiently answering all my questions. Endre, thank you for supporting us with your creative measures which always make our work efficient.

Dandan, my dearest friend and family, I deeply thank you for your infinite support. You teach me programming, you share with me your daily discoveries, you bridge me to a wider world. This PhD has not been an easy journey. Thank you for always being there for me.

Finally, I express my deepest appreciation to my parents for always encouraging me to believe in myself and follow my dreams. I could not imagine myself going through that far without your love.

Xiaozheng Liu

September 2021

Abstract

Obesity is an independent risk factor for more than thirteen cancer types. In breast cancer context, obesity is strongly associated with higher incidence and poorer prognosis in postmenopausal patients. The management of breast cancer in obese patients is faced with numerous challenges, as dosing, toxicity and drug efficacy. Thus, it is highly demanded to investigate the in-depth mechanisms underlying the obesity-cancer connection and translate these fundamental mechanisms into clinical applications.

To this end, we first searched for breast cancer phenotypes driven by the obese environment. In both patients and preclinical models, we identified enhanced stem-like traits in cancer cells exposed to the obese settings and it independently predicted breast cancer risks. Next, as obesity-induced elevated free fatty acid is a central phenotype in obese individuals, we established *in vitro* cancer cell models by long-term exposure to palmitic acid. Under this condition, cancer cells consistently dedifferentiated towards a cancer stem-like phenotype and displayed enhanced tumor initiation capacity. Mechanistically, we found that the obesity-induced phenotype was governed epigenetically through remodeling the chromatin landscape in cancer cells, specifically through increased chromatin occupancy of the transcription factor CCAAT/enhancer-binding protein beta (C/EBPB). We further identified Lipocalin 2 and Claudin 1 as the key downstream target genes of C/EBPB and functionally demonstrated the critical roles of C/EBPB and its targets in obesity-induced breast cancer initiation phenotypes.

The second part of this thesis focuses on the investigation of endotrophin-induced physiological responses in human cells and the clinical translational potential. The cleavage product of collagen VI alpha 3 chain, endotrophin, is upregulated in obese states and has been previously reported as a stimulator for oncogenic transformation of mammary ductal epithelial cells in rodents. In addition, endotrophin presents pro-fibrotic, chemoattractant, pro-angiogenic and epithelial to mesenchymal transition induction properties in murine cells. Here, we demonstrated that the circulating levels of endotrophin is increased in breast cancer patients compared with healthy individuals.

Recombinant human endotrophin exerted similar effects on human cell lines as in murine cells, suggesting that endotrophin is a viable target for breast cancer therapy. As such, we developed neutralizing antibodies targeting human endotrophin, and ultimately validated the effectiveness of such antibody in human breast cancer cell lines and human cell line-derived nude mouse models. In both *in vitro* and *in vivo* models, the antibody showed potent tumor growth inhibition and anti-cisplatin resistance functions, and thus provided the important stage towards developing targeted therapies for obese breast cancer patients.

List of publications

Paper I

C/EBPB-dependent Adaptation to Palmitic Acid Promotes Tumor Formation in Hormone Receptor Negative Breast Cancer

Xiao-Zheng Liu*, Anastasiia Rulina*, Man Hung Choi*, Line Pedersen, Johanna Lepland, Sina T Takle, Noelly Madeleine, Stacey D'mello Peters, Cara Ellen Wogsland, Sturla Magnus Grøndal, James B Lorens, Hani Goodarzi, Per Eystein Lønning, Stian Knappskog, Anders Molven, Nils Halberg. *Equal contribution. (Submitted manuscript)

Paper II

Human endotrophin as a driver of malignant tumor growth

Dawei Bu, Clair Crewe, Christine M. Kusminski, Ruth Gordillo, Alexandra L. Ghaben, Min Kim, Jiyoung Park, Hui Deng, Wei Xiong, **Xiao-Zheng Liu,** Per Eystein Lønning, Nils Halberg, Adan Rios, Yujun Chang, Anneliese Gonzalez, Ningyan Zhang, Zhiqiang An, Philipp E. Scherer. *JCI Insight*, 2019. **5.**

Additional work not included in the thesis:

Cellular mechanisms linking cancers to obesity

Xiao-Zheng Liu*, Line Pedersen*, Nils Halberg. *Equal contribution. *Cell Stress*, 2021. **5**(5): p. 55-72.

Upregulated PDK4 expression is a sensitive marker of increased fatty acid oxidation

Ina Katrine Nitschke Pettersen, Deusdedit Tusubira, Hanan Ashrafi, Sissel Elisabeth Dyrstad, Lena Hansen, **Xiao-Zheng Liu**, Linn Iren Hodneland Nilsson, Nils Gunnar Løvsletten, Kjetil Berge, Hege Wergedahl, Bodil Bjørndal, Øystein Fluge, Ove Bruland, Arild Christian Rustan, Nils Halberg, Gro Vatne Røsland, Rolf Kristian Berge, Karl Johan Tronstad. *Mitochondrion*, 2019. **49**: p. 97-110.

Table of Contents

Abbreviations

2HG	2-hydroxyglutarate
Acetyl-CoA	Acetyl coenzyme A
AF2	Activation function 2
AMPK	AMP-activated protein kinase
BMI	Body mass index
BRCA	Breast cancer gene
BRM	Brahma
C/EBPB	CCAAT/enhancer binding protein beta
CDK	Cyclin-dependent kinase
CHD	Chromodomain helicase DNA binding
CLDN1	Claudin 1
COL6	Collagen VI
CSCs	Cancer stem cells
CSK	C-terminal SRC kinase
CTLA-4	Cytotoxic T lymphocyte-associated antigen 4
DAG	Diacylglycerol
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
ER	Estrogen receptor
ERK	Extracellular signal-related kinase
ESCs	Embryonic stem cells
ETP	Endotrophin
FABP4	Fatty acid binding protein 4
FFA	Free fatty acid
FOXA1	Forkhead box A1
GATA	GATA binding protein
GC-MS	Gas chromatography–mass spectrometry
HAT	Histone acetyltransferase

HDAC	Histone deacetylase
HDM	Histone demethylase
HER2	Human epidermal growth factor 2
HFD	High fat diet
HMT	Histone methyltransferase
ICB	Immune checkpoint blockade
IHC	Immunohistochemistry
IL-1β	Interleukin 1β
INO80	Inositol
ISWI	Imitation switch
JHDM	JmjC-domain containing histone
KLF4	Kruppel-like factor 4
LAP	Liver activating protein
LCN2	Lipocalin 2
LIP	Liver inhibitory protein
MAPK	Mitogen-activated protein kinase
MaSC	Mammary stem cell
MAT	Metastasis-associated gene
MBD2	Methyl-CpG-binding domain 2
MDR1	Multidrug resistance protein 1
MMe	Metabolically activated
MMTV	Mouse mammary tumor virus
MYC	Myelocytomatosis viral oncogene homolog
NG2	Neuron-glial antigen 2
NuRD	Nucleosome remodeling and histone deacetylase
OCT4	Octamer-binding transcription factor 4
PAK2	p21 protein-activated kinase 2
PAM50	Prediction analysis of microarray 50
PD-1	Programmed cell death receptor 1
PD-L1	Programmed cell death ligand 1
PlGF	Placental growth factor

PR	Progesterone receptor
PTEN	Phosphatase and tensin homolog
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SOX2	Sex determining region Y box 2
SWI/SNF	Switching/Sucrose non-ferme
TNBC	Triple-negative breast cancer
TP53	Tumor protein p53
uORF	upstream open reading frame
VEGFA	vascular endothelial growth factor A
WHO	World Health Organization
αKG	α-ketoglutarate

1. Introduction

1.1 Obesity

Obesity is defined as an abnormal and excessive accumulation of body fat[1]. In 2016, more than 50% of adults were reported as overweight or obesity, accounting for nearly 2 billion people worldwide. The number among children and adolescents was over 340 million[1]. Obesity has profound negative impacts on both physical and psychological health. It is the leading risk factor for the increased morbidity and mortality of many diseases, including cardiovascular diseases, diabetes and cancer[2]. With the steadily increasing prevalence, overweight and obesity has become a severe public health burden worldwide.

Obesity is the result of energy disequilibrium. As caloric intake exceeds energy expenditure, the excess energy is deposited in the adipocytes of adipose tissue. If this imbalance is maintained over time, it will lead to an increase in body mass[3]. Body mass index (BMI) represents the most commonly used parameters to classify overweight and obese individuals. It is calculated as weight (in kilograms) divided by the square of height (in meters)[4]. According to world health organization (WHO) classification, individuals with a BMI between 25-29.9 kg/m^2 are defined as overweight and individuals with BMI above 30 kg/m^2 are considered obese[1]. In some circumstances, BMI may not be a perfect measure. Given the differences between fat and nonfat mass (e.g., muscle and bone), BMI could overestimate body fat in the population with higher muscle percent, for example athletes, and underestimate it in some elderly persons who have low bone density. Thus, in addition to BMI, the measures of abdominal obesity, such as waist circumference and waist-hip ratio, as well as direct body fat measures have been recommended by clinical practice guidelines[5, 6].

Extensive epidemiological studies have indicated that obesity is associated with an increased risk of numerous health complications, such as type 2 diabetes, cardiovascular diseases, inflammation, fibrosis and cancers (Figure 1)[2, 5, 7, 8]. Most obesity-related

comorbidities are attributed to the metabolic alterations, perhaps most notably, lipid metabolism. After long periods of excess energy intake, adipose tissue reaches its storage limit and the additional energy in the form of lipids spill into the circulation where it will be transported to lipid-intolerant organs as heart, liver and pancreas for ectopic deposition. The accumulated lipids are converted into toxic derivatives, such as ceramides, acylcarnitines and diacylglycerols (DAGs)[9]. In liver, the lipid-induced DAG accumulation activates protein kinase C which inhibits insulin receptor tyrosine kinase activity, and consequently result in hepatic insulin resistance[10]. Similarly, in skeletal muscles, the lipid-induced ceramides accumulation results in insulin resistance by inhibiting both insulin receptor signaling and insulin-stimulated glucose uptake[11].

In addition to lipotoxicity, long-term exposure to hyperglycemia may cause β-cell dysfunction and result in a decreased insulin production, which is termed glucotoxicity[12]. Furthermore, the dysfunctional adipocytes abnormally produce circulating factors such as leptin, adiponectin and cholesterol (Figure 1), which cause both systemic and local impacts. Collectively, all the obesity-related systemic and local alterations could contribute to cancer initiation and progression. The relevant mechanisms will be discussed in Section 1.3.2.

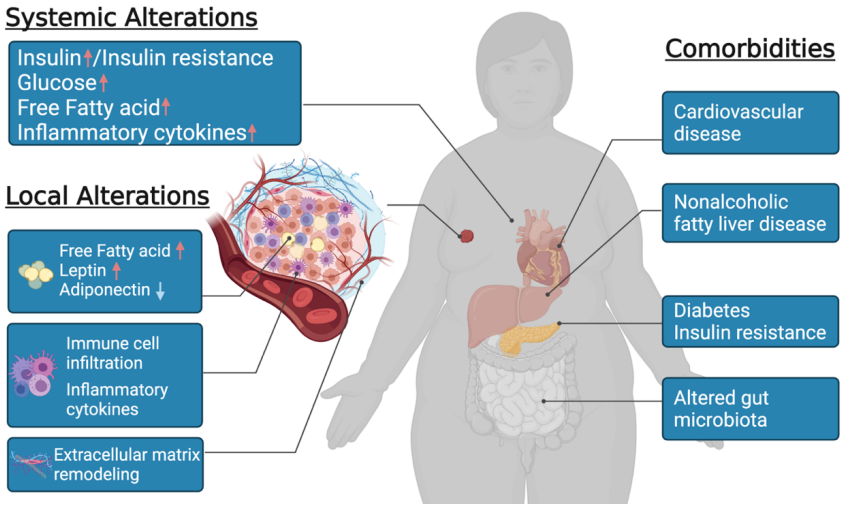


Figure 1. An overview of obesity-related comorbidities and associated systemic and local alterations.

1.2 Breast Cancer

Breast cancer is the most commonly diagnosed cancer and a leading cause of cancer mortality in women. In 2020, an estimated 2.3 million new breast cancer cases were diagnosed and 685 000 deaths reported worldwide[13]. The incidence rate of breast cancer is higher in high-income countries (55.9 per 100 000) compared with low-income countries (29.7 per 100 000)[13]. However, the mortality rates are lower in high-income countries due to the inadequacy of early diagnosis and access to treatment in low-income countries[14].

1.2.1 Breast cancer classification

Breast cancer classification is important to characterize the cancer properties at diagnosis and follow the clinical behaviors during cancer progression. Moreover, it is central for treatment decisions.

Breast cancer classification has been based primarily on histological features and growth patterns of tumor. Preinvasive (in situ) carcinoma and invasive carcinoma[15] are the two main classifications of breast cancer. According to WHO classification, invasive breast carcinoma can be further divided into more than 10 subtypes[16]. Among these subtypes, infiltrating ductal carcinoma and invasive lobular carcinoma are the most frequently diagnosed lesions, which account for more than 70% of all invasive lesions[15, 17].

With the development of molecular analytical techniques, several systematic and comprehensive molecular classifications have been developed, which have become more and more important in term of dictating breast cancer treatment strategies. In 2000, a study by Perou *et al*.[18] suggested a breast cancer molecular classification according to the expression patterns of an intrinsic gene subset (including 496 genes; their expressions exhibit greater difference between different tumor samples than between samples from same tumor). In this study, breast cancer was clustered into four subtypes, which was further developed to five intrinsic subtypes: luminal A, luminal B, human epidermal growth factor receptor 2 (HER2) positive, basal-like, and normal breast-like,

due to the distinctions between luminal A and luminal B subgroups[19, 20] (Figure 2). Prediction analysis of microarray 50 (PAM50) is another gene expression-based molecular classification[21]. Breast cancer subtypes can be identified by quantitative measurement of a 50-genes signature. The performance of PAM50 classifications has been evaluated across multiple cohorts[22], and has significantly added prognostic value to current histological and intrinsic classifications.

Immunohistochemistry (IHC)-based surrogate classification brings molecular classification into clinical practice. A four markers surrogate panel, including estrogen receptors (ER), progesterone receptors (PR), HER2 and proliferation marker Ki-67[23], is widely accepted and classify breast cancers into five subtypes (Figure 2). This classification generally overlaps with molecular classification and plays major roles in treatment decisions[24]. Overall, due to the heterogeneity of breast cancer, the combination of different classification methods can better characterize the status of breast cancer, which form the basis to guide future treatment strategies.

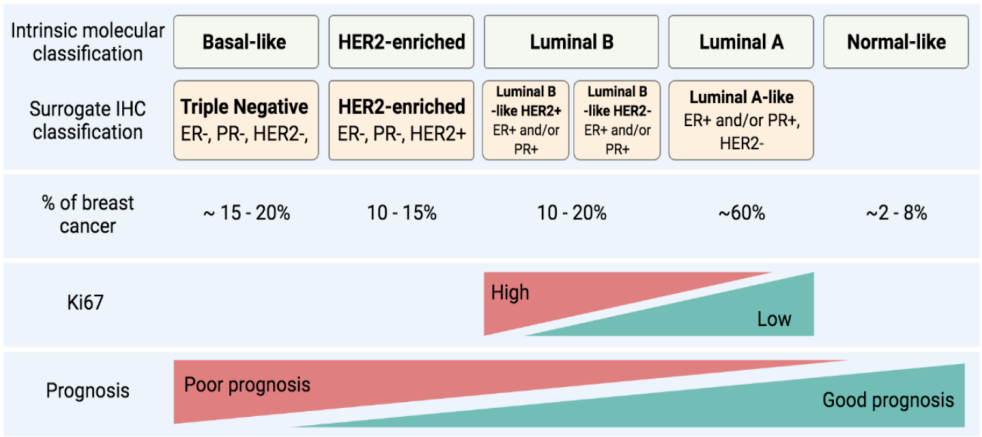


Figure 2. Breast cancer classifications with proliferation and prognosis information. Five breast cancer subgroups have been identified by intrinsic molecular signature, which are highly overlapped with surrogate IHC classifications. For instance, 80% of triple-negative breast cancer (TNBC) cases are basal-like. Luminal A breast cancer is the most common subtype. Basal-like/TNBC has the worst prognosis and accounts for around 15 – 20% of all breast cancer cases. Figure made based on [24-26].

1.2.2 Mammary gland development

Human mammary gland is a complex secretory organ with the function of synthesizing and secreting milk for nourishment of newborn[27]. Extensive lineage tracing studies have identified the existence of distinct mammary stem cells (MaSCs) and progenitor cell populations. They drive mammary gland development and also contribute to the morphological and functional changes in adult mammary gland, especially during pregnancy-lactation-involution cycles[28].

Human mammary gland development follows a hierarchical differentiation structure (Figure 3A). MaSCs reside at the top of the hierarchy and give rise to basal and luminal progenitor cells, which further subdivided into mature basal myoepithelial, ductal and alveolar cells[29]. However, it is still debated if both bipotent and unipotent stem/progenitor cells exist in the postnatal mammary gland[30].

During the early stage of mammary gland development, a wide range of factors and signaling pathways are involved. For instance, Wnt signaling plays key roles in mammary placode formation. This signaling is initiated when Wnt proteins bind to Frizzled/low-density lipoprotein receptor-related protein (LRP) receptor complex, and the downstream signaling of Wnt/Frizzled/LRP complex protects β-catenin from degradation. Thus, the accumulated β-catenin interacts with TCF/LEF transcription factor to activate the transcriptions of Wnt targets[31]. Wnt/β-catenin signaling is specifically activated at placode-forming regions and the inhibition of Wnt is able to completely block mammary placode formation[32].

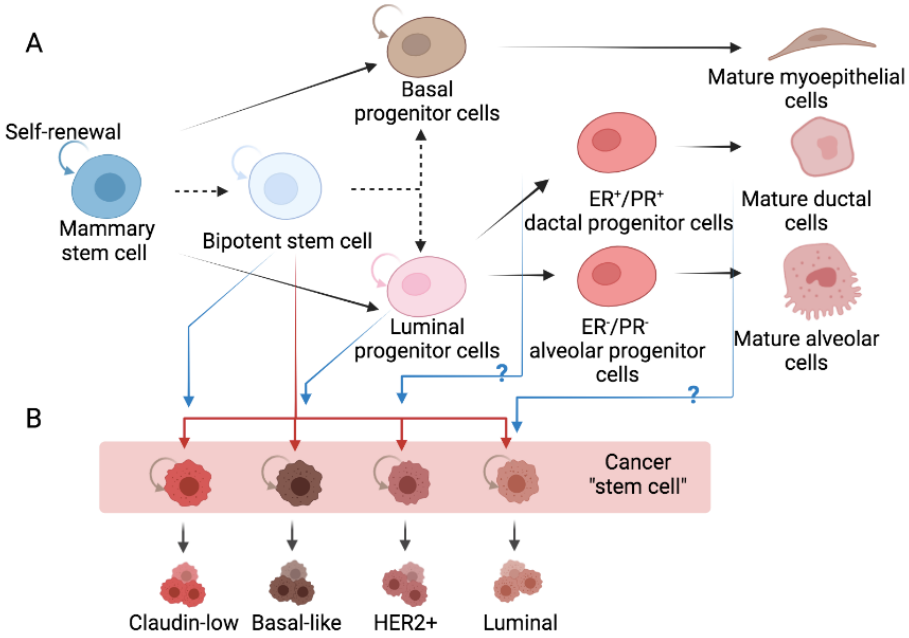


Figure 3. Schematic model of mammary epithelial differentiation hierarchy and potential relationships with breast cancer subtypes. (A) Multipotent stem cells have self-renewal ability and give rise to the lineage-restricted progenitor cells. Basal and luminal progenitor cells further differentiate to mature myoepithelial cells, ductal cells and alveolar cells respectively. The existence of bipotent MaSCs has also been reported and it can develop to unipotent luminal or basal stem/progenitor cells (the lineages of bipotent stem cells are marked with dotted line) which follow either luminal or myoepithelial lineage. (B) Two models (marked in blue and red lines) explain the development of different breast cancer subtypes. In the first model (red), mammary stem cell is the origin of breast cancer. It develops to different subtypes of breast cancer by acquiring various genetic or epigenetic changes. The other model (blue) suggests that each breast cancer subtype is from a different stage of mammary stem cells or progenitors. It is still debated where is the origin of $HER2^+$ and luminal subtypes (question marks). Figure made based on [28, 29, 33].

MaSCs and distinct progenitors also exist in adult mammary gland, and together with mature epithelial cells and non-epithelial components, such as fibroblasts, adipocytes and immune cells maintain adult mammary gland functions (Figure 4)[34].

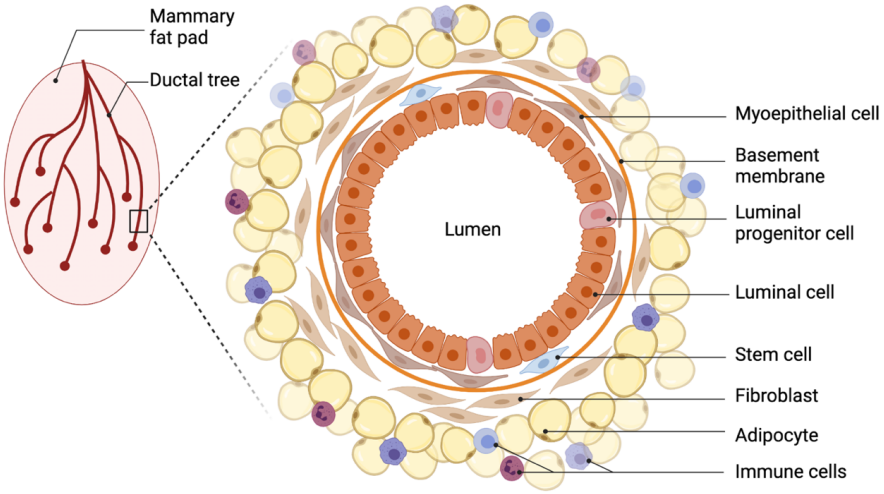


Figure 4. Schematic representation of a duct cross section and its environment. Figure made based on [33].

In adult MaSCs, epigenetic programs play a fundamental role to control the balance between self-renewal and differentiation[28]. For instance, it has been identified that cells isolated from normal breast tissue of healthy women present distinct DNA methylation patterns. These patterns closely correlate with cellular differentiation status, and the more differentiated cells display higher degree of DNA methylation[35]. It has been consistently observed that human MaSCs harbor more hypomethylated chromatin elements compared with differentiated luminal cells, which is important to keep high expression of several transcription factors to maintain stem cell states[36]. Specifically, transcription factor 3 (TCF3), homeobox protein Hox-A10 (HOXA10) and forkhead factor C1 (FOXC1) are hypomethylated and highly expressed in mammary progenitor ($CD44^{+}$) cell populations and all of them have been reported to play key roles in the maintenance of stem cell homeostasis and self-renewal capacity[35]. Epigenetic landscape significantly changes during pregnancy. For example, H3K27me3

modifications in luminal cells are globally increased during pregnancy. As a repressive modification marker, the increased H3K27me3 modifications correlate with repression of luminal genes, which restrict the luminal lineage and facilitate the expansion of alveolar compartments[37]. Interestingly, the decreased H3K27me3 modification only occurs on specific luminal genes, including *Elf-5*, *Wap*, and *Csn2,* which causes their upregulations. These genes are required for milk production[37].

1.2.3 Origin of breast cancer

MaSCs have been considered as leading candidates for the origin of breast cancer due to their self-renewal and differentiation capability. Through acquiring specific genetic and epigenetic aberrations, MaSCs are able to transform to heterogeneous breast cancer stem cells and further differentiate to different subtypes[38]. In addition, emerging evidence have demonstrated that diverse progenitor cells and even lineage committed cells are able to dedifferentiate and acquire cancer stem-like properties (Figure 3B). Here, the different cancer subtypes are derived from specific populations of stem/progenitor cells. For example, the gene expression profile of the claudin-low breast cancer subtype is similar to MaSC signature[39], and luminal progenitor gene signature strongly correlates with the basal-like breast cancer. As such, it is believed that basal-like breast cancer is derived from luminal progenitor cells, but not basal progenitors[40].

1.2.4 Mechanisms of breast cancer development

As discussed in Section 1.2.2, mammary gland development is regulated by both genetic and epigenetic mechanisms. The aberrant alterations in both aspects are also implicated in breast cancer initiation and progression.

Genetic alterations in breast cancer

Genetic predisposition has been consistently confirmed as a risk factor of breast cancer development. Pathogenic germline mutations in genes involved in DNA repair, cell cycle and apoptosis control, such as breast cancer gene (*BRCA) 1*, *BRCA2*, phosphatase

and tensin homolog (*PTEN)* and tumor protein p53 (*TP53)*, significantly increase breast cancer risk[41]. Epidemiological studies have shown that the lifetime risk of breast cancer among women with inherited mutations in *BRCA1* and *BRCA2* genes are around 82%[42]. Germline *BRCA1/2* mutation testing is therefore recommended to the females with a personal or family history of *BRCA*-associated cancers[43]. An early study identified that BRCA1 is required for the differentiation of MaSCs and progenitors into ER^+ luminal epithelial cells[44]. This might support the later findings that luminal progenitors and not basal stem cells are the target cell population for the transformation in *BRCA1*-associated tumorigenesis[40, 45]. This may further explain that the majority of *BRAC1*-mutation carriers develop the breast cancers with basal-like phenotype[41].

Recent large scale genomic sequencing efforts have revealed breast cancer to have few somatic mutations (point mutations and small insertions/deletions) but many large copy number alterations, relative to other cancer types[46]. As such, one may assume germline copy number alterations to have an impact on breast cancer development, but this is still an understudied topic.

Several essential signaling pathways in mammary gland development, such as Notch, Hedgehog, Wnt signaling etc. are frequently dysregulated during tumorigenesis[33, 47]. For instance, activating mutations in key components of Wnt/β-catenin signaling result in the overactivation of this signaling[48] which promotes the dedifferentiation of progenitor cells, and consequently the expanded stem cell pool increases the frequency of tumorigenic transformation[49]. The *Wnt1* locus was originally identified as a strongly favoured integration site of mouse mammary tumor virus (MMTV). Based on this, MMTV-*Wnt1* transgenic mouse model has been developed and widely used as a spontaneous breast cancer mouse model in breast cancer studies[31].

Epigenetic alterations in breast cancer

Genomic DNA interacts with histones to form the DNA-protein complex known as chromatin. As the fundamental unit of chromatin, the nucleosome consists of a histone octamer wrapped by 146 bp of DNA[50] (Figure 5A). The term epigenetic broadly refers to the heritable changes in gene expression without changes in DNA

sequences[51]. DNA methylation and histone modification are two major areas of epigenetic regulations[52]. Today, noncoding RNAs and nucleosome (chromatin) remodeling have also been recognized to participate in epigenetic landscape[53]. Abnormal epigenetic events frequently complement genetic changes and contribute to cancer initiation and progression[54]. This section will focus on DNA methylation, histone modification, chromatin structure remodeling and transcriptional regulatory proteins to introduce the epigenetic changes in breast cancer.

DNA methylation

DNA methylation is the earliest and the most broadly studied epigenetic modification in cancer[52, 55]. DNA hypermethylation frequently occur in the CpG-rich promoter regions to silence gene expression by altering chromatin structures to regulate DNA-protein binding[56] (Figure 5B). DNA hypermethylation is detected in various tumor-suppressor genes and involved in different stages of breast cancer development. It has been discussed above that aberrant BRCA1 function is associated with increased breast cancer risk. In addition to genetic mutations, BRCA1 function can also be epigenetically regulated. The promoter hypermethylation of *BRCA1* gene was observed in around 10 - 15% of breast tumors in women without family history of BRCA-associated breast cancers[57]. Also, TNBCs with *BRCA1* promoter hypermethylation present a similar phenotype as *BRCA1*-mutated TNBCs, and response to the chemotherapy designed for *BRCA1*- mutated TNBCs[58]. Recently, it has been shown that low mosaic levels of *BRCA1* methylation in normal cells, presumably occurring in early embryonic life, is significantly associated with increased risk of ovarian cancer later in life[59]. How such methylation affects the risk of breast cancer is still unknown.

In contrast to DNA promoter hypermethylation, DNA hypomethylation has no specific site preferences. Global DNA hypomethylation has been accepted as a hallmark of human cancer, which promotes cancers mainly through affecting chromosomal stability[60].

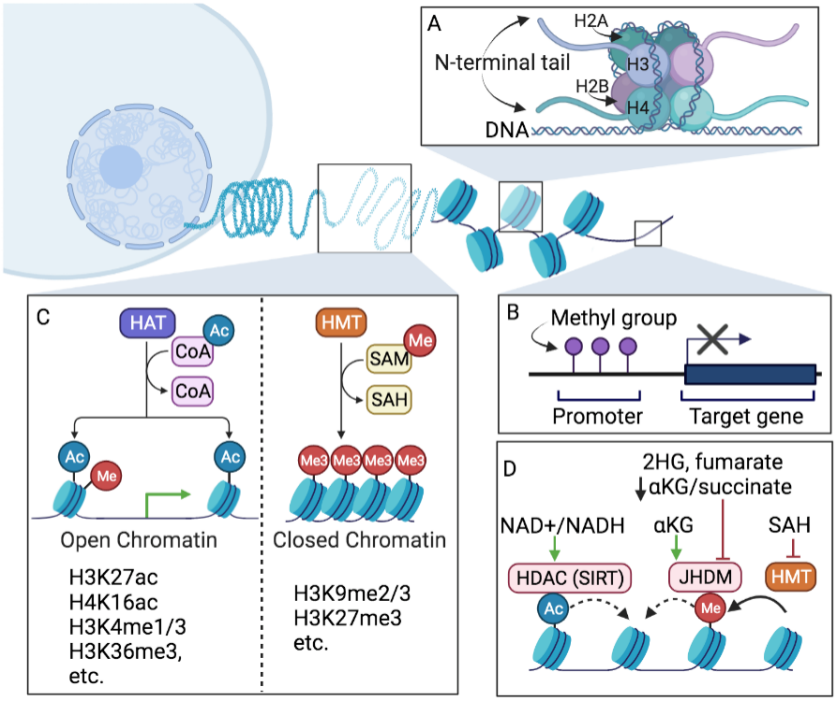


Figure 5. The illustration of nucleosome organization, epigenetic alterations and metabolites effects on modification enzymes. (A) The core of nucleosome is an eight-histone protein complex which contains two copies of each of histones H2A, H2B, H3 and H4. Epigenetic modifications can occur on both DNA and histone N-terminal tails. (B) DNA methylation frequently occurs within CpG islands. Promoter DNA hypermethylation usually inhibits gene expression. (C) Histone modifications are catalyzed by histone modifying enzymes and affect chromatin structures. Histone acetyltransferase (HAT) and histone methyltransferases (HMT) catalyze histone acetylation and methylation respectively; High acetyl-CoA level enhances HAT activity. Likewise, high S-adenosylmethionine (SAM): S-adenosylhomocysteine (SAH) ratio increases the activity of HMT. Different histone modifications in variant locations may regulate "open"/ "closed" chromatin status and further affect the transcription of downstream genes. (D) Metabolites affect the activities of histone modifying enzymes. The increased NAD^+/NADH ratio enhances the activity of NAD^+ - dependent histone deacetylase SIRT1. High level of α-ketoglutarate (αKG) activates JmjC-domain containing histone demethylase (JHDM). On the contrary, decreased 2-hydroxyglutarate (2HG), fumarate and αKG/succinate ratio inhibit JHDM activity. Figure made based on[52, 61, 62].

Histone modifications

Histone modifications frequently occur on histone lysine, arginine and serine residues. According to different chemical groups, histone modifications mainly include methylation, acetylation, phosphorylation, ubiquitination, sumoylation and adenosine diphosphate ribosylation in different degrees (e.g., di- and trimethylation)[50]. These modifications alone or in combination with DNA methylations play crucial roles to regulate the status of chromatin structure (Figure 5C). For example, histone acetylation is commonly associated with transcriptional activation. This modification unfolds chromatin by neutralizing the basic charge of the lysine[63]. Moreover, other histone modifications, such as trimethylated histone 3 lysine 4 (H3K4me3), and H3K36me3 have also been recognized as transcriptional activation markers, whereas H3K9me2/3, H3K27me3 and DNA hypermethylation are linked to closed chromatin[64-66].

Addition or removal of modifications are normally catalyzed by histone-modifying enzymes. The activities of such writer and eraser enzymes are largely dependent on the availability of substrates and cofactors in a microenvironment[61]. A good example is the pair of enzymes working oppositely on lysine acetylation – histone acetyltransferases (HATs) and histone deacetylases (HDACs). HATs use acetyl-CoA (acetyl coenzyme A) as an acetyl donor to transfer acetyl groups to lysine residues. This process is reversed by HDACs which remove acetyl groups and restore the positive charge of the lysine. The level of acetyl-CoA is therefore essential for the activity of these enzymes. Increased intracellular acetyl-CoA level enhances the activity of HATs, which globally promote histone acetylations[67]. In contrast, inhibition of acetyl-CoA synthesis or decreased availability of acetyl-CoA sources, such as glucose, citrate and acetate, induce a rapid deacetylation[68] (Figure 5C). Likewise, histone methylation is also controlled by enzymes with opposing activities - histone methyltransferases (HMTs) and demethylases (HDMs). As a universal methyl donor, S-adenosylmethionine (SAM) is converted to S-adenosylhomocysteine (SAH) during histone methylation. Thus, besides the cellular SAM abundance, the ratio of SAM:SAH is also a rate limiting factor for methylation process[69] (Figure 5C). In addition, a variety of metabolic intermediates, such as α-ketoglutarate (αKG), 2-hydroxyglutarate

(2HG), succinate, fumarate, and nicotinamide adenine dinucleotide (NAD^+/NADH), also serve as cofactors that affect the activities of histone-modifying enzymes (summarized in Figure 5D).

Unlike genetic mutations, most epigenetic modifications are reversible and regulated by enzymes, which allows cancer cells to continuously adapt to alterations in environment. This feature formed the basis for the development of epigenetic therapies[52]. Currently, most epigenetic anti-cancer drugs that have been approved or in clinical trials are small molecular inhibitors of HDACs or DNA methylation. For instance, 5-azacytidine and 5-Aza-2′-deoxycytidine have been approved by FDA for myelodysplastic syndrome and leukemia[70]. The HDACs inhibitors, entinostat and panobinostat are in clinical trials for the use in treatment of solid tumors, including breast cancer[71]. In addition to the epigenetic therapies, epigenetic profiles can be used for the prediction of cancer risk and prognosis, as well as the diagnosis of multiple cancer types. For example, the hypermethylated glutathione S-transferase P1 (GSTP1) has been identified as a biomarker for prostate cancer[70]. In addition, global DNA hypomethylation is a hallmark of cancer[60].

Nucleosome remodeling

In addition to above-mentioned histone modification-induced chromatin remodeling, a group of ATP-dependent chromatin remodelers can also drive such events by repositioning nucleosomes[72]. Since histone modifying enzymes and regulatory proteins can be recruited by chromatin remodeling complexes, ATP-dependent complexes and histone modifications are likely to work together to regulate chromatin structures.

Based on distinct domains, ATP-dependent chromatin remodeling complexes can be classified into four major families: switching/sucrose non-fermenting (SWI/SNF), nucleosome remodeling and histone deacetylase (NuRD)/chromodomain helicase DNA binding (CHD), imitation switch (ISWI) and inositol (INO80)[72, 73]. The fundamental mechanism for all remodeler-mediated chromatin changes is DNA translocation which refers to the process that remodelers slide histone octamers to alternative positions along

DNA[74]. This process can result in the assembly, disassembly, reposition, exchange and removal of histone proteins according to the specific subunits recruited in the remodeling complexes[73].

SWI/SNF remodelers and NuRD complexes have been implicated in human cancer progression[75, 76]. SWI/SNF remodelers are normally associated with an open chromatin. Consequently, inactivating mutations in SWI/SNF subunits may lead to the loss of chromatin accessibility, which has been widely identified in several cancer types. In the context of breast cancer, more than 25% of patients harbor mutations in SWI/SNF remodeler subunits[74]. The inactivating mutation on SWI/SNF subunit encoding gene *ARID1A* is frequently observed in advanced ER^+ breast cancer[77]. *ARID1A* mutations impair the binding of SWI/SNF complex to the luminal lineage-determining transcription factor loci. This has been suggested to affect the chromatin accessibility of transcription factor motifs involved in luminal differentiation, which in turn promotes a luminal to basal-like transformation of breast cancer cells. In addition, ER- and forkhead box A1 (FOXA1)- chromatin interactions are inhibited by *ARID1A* mutation, which could drive endocrine therapy resistance[77].

NuRD complex subunits, metastasis-associated gene (MTA) 1 and MTA3 have also been reported to affect estrogen functions and breast cancer progression[76]. The increased levels of MTA1 are frequently detected in breast cancer and closely associated with high tumor grade[78]. In a NuRD complex, MTA1 is able to interact with HDACs and guide the complex to the activation function 2 (AF2) binding domain of ERα, and repress ER-mediated transcriptions[79]. Another MTA family member, MTA3 exhibits opposite expression patterns of MTA1 in breast cancer. MTA3 is an estrogen-dependent component of NuRD complex, and it is gradually lost during breast cancer progression[80]. Functionally, MTA3 interacts with HDACs in the NuRD complex to repress the expression of *Snail* and promote the E-cadherin expression[81]. Thus, the loss of MTA3 promotes epithelial to mesenchymal transition (EMT) and enhances the invasive potential of breast cancer cells[80]. When NuRD complex contains a methyl-CpG-binding domain 2 (MBD2), the complex can be recruited into the hypermethylated promoter of tumor suppressor genes to ensure gene silencing[76].

The distinct roles of MTA1-, MTA3- and MBD2-NuRD complex in cancer progression reflect that the genomic targeting and functional selectivity of a chromatin remodeler largely depend on the assembled subunits in the complex.

Transcription factors in breast cancer

The epigenetically controlled chromatin remodeling potently regulates chromatin accessibility for the transcription factors and other regulatory proteins involved in transcriptional regulation. Transcription factors contain specific DNA binding motifs which guide the recruitments of transcription factors to defined gene promoter or enhancer regions and thereby regulate gene expression[82]. Several major transcription factor families play important roles in different stages of breast cancer development. For instance, transcription factors octamer-binding transcription factor 4 (OCT4), Sex determining region Y box 2 (SOX2), NANOG, Kruppel-like factor 4 (KLF4) and myelocytomatosis viral oncogene homolog (MYC) are essential for maintaining the self-renewal and pluripotency of embryonic stem cells (ESCs). Moreover, ectopic expression of these transcription factors is able to convert somatic cells to pluripotency[83, 84]. High expression levels of these transcription factors have also been detected in cancer stem cells (CSCs) and most of them act similarly as in ESCs. For instance, *OCT4* expression is significantly increased in $CD44^{+}/CD24^{-/low}$ breast CSC-like cells[85] and is associated with poor prognosis in hormone receptor-positive breast cancers[83]. The high expression of SOX2 in basal-like breast cancer promotes a less differentiated stem-like phenotype[86]. Also, NANOG levels are gradually decreased along with the differentiation of stem cells and is silenced in normal somatic cells. Interestingly, the re-expression of NANOG has been observed in many CSCs[83]. The functional similarities of these pluripotent transcription factors in ESCs and CSCs provide evidence to link the functions of transcription factor CCAAT/enhancer binding protein beta (C/EBPB) in mammary gland development and in the breast cancer stemness phenotype.

C/EBPB, a member of C/EBP transcription factor family, contains a highly conserved basic-leucine zipper dimerization and DNA-binding domain[87]. Three distinct protein

isoforms of C/EBPB have been identified, 44 kDa (38 kDa for mouse) liver activating protein* (LAP* or LAP1), 42 kDa (34 kDa for mouse) LAP (or LAP2) and 20 kDa liver inhibitory protein (LIP), and they are encoded by an intronless gene with different in-frame translation start sites (Figure 6)[87].

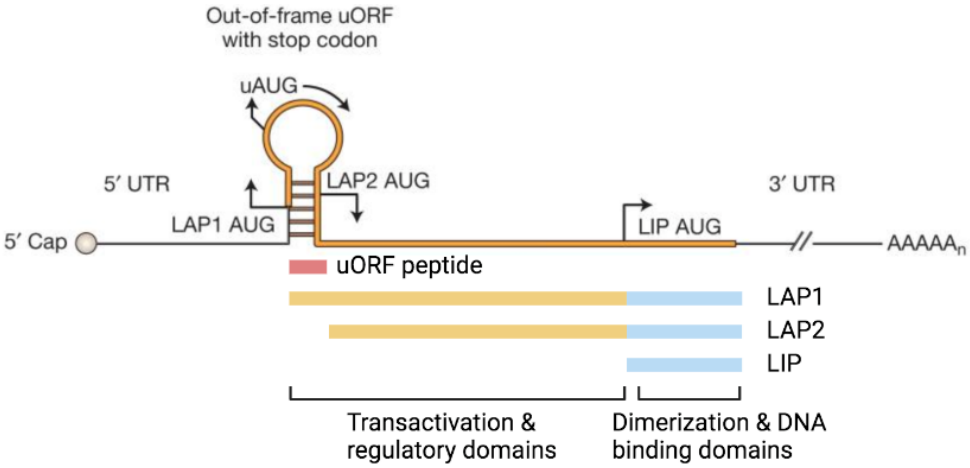


Figure 6. mRNA structure of C/EBPB and translated C/EBPB isoforms. *C/EBPB* mRNA contains several in-frame translation start codons and an out-of-frame upstream open reading frame (uORF). LIP translation initiates at LIP AUG codon following the leaky ribosome scanning over the upstream start codons. The expression of uORF peptide is crucial for LIP translation, and the mutation at the uORF site abrogates LIP expression. For the three isoforms, LAP1 and LAP2 contain both transactivation and inhibitory domains, while LIP lacks the N-terminal activation domain and frequently functions as a dominant-negative inhibitor. Figure modified from [88, 89]

C/EBPB target genes contribute to diverse biological processes, such as cell proliferation, differentiation, inflammation, metabolism and determination of stem cell fate in a variety of tissues, including the mammary gland[88]. Expression of C/EBPB is dramatically changed during mammary gland development and pregnancy-lactation-involution cycles, which tightly reflects its regulatory functions on ductal morphogenesis and lobuloalveolar proliferation[90]. In the virgin mammary gland, all three C/EBPB isoforms are present at low level, which increases during pregnancy. In particular, the LIP isoform increases more than 100-fold before declining again at

parturition. Afterwards, LIP expression remains low throughout lactation and involution, whereas LAP expression dramatically increases during involution stage[90]. It has been reported that *C/ebpb*-deficient mice display delayed ductal outgrowth, enlarged ducts with decreased branching[91], and impaired lobuloalveolar development in pregnant[92]. A study by LaMarca *et al.* identified the roles of C/EBPB in mammary stem/progenitor cell self-renewal and maintenance[93]. By using both germline *C/ebpb*$^{-/-}$ mice and a conditional knockout strategy, they showed that the depletion of *C/ebpb* on mammary epithelial cells impaired mammosphere formation ability. Moreover, the MaSCs frequency significantly decreased in *C/ebpb*$^{-/-}$ mice[93]. In addition to the broad roles in mammary gland development, C/EBPB has been implicated in hematopoiesis, osteogenesis, adipogenesis as well as breast cancer progression[90].

Whereas functional mutations in *C/EBPB* are rarely detected in solid tumors[94], the altered *C/EBPB* expressions are frequently observed in breast cancer, especially in TNBC[95, 96]. A study by Bundy *et al*. reported that the overexpression of human *LAP2* in MCF10a epithelial cells induced tumorigenic transformation through a C/EBPB-COX2-mediated EMT process[97]. This phenotype was consistently observed by another study, in which the upstream regulation of *C/EBPB* was further investigated. In this study, they discovered that the Ras/Raf/MEK/ERK signaling activated c-MYC to reduce the expression of brahma (BRM), the core enzyme of SWI/SNF chromatin remodeling complex. The decreased BRM epigenetically promoted *C/EBPB* expression, thereby contributing to the malignant transformation[98]. Other mechanisms, such as AMP-activated protein kinase (AMPK)- LAP-mediated breast cancer immune suppression[95] and C/EBPB-JAK/STAT-induced aggressive TNBC[96] were also demonstrated. However, all of the above C/EBPB involved mechanisms in mammary gland development and breast cancer progression do not directly address how C/EBPB signaling and which specific isoforms contribute to breast cancer stemness. In our study (Paper I), the mechanism that C/EBPB epigenetically regulates target genes to drive the breast cancer stem-like property is identified in obese

scenarios. Moreover, the specific isoform LAP2 is the crucial player in this transcriptional network.

Pioneer factors are a group of special transcription factors which are able to access their DNA target sites in compacted chromatin, thereby competing or recruiting other factors to regulate gene transcription[99]. Thus, pioneer factors may act both active and passive roles (Figure 7A). FOXA and GATA binding protein (GATA) families were the first described and are the most extensively studied pioneer factors. During early development, several FOXA and GATA factors, such as FOXA1, FOXA3, GATA4 and GATA6 are able to bind highly compacted chromatin. Binding of these pioneer factors could either stabilize nucleosome for other factors binding to form a transcription factor complexes[100] or open chromatin to allow transcription factors, chromatin modifiers, and nucleosome remodelers to bind[101, 102]. This process is necessary to initiate downstream regulatory events, and normally in a transient manner, which reflects that pioneer factors play more crucial roles for initiating developmental lineage rather than maintaining it[103]. In contrast, such prior binding may also act passively. In this condition, the stable binding of pioneer factors can reduce the number of subsequent factors binding to the specific chromatin regions, which in turn inhibit target gene transcription. This passive mode is frequently observed in undifferentiated stem/progenitor cells[101].

In addition to the early-stage development, pioneer factors have been investigated to govern ER and other nuclear hormone receptor binding events in differentiated cells, including breast cancer cells. Intriguingly, in ER^+ MCF-7 breast cancer cells, over 50% of ER-binding sites overlap with a FOXA1 binding regions[104]. The chromatin compression is relieved upon FOXA1 binding, which allows ER to bind to the estrogen-responsive elements and mediate transcription of its target genes. FOXA1 displays preferences for particular binding sites, for example the H3K4me1 and H3K4me2 enriched domains, while DNA hypermethylation may impair FOXA1 binding[105] (Figure 7B). GATA3 has been reported to act in a similar manner in MCF-7 cells, and GATA3/ER-mediated transcription of target genes are essential for the growth of breast

cancer cells[106]. In fact, this pioneer factor-mediated ER binding has been suggested as the reason of drug resistance in hormone-dependent cancers[99].

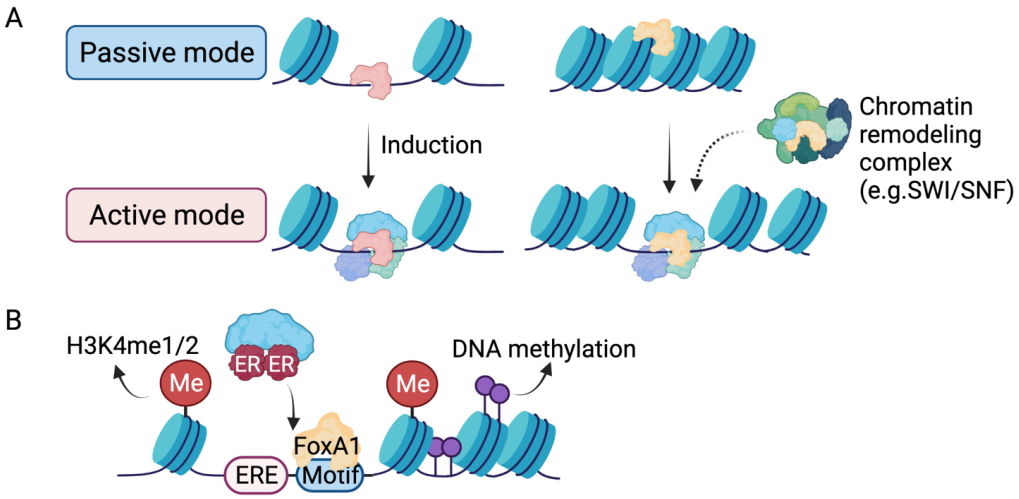


Figure 7. Illustration of active and passive roles of pioneer factors and working model in breast cancer. (A) Pioneer factors play both active and passive roles in different transcriptional events. The prior binding of pioneer factors can reduce the binding of other factors, thereby maintaining the passive transcriptional mode. Under certain inductions, such as estrogen treatment, pioneer factors facilitate the recruitment of cofactors to induce an active mode of transcription. Pioneer factors binding can relieve highly compact chromatin and prepare the chromatin platform for the binding of other regulatory factors. Chromatin remodeling complexes may facilitate this transition, but are not necessary. (B) Schematic representation showing the binding preferences and function of pioneer factor FOXA1 in ER^+ breast cancer cells. Figure modified from [101].

Pioneer factor mechanisms provide basis for the design of new anti-cancer strategies, especially for hormone-dependent cancers. To this end, further explorations are required to form a comprehensive understanding of the dynamic interactions among pioneer factors, hormone receptors, co-factors and cancer chromatin.

1.2.5 Breast cancer heterogeneity and evolution

Cancer heterogeneity refers to the variation among tumors from different patients (intertumor heterogeneity) as well as within a single tumor (intratumor heterogeneity)[107]. Breast cancer intertumor heterogeneity is critical for classifying breast cancer subtypes, while the intratumor heterogeneity reflects the diversity of different subpopulations and even individual cancer cells in a tumor mass (Figure 8A)[107]. The multiregional sequencing study adds further information about spatial and temporal heterogeneity in breast tumors[108]. The uneven distribution of diverse subclones across different regions is the major cause of spatial heterogeneity, and temporal heterogeneity refers to the variations generated from natural progression or selective pressures over time[109]. In the above-mentioned study, it was identified that the extent of breast tumor diversification followed spatial pattern, however no strict temporal order was identified[108].

Cancer progression can be viewed as an evolution process. During this process, both genetic and epigenetic alterations contribute to the breast cancer heterogeneity - a key contributor for cancer cell survival, especially under certain pressures, such as anti-cancer treatments and metastasis[110]. With the development of molecular profiling technologies, extensive potential models have been proposed to illustrate this process, which are generally classified as Darwinian selection or Lamarckian induction[111] (Figure 8B).

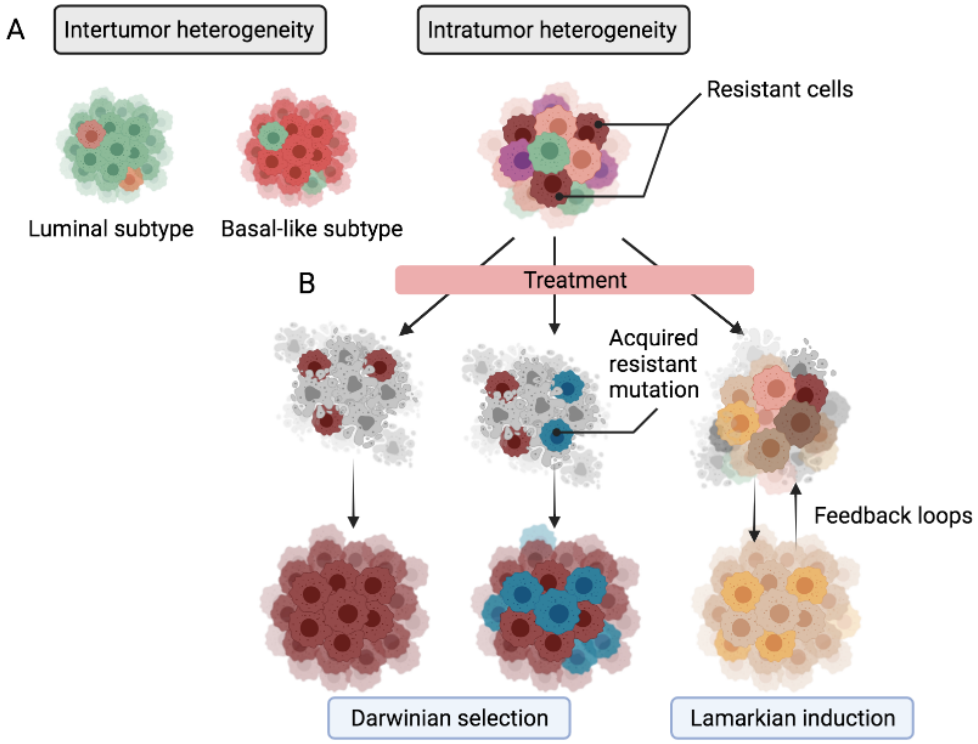


Figure 8. Illustration of tumor heterogeneity and potential models for drug resistance development. (A) Tumor heterogeneity includes inter- and intratumor heterogeneity. (B) Darwinian selection and Lamarckian induction are widely accepted models underlying the formation of drug resistance. For Darwinian selection model, the pre-existing resistant cells or the acquired resistant cells during treatment can survive and outgrow under a selection pressure. Lamarkian induction refers to the process that cells adapt to the altered environment through cell plasticity or treatment-induced feedback loops. Figure modified from [107, 111].

Darwinian selection

In the context of drug resistance, Darwinian selection originally referred to the survival and outgrowth of pre-existing resistant cell clones under a potent therapy pressure, which is normally driven by the pre-deposited resistant mutations[112]. Later, this theory was expanded to explain acquired resistance. For example, in many cancer types, such as ovarian cancer, glioblastoma and breast cancer, unique mutations have been observed in the post-treated or relapsed tumors and associated with treatment

resistances. However, these mutations are absent or rarely detected from pretreatment tumors[113-115]. This proves that *de novo* acquired mutations are induced during treatment, and certain subclones carrying these mutations are selected and outgrow. Interestingly, a large proportion of patients who acquired drug resistance did not show clear evidence of genetic evolutions, which raises the possibility that acquired resistance may result from non-genetic evolutions (termed non-genetic gradual Darwinian selection)[111]. For example, bromodomain and extraterminal domain (BET) bromodomain inhibitors have been used for TNBC treatment. These inhibitors work as a competitor of BET bromodomain proteins, such as bromodomain containing 4 (BRD4) and thereby inhibit oncogenic transcriptional events. During treatment, a population of cancer cells through hyperphosphorylating BRD4 facilitate a MED1 mediated bromodomain-independent chromatin binding and escape from the inhibited states. This population of cells are selected and outgrown, and form BET bromodomain inhibitor-resistant phenotype[116].

Lamarckian induction

Unlike the Darwinian selection process, Lamarckian induction does not include a clonal selection step; instead, it occurs through the induction of an altered environment and is mainly driven by cell plasticity[117]. During the drug treatment, cancer cells change the state toward a more drug-tolerant state, such as stem-like and mesenchymal states[118]. A longitudinal single-cell analysis identified the adaptation in individual resistant leukemic cells. Drug treatment rapidly stimulated the cell-individual induction of multidrug resistance protein 1 (MDR1) expression in a large population of cells (comprising 30-40% of the cells). The MDR1-dependent drug resistance was unlikely induced by a selection of pre-existing MDR1 highly expressed clones[118]. Another study in ER^+ breast cancer shows that the resistance of ER-targeted therapies could be driven by the estrogen-induced feedback loops[119]. ER-targeted therapies inhibit downstream C-terminal SRC kinase (CSK) expression, and the low level of CSK promotes p21 protein-activated kinase 2 (PAK2) expression which is associated with endocrine therapy resistance. This case raises a compensation-based model for drug resistance, which has been consistently observed in different cancer types[120].

Darwinian selection or Lamarckian induction are not completely independent processes, and they interact with each other in numerous cases[121-123]. During the early stage of effective treatment, although the drug-tolerant state can be induced rapidly, most of cells will still be killed by the potent selection pressures. In addition, the Lamarckian induction-induced drug-tolerant state may provide cells more opportunities to acquire genetic and epigenetic changes[111]. Thus, to overcome drug-resistance both mechanisms need to be considered.

1.2.6 Breast cancer treatments

The therapeutic strategies for breast cancer depend on multiple factors including both histopathological and molecular characteristics. The phase and stage of cancer progression are especially important for making the principles of therapy.

Nonmetastatic breast cancer

The early-stage breast cancer without distant metastasis accounts for more than 90% of total breast cancer cases at diagnosis[124]. The principle of treatment for this stage is tumor eradication and preventing recurrence[124]. To this end, surgical resection is normally adopted, followed by systemic therapies and/or radiation therapy. For some specific cases, for example, large tumor size at diagnosis, systemic therapies are necessary before surgery. The systemic therapies, including endocrine, chemo and targeted therapies, are determined by the molecular subtype of breast cancer, tumor burden and recurrence risks. The patients with hormone receptor positive tumors should receive a standard adjuvant endocrine therapy for at least 5 years after surgery to block ER activity or reduce the circulation level of hormone, and the combination with chemotherapy can be considered according to the risk of recurrence[25]. HER2-targeted therapy is necessary for the patients with $HER2^+$ breast cancer, and the combination with chemotherapies significantly improves patients' survival[125]. TNBC has the fewest therapeutic options due to the lack of hormone receptors expression. Therefore, chemotherapy is generally given to all patients with this subtype. Still, TNBC patients have the lowest response rate and the worst prognosis. Around 70% of TNBC deaths occurs in 5 years after diagnosis[126]. However, recent data indicate TNBC to be the

breast cancer subtype benefiting most from novel treatment strategies such as immunotherapy (see below) and poly (ADP-ribose) polymerase (PARP) inhibition.

Metastatic breast cancer

Metastasis is the leading cause of cancer mortality and responsible for more than 90% of cancer deaths[127]. For the patients with a diagnosis of metastatic breast cancer, the main goal of treatments is prolonging survival and alleviating treatment-related toxicity. To this end, surgery to remove the primary tumor is normally not recommended, and systemic therapies are administrated according to multidisciplinary evaluations, especially the types of breast cancer[25, 124]. For instance, endocrine therapies are still necessary for the metastatic hormone receptor positive breast cancer patients, and it is frequently combined with cyclin-dependent kinase (CDK) 4/6 inhibitors to delay the endocrine resistance[128]. Several clinical studies have reported that the combination treatment significantly increases the progression-free survival by 5-10 months[129, 130]. Generally, the median overall survival of patients with hormone receptor positive metastatic breast cancer is around 4-5 years, whereas for the patients with metastatic TNBC it is only around 10 months[124].

In recent years, an increasing number of immunotherapies have been approved for breast cancer treatment, which especially are benefits for the patients with advanced TNBC. Immune checkpoint blockade (ICB) agents, such as anti-programmed cell death receptor 1 (PD-1) antibodies, anti-programmed cell death ligand 1 (PD-L1) antibodies and inhibitors of cytotoxic T-lymphocyte–associated antigen (CTLA-4), are the most investigated immunotherapies for breast cancers. In 2019, the first ICB agent, atezolizumab (anti-PD-L1 antibody) was approved by FDA for the use in combination with nab-paclitaxel in metastatic TNBC treatment[131]. Also, several ICB monotherapies are in phase III clinical trials for metastatic breast cancer treatments[131]. Besides the ICB agents, several kinds of adoptive cellular therapies (ACT), including administration of tumor-infiltrating lymphocyte (TIL), administration of genetically engineered lymphocytes to express a specific T-cell receptor (TCR) or express chimeric antigen receptors (CARs) have been approved for leukemia treatment

and are currently being investigated in phase I and II clinical trials for the use of breast cancer treatment[132].

1.3 Obesity breast cancer connection

1.3.1 Epidemiology

Obesity is a well-established risk factor for cancers and contributes to the increased incidence of more than 13 cancer types[133]. Generally, obese individuals have 1.5 to 4 times higher risks to develop different cancer types compared with non-obese individuals[134]. In addition, obesity is also positively associated with increased cancer mortality. It has been estimated that when BMI exceeds 40 kg/m^2, the death rates from all cancer types are around 60% higher than the rates in normal weight patients. Overall, around 15-20% of cancer-related deaths can be attributed to overweight and obesity[135].

However, in the context of breast cancer, this association is more complex and largely depends on patients' menopausal status and the hormonal receptor status of breast cancer. It has been consistently shown that overweight and obesity are positively associated with increased breast cancer risks in postmenopausal women, especially in older postmenopausal women[136, 137]. In this population, the risk of breast cancer increases by 12% for every 5 BMI units[138]. In addition, obesity is associated with reduced 5-year disease-free survival and overall survival in postmenopausal patients, and the risks of cancer relapse and death are strongly increased in this group[139]. In premenopausal women, the association is less clear. Interestingly, some studies even demonstrate that obesity may be protective for premenopausal women [137]. Other measures of obesity, such as waist circumference and waist-hip ratio can better estimate the fat distribution, but they are not consistently associated with breast cancer risk[136].

Several epidemiological studies specifically analyzed the association between elevated circulating free fatty acid (FFA) level and breast cancer risk. A study by Hirko *et al*. claimed that among overweight and obese women, the level of saturated fatty acids and dairy-derived fatty acids were positively associated with breast cancer risk, and this

association was reversed in normal weight women[140]. Moreover, a meta-analysis study analyzed 4365 patient cases from three cohort and seven case control studies. The result showed that total monounsaturated fatty acids, oleic acid and palmitic acid (PA) were significantly associated with the increased breast cancer risks. However, the total saturated fatty acid was only associated with postmenopausal breast cancer risk but not in premenopausal women[141].

1.3.2 Molecular mechanisms

Although epidemiological studies have consistently shown a positive relationship between obesity and postmenopausal breast cancer risk, the mechanisms underlying this association are not well defined. In recent years, an expanding number of mechanisms have been identified to mediate this connection (summarized in Figure 9).

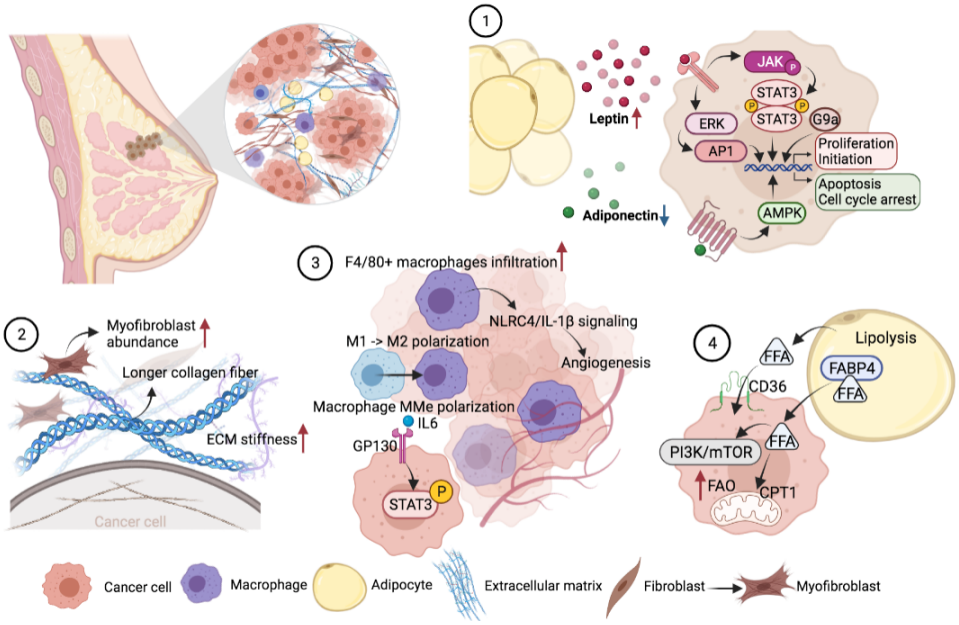


Figure 9. Summary of potential molecular mechanisms connecting obesity and breast cancer development. Previously reported molecular mechanisms are summarized and classified into four aspects: (1) dysfunctional adipocytes, (2) ECM remodeling, (3) Inflammation and (4) Metabolic reprogramming.

Dysfunctional adipocytes

The primary site of breast cancer is rich in adipose tissue; thus, obesity-induced dysfunctional adipocytes may abnormally secrete multiple circulating factors to drive an altered tumor inducing microenvironment.

Leptin

Leptin is predominantly secreted by adipocytes. After getting into circulation, leptin regulates multiple biological processes through binding to tyrosine kinase leptin receptors. One of its best known functions is to control energy balance through regulating food intake[142]. Obese individuals normally present higher serum leptin level compared with normal-weight people[143], and they are frequently insensitive to endogenous leptin[144]. In addition, leptin also regulates the generation of reproductive hormones[145]. Thus, leptin has been implicated as an important factor mediating the obesity-breast cancer connection. Epidemiological studies have shown that the increased level of circulating leptin is associated with the increased risks of breast cancer and the more aggressive cancer phenotypes[145, 146]. Moreover, leptin and leptin receptor have been observed significantly overexpressed in breast cancers[147], which is able to enhance the proliferation of malignant breast epithelial cells *in vitro*[148]. Mechanistically, leptin induces the phosphorylation of STAT3 in a time- and dose-dependent manner, and it also activates extracellular signal-related kinase (ERK) 1/2. Both pathways are implicated in leptin-induced proliferation of breast cancer cells[148, 149]. In addition to cell growth, leptin has also been proposed to enhance tumor initiation in obese environment through leptin-STAT3-G9a HMT axis[150] or by promoting EMT process and CSC survival[151]. However, the larger tumors size, higher tumor incidence and greater number of metastases have been extensively observed in leptin-deficient *ob/ob* mouse model across different cancer types[152-154], suggesting that the leptin signaling may not be required for obesity-induced cancer progression.

Adiponectin

Adiponectin, another adipocyte-derived hormone, has also been suggested to play roles in the obesity-breast cancer connection. In contrast to leptin level, circulating adiponectin concentration is negatively correlated with body fat mass[155]. Epidemiological studies have indicated that adiponectin level is inversely associated with breast cancer risk[156]. In healthy individuals, adiponectin mainly functions as a starvation hormone, which is able to stimulate AMPK pathway and thereby promotes food intake and fat storage in fasting conditions[157]. The expression of adiponectin receptors have been observed in several human breast cancer cell lines, and *in vitro* studies have further revealed anti-proliferative, inhibition of invasion and migration effects of adiponection[158-160]. Adiponectin-stimulated AMPK signaling pathways are implicated in all of these effects. For instance, adiponectin exposure can induce cell apoptosis through AMPK-mediated p53-bax pathway, inhibit cell invasion and migration via AMPK–S6K axis, and block cell cycle by AMPK-inhibited-cyclin D1[158-160]. In addition, adiponectin has also been reported to inhibit tumorigenesis through blocking the phosphorylation of AKT-GSK-3β signaling[161] .

Other adipokines

In addition to the abnormal leptin and adiponectin secretions, a recent study reported that obesity-induced elevated level of resistin was able to promote breast tumorigenesis[162]. In obese state, the increased circulating FFA simulates peroxisome proliferator-activated receptor γ signaling, thereby activating the TAZ-mediated resistin expression[162]. Furthermore, the altered estrone (E1): estradiol (E2) ratio has been proposed to explain the paradoxical effects of obesity in pre- and postmenopausal breast cancer progression[163]. After menopause, the high E1:E2 ratio could stimulates NFκB-mediated signaling to promote tumor stemness properties in obese patients[163].

Extracellular matrix (ECM) remodeling

As a major component of tumor microenvironment, ECM provides mechanical and physicochemical supports for all the cells embedded into it, including carcinoma

cells[164]. It has been identified that obesity can induce ECM remodeling in breast tissue[165]. Cancer-free breast tissues isolated from obese individuals display increased myofibroblast content compared with the tissues from non-obese individuals, which further enhance the deposition of ECM components and the stiffness of matrix. All of these alterations provide a suitable environment for the transformation of premalignant cells[166]. In addition to the tumorigenic effect, obesity-associated ECM remodeling also contributes to breast cancer progression. Tumor specimens collected from obese breast cancer patients display severe fibrosis features, with thicker and longer collagen fibers compared with the tissues harvested from non-obese patients. These structural alterations of ECM have been suggested to promote the migration and invasion of breast cancer cells[165, 166].

Collagen VI (COL6), one of the important ECM proteins, is secreted by adipocytes and abundantly expressed in white adipose tissue[167]. In humans, *COL6A1*, *COL6A2* and *COL6A3* genes encode three distinct COL6 chains (α1, α2, and α3) which assemble to the complex COL6 tetramers and microfibrils[168]. COL6 interacts with other ECM components to organize and maintain tissue architecture[169]. Upregulated COL6 has been observed in both obese mammary gland and breast tumor lesions[170]. In addition to the architectural supports, COL6 also functions as a signaling molecule, and through interacting with neuron-glial antigen 2 (NG2) - epidermal growth factor receptor (EGFR) drives the mitogen-activated protein kinase (MAPK)-mediated migration and invasion of TNBC cells[170]. Furthermore, the interaction with NG2/chondroitin sulfate proteoglycan (CSPG) receptor can stabilize cyclin D1, which in turn facilitates breast cancer cell proliferation[167]. In addition, the cleavage product of COL6A3 chain, endotrophin (ETP), has also been reported to stimulate mammary tumor initiation and progression in the obese preclinical models[171, 172].

Inflammation

Low-grade chronic inflammation is a well-recognized feature of obesity and happens both systemically and at specific tissues[173]. Obesity-related systemic inflammation is characterized by increased circulating inflammatory cytokines, which not only

promote breast cancer initiation and proliferation locally, but also act globally to support tumor spread[173]. Locally, the chronic inflammatory state is caused by the recruitment of macrophages with abnormal secretion of inflammatory cytokines[174]. In transplant breast cancer animal models, high fat diet (HFD) exposure significantly increases infiltration of $F4/80^+$ macrophages in tumors[175-177]. In addition to the abundance, functionally, the elevated infiltration of tumor-associated macrophages is able to activate the NLRC4 inflammasome which further promotes the Interleukin 1β (IL-1β) activation. The NLRC4/IL-1β signaling does not directly act on tumor cells, however it can induce vascular endothelial growth factor A (VEGFA) expression and angiogenesis, which accelerate breast cancer progression in an obese environment[178]. Furthermore, the altered phenotypes of macrophage also play roles in the obesity-breast cancer connection. In both obese breast cancer patients and mouse models, the circulating placental growth factor (PlGF) levels are significantly increased, which through PIGF/VEGFR-1 axis promote macrophage polarization and contribute to the tumor progression in obesity[179]. Another recent study uncovered that obesity-induced elevated fatty acids drove a metabolically activated (MMe) macrophage phenotype. Unlike M1 macrophages, the pro-tumorigenic MMe macrophages through IL-6/GP130 signaling, promotes TNBC stemness and tumorigenesis[180].

Metabolic reprogramming

Obesity frequently accompanies with altered whole-body metabolism and metabolic syndrome, such as, hyperglycemia, dyslipidemia and insulin resistance[181]. As a hallmark of cancer, the metabolic reprogramming contributes breast cancer progression in multi-stages, including tumor initiation, growth and metastasis according to the availability of nutrient in tumor microenvironment. As what has been introduced in Section 1.1, during obesity, the expanded adipose tissue spills out the extra lipids beyond its storage limit. Thus, elevated FFA levels occur both locally and systemically. The physiological serum FFA concentration of obese women has a large range. It has been reported that the concentration varies from 70 -700 μM according to different measurement methods, the level of obesity and obesity-induced comorbidities [182, 183]. A study by Madak-Erdogan *et al*. systematically analyzed metabolite composition

of plasma samples from 37 non-obese and 63 overweight and obese postmenopausal women by using gas chromatography–mass spectrometry (GC-MS)[184]. The results revealed that several FFAs including linoleic acid, oleic acid, PA and stearic acid were increased in high BMI women, and all of these FFAs were decreased upon weight loss[184].

Several recent studies have pointed out that the altered FFAs could be the link between obesity and breast cancer progression. As a FFA transporter, CD36 overexpression was observed in a small population of cancer cells and characterized the head and neck cancer stem cells[185]. In the HFD-induced obese mouse model, CD36 positive cells were significantly increased in both primary tumor and metastatic lesions[186] This demonstrated that CD36 played an essential role in facilitating cells to adapt an altered microenvironment and drive metastasis[185]. In addition to the FFA transporter, fatty acid binding protein 4 (FABP4) has also been identified to promote cancer progression[187, 188]. Locally, FABP4 supports the interaction between ovarian cancer cells and surrounding adipocytes, and by using the adipocyte-derived fatty acids as an energy source to support a rapid tumor growth[188]. As human breast is also an adipocyte-rich tissue, these findings may also apply to breast cancer progression in an obese environment. Moreover, the FABP4 roles have been expanded to the global level. In obese state, circulating FABP4 level is significantly increased. Here, in addition to the FFA transporter roles, FABP4 acts as a signaling molecule to activate IL-6-STAT3-ALDH1 pathway to induce the breast cancer stemness properties[188]. Obesity-induced elevated circulating FFAs have also been suggested to activate the ERα and mTOR pathways, which further promote metabolic reprogramming and breast cancer progression[184]. Within the tumor microenvironment, cancer cells interact with stromal cells, and different cell types require distinct nutrients from a common metabolites pool[189]. A recent study reported that obesity induced a lipid competition between cancer cells and T cells, and these two types of cells showed distinct metabolic reprogramming. Increased fatty acid oxidation and accelerated fatty acid uptake were observed in cancer cells but not in T cells, which could explain the accelerating tumor growth and the impaired anti-tumor immunity in obese condition[190].

In addition to the mechanisms mentioned above, an expanding number of studies has also proposed other types of obesity-induced metabolites, adipokines and immune cells as potential mechanisms linking obesity and cancer progression. This reflects the complexity of obese environment. The current achievements have greatly advanced our understanding of breast cancer development in obese environment. However, the general mechanisms underlying the connection remain open for discovery. In addition, further studies are still required to form a highly comprehensive map depicting the interaction between these two dynamic pathological conditions, which further urges us to improve *in vitro* and *in vivo* models to mimic both obese conditions and cancer development processes in humans. Consequently, the comprehensive understanding of breast cancer progression in an obese environment will be translated to the development of novel targeted therapies for the benefits of obese breast cancer patients.

1.3.3 Obese patients breast cancer management

In addition to the standard therapies mentioned in Section 1.2.6, specific clinical practices need to be considered for the treatments of obese breast cancer patients. Extensive studies have demonstrated that obese breast cancer patients exhibit significantly higher risk of medical and surgical complications, such as infections, wound dehiscence and systemic venous thromboembolism[191, 192]. Furthermore, when chemotherapy doses are calculated based on patients' actual body weight, clinicians often reduce the doses for overweight and obese cancer patients to avoid the risks of overdose toxicity. It has been widely reported that obese breast cancer patients are often undertreated when they receive systemic anti-cancer chemotherapy[193, 194]. Lyman *et al*. demonstrated in their study that a more than 15% chemotherapy dose reduction occurred in 36.5% of patients[193]. Moreover, severely obese breast cancer patients have four-times higher risk to receive a reduced chemotherapy dose compared with normal BMI patients[195]. The chemotherapy dose reductions frequently result in undertreatment in obese breast cancer patients and contribute to the worse prognosis and poor survival[196]. Importantly, emerging evidence support that overweight and obese breast cancer patients do not experience increased toxicity by using normal dose

chemotherapy. Thus, the reduced chemotherapy dose is not recommended for obese patients in clinical practice[197].

2. Aim of the study

2.1 Overall aim

The overall objective of this project is to investigate the molecular mechanisms of how breast cancer cells adapt to and progress in an obese environment and to investigate the translational potential of an established mechanisms underlying the obesity-cancer connections.

2.2 Specific aims

Specific aim 1

Identify the malignant cellular phenotypes of breast cancer adapted to an obese environment, and systematically extract the molecular mechanisms underlying this connection.

Specific aim 2

Validate the established endotrophin functions in human cell lines, and ultimately to develop targeted therapies and investigate the potential of clinical applications.

3. Methodological considerations

3.1 Patient materials and analysis

The patient's data and materials in the presented projects were collected either from a human breast cancer biobank or public databases.

In paper I, the survival analysis was performed by using a breast cancer patient cohort containing 223 patients with primary stage III breast cancers[198]. The basic clinical information, including patients' BMI, age, cancer hormone receptor status and patients' survival situations were well documented. Tissue microarray experiments were performed on the tissue specimens from the same patient cohort. The formaldehyde fixed and paraffin embedded tissue samples were subject to tissue microarray construction. Microtome sectioned slides were stored at 4°C before use. The study was approved by the regional committees for medical and health research of Western Norway (REK-Vest; approval number 273/96-82.96).

In paper II, to measure circulating ETP levels in human patients, the plasma samples were collected from 195 breast cancer patients with informed consent under the approved Institutional Review Board protocols # HSC-MS-10- 0580 and HSC-MS-11-0559. The mRNA expression of *COL6A3* in tumors and paired healthy tissues was analyzed by using data from 111 breast cancer patients extracted from the cancer genome atlas (TCGA) database. The *COL6A3* expression-based survival analysis was performed by using the same patient cohort as in paper I.

3.2 *In vitro* models

Breast cancer development is a complex multistep process and almost all of the components in the tumor microenvironment participate in this process. With the traditional two-dimensional culture on plastic, it is hard to capture the complexity of tumor progression in a living system, due to the lack of ECM components, physiological substratum and cell-cell/stromal interactions[199]. In addition, with a constant supply

of nutrients in culture media, breast cancer cells will not be challenged by either hypoxia or nutrient deficiency, which are frequently present *in vivo*. However, the ease by which most experimental variables can be easily modified and quantified during standard *in vitro* settings have made this approach popular for the past decade. Especially, when study the effects of a single factor, it is possible to exclude the influences from most cofactors. For instance, in paper I, to assess how breast cancer cells adapt to the obesity-induced fatty acid enriched environment, an *in vitro* model for long-term adaptation to high concentrations of PA was established on both murine and human breast cancer cell lines. The transcriptomic profiles and phenotypes of this *in vitro* model is highly overlapped with human breast tumors. The phenotype of PA-adapted cells was further validated in *ex vivo* cells derived from both chow diet and HFD-fed mice. Since *ex vivo* cells are dissociated from fresh tumor tissues and cultured *in vitro* for a short-term, they can closely resemble *in vivo* tumor characteristics.

Choosing the appropriate cell lines is a key step to establish an *in vitro* model. Numerous breast cancer cell lines have been well-defined and available in commercial cell banks[200]. Due to the heterogeneity of breast cancer, no single cell line can truly represent the primary breast cancers[201]. Thus, to make a concrete conclusion, the results should be verified in multiple cell lines and living systems. Both human and murine cell lines used in paper I and II have been summarized in Table 1. In paper I, to investigate the molecular mechanisms in identified hormonal receptor negative breast cancer patient group, only TNBC cell lines were used. In paper II, due to the broad implications of anti-ETP antibody, both hormonal receptor positive and negative breast cancer cell lines were used to validate the effectiveness of the antibody.

Table 1. The overview of cell lines used in paper I and II.

Cell line	Source	Hormone receptor status
HCC1806	Human breast cancer cell line	Triple negative
MDA-MB-231	Human breast cancer cell line	Triple negative
MDA-MB-453	Human breast cancer cell line	ER-, PR-, HER2+
T47D	Human breast cancer cell line	ER+, PR+, HER2-
MCF-7	Human breast cancer cell line	ER+, PR+, HER2-
ZR-75-1	Human breast cancer cell line	ER+, PR +/-, HER2-
SKOV3	Human ovarian cancer cell line	-
HUVEC	Human umbilical vein endothelial cell line	-
E0771	Murine breast cancer cell line	Triple negative

Another major benefit of using *in vitro* models for cancer studies is that they offer different functional assays to investigate the properties of cancer cells. In paper I and II, both gain- and loss-of-function assays were performed to investigate the functions of specific genes. In addition, in paper I, the *in vitro* tumorsphere formation assay was adopted to evaluate the stem-like property of PA-adapted cells. Here, cancer cells were seeded in a serum-free and non-adherent condition, in which only stem/progenitor cells can survive and outgrow[202]. Following certain *in vitro* culture times, the number of formed tumorspheres can reflects the level of cancer stem-like traits. Thus, depending on the specific questions, the appropriate *in vitro* models can serve as a powerful research tool.

3.3 *In vivo* models

In vivo animal models have become the essential research tools for molecular mechanistic and translational studies in cancer biology. Cancer and obesity are both complex pathological conditions which induce both local and systemic alterations[173]. Despite the great improvements of *in vitro* models, such as conditional media and three-dimensional culture, these models still cannot replace animal models in cancer-obesity studies. In paper I, both breast cancer syngeneic mouse models and an HFD-induced

obese mouse model were used to verify the hypothesized mechanisms *in vivo.* Animal experiments were approved by the Norwegian Animal Research Authority (FOTS ID 24723). In paper II, the human cell-derived xenograft nude mouse model was used to investigate the effectiveness of targeted therapy on human breast cancer cells. This study was approved by the Institutional Animal Care and Research Advisory Committee at The University of Texas Southwestern Medical Center (Protocol Number 2015-101207-G).

3.3.1 *In vivo* models for obesity research

The genetic obese mouse and diet-induced obese (DIO) mouse are widely used in obesity related cancer studies. Among genetic models, obesity is most commonly caused by the disruption of leptin signaling, either by the lack of leptin (*ob/ob* mice) or by mutations in the leptin receptor gene (*db/db* mouse model). The deficient leptin signaling acts by interacting with central nervous system to increase food intake and result in an obese condition[203]. Although *ob/ob* and *db/db* mouse models display multiple obesity-induced comorbidities, the tumor developments observed in these models are controversial, and some of them are in conflict with epidemiological findings[204, 205]. This points out the possibility that leptin signaling is involved in cancer development in the obese condition. Additionally, leptin mutations are not common in humans, and human obesity is rarely caused by a monogenic trait[206]. Thus, genetic obese mouse models are less optimal for the mechanistic studies in obesity-cancer connection.

The DIO mouse models can better mimic most human obese cases. To establish DIO mouse models, the mice are given free access to the high calorie diets containing high proportion of fat, sugar or other ingredients. Among different feeding regimes, the most commonly used diet contains 30-60 kJ% energy from fat (known as, HFD). After a long-term (normally 10-12 weeks) exposure to HFD, mice gradually accumulate fat and develop obesity which is illustrated by increased body weight and obesity-induced comorbidities[207, 208]. However, not all the mouse strains are suitable for establishing DIO models, because different strains, even littermates (within the same cohort), may

respond to HFD feeding differently. For instance, Balb/c mice are resistant to DIO[209], and HFD feeding only has little effects on their body weight[210]. Due to the stably increased body weight and development of obesity-induced comorbidities after a long-term HFD exposure, C57BL/6J strain is the most commonly used DIO mouse strain[211]. In paper I, female C57BL/6J mice were fed with a HFD (60% kcal from fat, 20% from protein and 20% from carbohydrates, Research Diets, D12492) for 10 weeks prior to tumor cells implantation. Chow diet-fed (7.5% kcal from fat, 17.5% from proteins and 75% from carbohydrates, Special Diet Services RM1, 801151) littermates were used as controls. The established obese mice display significantly body weight gain, liver steatosis, hyperinsulinemia, hyperglycemia and reduced glucose clearance compared with the regular chow diet-fed mice (data is shown in paper I).

3.3.2 *In vivo* models for breast cancer research

In breast cancer studies, several mouse models have been established to mimic breast cancer progression *in vivo*. Generally, the models can be classified to four main groups, cell line-derived mouse models, patient-derived xenograft models (PDXs), genetically engineered mouse models (GEMMs) and environmentally induced cancer models[212]. Among these models, cell line-derived mouse model is the first established and employed by the majority of breast cancer studies. The advantage is that both tumorigenic human (xenograft) and murine cell lines (syngeneic) can be used in this model. In addition, it is relatively easy to study gene functions *in vivo* by manipulating the implanted cells. Moreover, the orthotopic injection of breast cancer cells into mammary fat pad is relatively straightforward and the tumor volumes can be easily monitored during the experimental period. It should be noted that the human breast cancer cell lines need to be studied in immunocompromised mice, which limits their use in tumor immunology related research. In paper I, given the important roles of the immune system in tumor progression, we decided to use immunocompetent mice. Moreover, as mentioned above, the obese model needs to be established in C57BL/6J mouse strain, which further limits the choices of appropriate cell lines. Therefore, the triple negative murine breast cancer E0771 cell line was used to establish the orthotropic breast cancer mouse model. In paper II, to investigate the effectiveness of neutralizing

ETP antibody on human breast cancer cells *in vivo*, two cell line-derived nude mouse models were adopted in the experiment.

3.3.3 Body weight, excess nutritions or obesity

As BMI is an important parameter to define human obesity, the body weight of animals is normally used to monitor the obese mouse model. However, several studies and our experiments have found that a small population of C57BL/6 mice are resistant to HFD-induced obesity and long-term HFD exposure fails to induce the increased body weight[213-215]. Additionally, there is a trend toward increased tumor incidence and tumor burden in the obesity-resistant mice compared with low-fat/chow diet-fed mice even though they have the same body weight gains. In another study, to investigate if the excess nutrition is important for cancer progression, the mice are exposed to HFD for four days prior to tumor challenge[180]. The short-term HFD feeding is able to induce a transient condition of nutrient excess but is not enough to induce a higher tumor incidence. Overall, the established state of obesity is required for affecting tumor outcomes. Therefore, in our study, a long HFD feeding period (10 weeks) was used, and in addition to the body weight gain and feeding period, the obesity-induced clinical and biochemical features had also been verified before tumor challenge.

3.4 Profiling chromatin regulatory landscape

The advancements of sequencing-based technologies in the recent years make it possible to capture the information on chromatin accessibility, histone modifications, chromatin 3D structures and regulatory factor bindings. The alterations at these levels have been shown to regulate the transcriptional capacity of a certain genetic locus[216]. Thus, the combination of epigenetic profiling technologies and transcriptomic analysis is able to depict a comprehensive chromatin regulatory landscape in breast cancer progression.

3.4.1 The Assay for Transposase-Accessible Chromatin with sequencing (ATAC-Seq)

Chromatin accessibility can be profiled by diverse approaches, such as DNase I hypersensitive site sequencing (DNase-seq), micrococcal nuclease sequencing (MNase-seq) and ATAC-seq. These approaches are all enzyme-based technologies, in which the chromatin accessibility is measured by quantifying the cleaved fragments of open chromatins (e.g., DNase-seq and ATAC-seq) or the fragments spanning single nucleosomes (e.g., MNase-seq). Due to the straightforward and efficient procedures, ATAC-seq has been adopted by the majority of studies of chromatin accessibility measurements in the recent years. Essentially, ATAC-seq significantly reduces the input cell number compared with other alternative assays, and the libraries can be generated with as few as 500 cells[217, 218]. ATAC-seq experiments are performed on purified nuclei. The accessible chromatin is randomly cleaved by a hyperactive Tn5 transposase and appended to sequencing adapters, which are preloaded in the transposase. The sequencing libraries are generated from purified fragmented DNA, and followed by next-generation sequencing. The locations and abundances of sequencing reads reflect the chromatin accessibility[219] (Figure 10).

Accessible chromatins provide the platform for transcription factors binding which regulates the expression of downstream genes. As the final readout, RNA-seq data are frequently integrated with ATAC-seq data to interpret the mechanisms of deregulated gene expression. The differentially accessible gene promoter regions are closely associated with the altered mRNA levels. In paper I, this combined analysis was adopted to predict the potential protein-DNA binding and activity of transcription factors. It was done by scanning the putative transcription factor binding sites on the accessible chromatin regions and linking it to the gene expression level[220]. However, it should be noted that the ATAC-seq coupled with RNA-seq analysis only provides indirect evidence of the interactions between chromatin and transcription factors. Therefore, in paper I, we used a direct protein-DNA interactions profiling strategies (discussed in Section 3.4.2) to validate our findings.

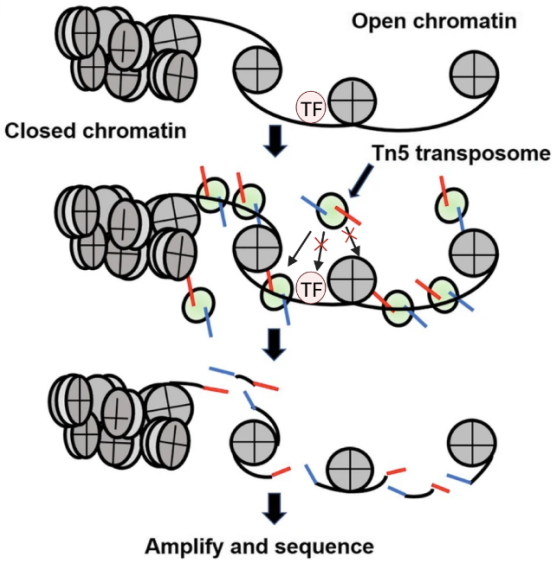


Figure 10. The workflow of ATAC-seq. The hyperactive Tn5 transposase preferentially fragments the unprotected chromatin regions and inserts sequencing adapters into it. The DNA regions occupied by nucleosome or transcription factors (TF) cannot be accessed by Tn5 transposase. The adapters-tagged DNA fragments are used to generate sequencing libraries. Figure adopted from [216] with modifications.

3.4.2 Cleavage Under Targets and Release Using Nuclease (CUT&RUN)

CUT&RUN is developed based on the chromatin immunoprecipitation followed by sequencing (ChIP-seq) assay which directly detects protein-chromatin interactions[221]. In CUT&RUN, unfixed permeabilized cells or nuclei are incubated with transcription factor-specific or histone modification-targeted antibodies, followed by the incubation of a protein A-MNase fusion protein (pA/MNase) and activation by calcium. The activated MNase cuts the DNA at both sides of the protein-DNA complexes which are further released into solution for subsequent analysis, while the remaining chromatin is trapped in the intact cells or nuclei[221] (Figure 11). This step dramatically reduces the background, thus only 1/10 of ChIP-seq sequencing depth is required for CUT&RUN samples[222]. In paper I, CUT&RUN was performed with both transcription factor C/EBPB and histone modification markers (H3K4me1 and

H3K27me3) specific antibodies to identify the mechanism of epigenetic activation of C/EBPB transcriptional activity in obese breast cancer cells.

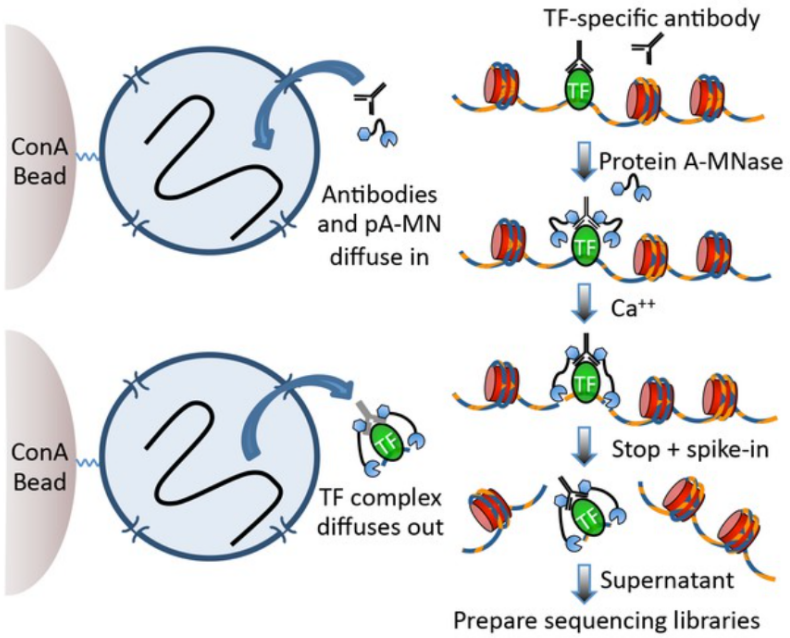


Figure 11. The workflow of CUT&RUN. Permeabilized cells or nuclei are attached to magnetic concanavalin A beads (ConA Bead) and incubated with transcription factor (TF)- or histone modification marker-specific antibodies and protein A-MNase fusion protein (pA-MN). MNase is activate by adding calcium (Ca^{++}) and cleaves chromatin. TF-DNA complex is then released into the supernatant. DNA is extracted from supernatant and used for preparing sequencing libraries. Figure adopted from [221].

4. Summary of results

Paper I

Title: "C/EBPB-dependent Adaptation to Palmitic Acid Promotes Tumor Formation in Hormone Receptor Negative Breast Cancer"

In Paper I, we identified a cancer cell-autonomous mechanism of obesity-induced breast cancer risk in postmenopausal patients. We found that in the DIO mouse model, the obese environment enhances tumor initiation compared with the chow diet-fed littermates. In human breast cancer patients, obesity was specifically associated with poor survival in postmenopausal patients with ER^-/PR^- tumors. In addition, the expression of human breast cancer stem-like markers, CD133 and CD44, were enhanced in tumor tissues from obese compared to non-obese postmenopausal ER^-/PR^- patients. Next, to investigate whether the obesity-induced cancer stem-like properties are a consequence of a selection or adaptation process, a cellular barcoding analysis was performed to monitor clonal dynamics during cancer progression in chow- and HFD-fed mice. The tumors derived from these mice displayed a similar barcode distribution, suggesting that exposure to an obese environment did not select for a pre-existing subclone. We therefore questioned what drives this cellular adaptation and what is the mechanism underlying this process. As PA is one of the predominant fatty acids in human circulation and dramatically elevated in obese individuals, we therefore established an *in vitro* model by a long-term adaptation to gradually increased concentrations of PA. By performing mass cytometry and *in vivo* extreme limiting dilution analysis, we found that *in vitro* adaptation model phenocopied obesity-induced stem-cell features. In addition, a mutual information-based transcriptomic data comparison analysis between patients' tumors and PA-adapted cells supported the phenotypic overlap at the patient level. High-coverage sequencing of 360 known cancer genes did not display any mutations correlating to obesity. We therefore speculated that the obesity-induced stem-like property could be driven by epigenetic alterations. Thus, an ATAC-seq experiment was performed to assess the chromatin accessibility. We found that the obese environment induced widespread chromatin remodeling and that

transcription factor C/EBPB was strongly associated with the more accessible chromatin in PA-adapted cells and E0771 *ex vivo* cells derived from HFD-fed mice. Consistently, the activating histone marker H3K4me1 and repressive marker H3K27me3 were respectively increased and decreased across the same chromatin regions in PA-adapted cells. Collectively, both ATAC-seq and CUT&RUN analysis suggested that the epigenetic activation of C/EBPB is critical for breast cancer development in obese conditions. To investigate the roles of C/EBPB in the tumor initiation phenotype, a series of *in vitro* and *in vivo* assays were performed using gain- and loss-of-function approaches. C/EBPB depletion inhibited tumorsphere formation and delayed the tumor initiation in obese mice, whereas overexpression of the C/EBPB isoform LAP2-promoted tumorsphere formation identified the functional relationship between C/EBPB and obesity-induced cancer stemness. We next wanted to determine the downstream signaling regulated by C/EBPB. By performing C/EBPB CUT&RUN and integrating the epigenetic and transcriptomic data from different *in vitro*/*in vivo* systems, we identified nine putative C/EBPB target genes. Among those genes, *Claudin 1 (CLDN1)* and *Lipocalin 2 (LCN2)* displayed similar expression patterns upon *C/EBPB* overexpression in both HCC1806 and MDA-MB-231 PA-adapted cells. We next functionally tested the roles of these two genes in tumor initiation capacity. Depletion of either *CLDN1* or *LCN2* significantly impaired tumorsphere formation *in vitro* as well as the tumor initiation capacity in HFD-fed mice, suggesting that *CLDN1* and *LCN2* are required for C/EBPB induced cancer stemness in the obese state. Taken together, long-term adaptation of breast cancer cells to an obesity-induced PA enriched environment promotes cancer stem-like property by epigenetically activating C/EBPB and stimulating the expression of C/EBPB downstream targets.

Paper II

Title: "Human endotrophin as a driver of malignant tumor growth"

In paper II, the major aim was to examine the basic physiological responses to ETP in human cells, to develop neutralizing antibodies against human ETP and to ultimately validate the effectiveness for potential clinical application. To this end, first, we

measured human plasma ETP levels and observed the higher circulating levels in breast cancer patients compared to healthy individuals. In addition, higher *COL6A3* (encoding ETP parental molecule) expression in primary tumors was associated with poor survival. We demonstrated that ETP exposure stimulated the induction of several EMT markers in both human cancer cell lines and mesothelial cells, indicating the potential EMT stimulatory function of ETP. To functionally test the chemoattractant properties, the *in vitro* scratch assay and transwell assay were performed on human endothelial cells, macrophages and several breast cancer cell lines. All cells displayed enhanced migration and invasion phenotypes upon ETP treatment. Especially, HUVEC endothelial cells displayed increased tube formation capacity upon ETP exposure, which supports an angiogenic role of ETP. We next generated and screened a panel of 132 monoclonal antibodies targeting human ETP. These antibodies were screened using cisplatin resistance and chemoattractant feature of ETP. In the panel, the ETPmAB4 antibody displayed the best performance, which completely neutralized ETP-mediated cisplatin resistance in MCF-7 cells and most effectively prevented macrophage migration. To examine the effectiveness of ETPmAB4 *in vivo*, ETP overexpressing MCF-7 cells were orthotopically implanted into nude mice. The ETPmAB4 treatment potently inhibited tumor growth and the induction of EMT signature was also impaired in tumor tissue. Moreover, the antibody inhibited the ETP-induced cisplatin resistance *in vivo*. These effects were further confirmed in MDA-MB-231 and ovarian cancer cell lines *in vivo*. In summary, we identified and described a novel ETP neutralizing antibody that effectively inhibit tumor growth and simulate cisplatin susceptibility on human cancer cells *in vitro* and *in vivo*.

5. Discussion

Paper I

The association between obesity and the increased postmenopausal breast cancer risk has been demonstrated in humans and pre-clinical models. However, the underlying mechanisms remain poorly understood. In paper I, we identified obesity-mediated stem cell-like properties in both DIO mouse model and human patients' materials. Such phenotype has been consistently observed on different cancer types, and diverse mechanisms have been proposed. The majority of studies have focused on the environmental factors, including altered adipokines (e.g. leptin and resistin) from dysfunctional adjacent adipocytes[151, 162], the elevated circulating FABP4[187, 188] and the altered immune landscape[180]. The diverse mechanisms reflect the complexity of an obese environment, meanwhile make it difficult to determine whether they are predominantly required for the obesity-associated tumorigenesis. Thus, in paper I, we sought to discover cell-autonomous mechanisms underlying the connection.

Over the past decades, numerous mutations in oncogenes and tumor suppressors, such as *PIK3CA, KRAS, TP53* and *PTEN*, have been extensively studied. The accumulation of mutations in somatic stem cells may lead to tumorigenic transformation[223, 224]. Although both diet and obese states have been proposed to affect genomic stability[223, 225, 226], due to the lack of direct evidence, it remains unclear how such exogenous stimuli induce cancer-related mutations, or if the driver mutations have already deposited in the cells before obesity was developed. In the obese state, the oxidative stress and chronic inflammation-caused DNA damage could be the potential inducers of genetic mutations[226]. We therefore hypothesized that the obesity-induced cancer stem-like properties could be dependent on obesity-induced genetic mutations. Unexpectedly, across a 360 known cancer genes panel, we did not identify any specific mutations associated with obesity in tumors from postmenopausal breast cancer patients. Consistently, another study in an obesity-associated pancreatic cancer mouse model demonstrated that the leptin-deficiency(*ob/ob*)-induced obese state drove pancreatic tumorigenesis in KC mice (*Pdx1-Cre; Kras*$^{LSL\text{-}G12D/WT}$). However, the

phenotype was not associated with any mutations in the known PDAC or pan-cancer driver genes as demonstrated by exome sequencing[227]. Additionally, the result from the *in vivo* barcode tracking study (Paper I Figure 2A) suggested an adaptation process for cancer cells to survive in an obese environment, which was unlikely driven by genetic alterations. Therefore, we focused our studies on epigenetic mechanisms rather than genetic mechanisms.

By systematically profiling the epigenetic landscape (ATAC-seq and CUT&RUN) in cancer cells, we found that the exposure to an obese environment caused extensive changes in chromatin accessibility. These findings parallel to what has been demonstrated in mouse liver tissue and colonic epithelium exposed to HFD-feeding[228-230]. Interestingly, all of these studies and our work identified significant metabolic changes, especially the changed lipid metabolism in obese settings. Lipid metabolism has been proposed to regulate chromatin accessibility through either changing the availability of key metabolites or triggering signaling pathways to affect histone modifications [231, 232]. Particularly, the lipid-derived acetyl-CoA, a central metabolite, connects different metabolic pathways as well as chromatin modifications[62, 231]. As such, obesity-induced elevated lipids do not simply serve as an energy source, it can also act as a selective molecule to affect chromatin remodeling.

Obesity-dependent chromatin remodeling appears to be associated with a stem cell like state compared with a published ATAC-seq data from different cell lineage states over the murine mammary gland development. Specifically, HFD-induced accessible chromatin regions enriched the binding motifs of transcription factor C/EBPB. As discussed in Section 1.2.4, *C/EBPB* expression is dramatically changed during mammary gland development[90] and is required to maintain the self-renewal of mammary stem/progenitor cells[93]. Several studies have also reported that the expression level of *C/EBPB* (or isoforms) is important for their regulatory functions and is normally associated with aggressive cancer phenotypes[95]. In addition, the overexpression of *C/EBPB* isoform *LAP2* induces oncogenic transformation of premalignant human breast epithelial cells[97]. In our study, we have demonstrated that C/EBPB is essential to promote cancer stem-like properties in obese states. However,

the key difference was that we did not observe the expression and cellular location changes of C/EBPB between obese and non-obese settings, reflecting that the activity of C/EBPB is likely regulated at the epigenetic level.

We further investigated C/EBPB downstream signaling in obesity-induced cancer stemness. By integrating epigenetic and transcriptomic profiles from both *in vivo* and *in vitro* models a panel of C/EBPB putative targets in breast cancer cells was identified. Interestingly, many of these genes are involved in ECM organization and remodeling. Obesity-driven changes in ECM components and stiffness have been extensively studied, which has also been proposed to link cancer progression in a few studies. For example, a non-cell-autonomous mechanism was proposed by Seo *et al*.[166]. Here, obesity-induced myofibroblasts enhanced the ECM stiffness in mammary adipose tissue in both DIO and *ob/ob* mouse models, which in turn promoted cancer cell growth and the tumorigenic potential of human MCF10A premalignant epithelial cells. Although our work did not address which cell types secrete ECM proteins and reconstruct the architectures, it has been shown previously that both cancer cells and stromal cells may contribute to the ECM remodeling[233]. In addition, tumor cell-intrinsic mechanisms have also been reported to drive ECM remodeling during metastatic progression[234], suggesting that the ECM remodeling-mediated cancer progression could be driven by both tumor cell-intrinsic and environmental factors.

In our study, *LCN2* and *CLDN1* were identified as two C/EBPB target genes. Both of them have been reported to be associated with a poor differentiated cancer phenotype[235, 236]. LCN2 is a secretory protein and produced by multiple organs. In obese individuals, LCN2 is abundantly generated by adipose tissue and involved in many biological processes, such as apoptosis and innate immunity[237]. Since its circulation concentration is sharply increased in obese state, LCN2 has been used as a marker for obesity and its associated pathologies[161]. In several cancer types, *LCN2* overexpression has been consistently observed, and it is frequently associated with advanced cancer stages and poor prognosis[235]. Mechanistic studies have revealed that LCN2-induced EMT[235] and angiogenesis[238] play important roles in tumor progression. Additionally, LCN2-promoted tumor formation has been reported in a

transgenic MMTV-PyMT mouse model[239], in which the significantly increased tumor incidence was observed in LCN2$^{+/+}$ PyMT mouse and the increased MMP9 activity was linked to this phenotype[239, 240]. However, more experiments are required to reveal the underlying mechanisms. CLDN1, a tight junction protein, is widely expressed in various cancer types and play important roles in cell proliferation and differentiation[241, 242]. Importantly, CLDN1 and its family members have been reported to induce cancer cell dedifferentiation through EMT or YAP signaling[241, 243]. In the present study, by performing loss-of-function assays in both *in vitro* and *in vivo* models, we demonstrated the essentialness of LCN2 and CLDN1 in obesity-induced cancer stem trait. Additionally, knockdown of these two genes epistatically rescued the *C/EBPB*-enhanced tumor initiation, which further confirmed that they are C/EBPB downstream targets in this signaling.

In sum, we demonstrate that cellular adaptation to the obese environment contributes to dedifferentiation of breast cancer cells towards a stem-like state, and that C/EBPB is a master regulator in this process. Our study provides a novel insight into the link between obesity and breast cancer, which could form the basis to design targeted therapy for obese breast cancer patients.

Paper II

As introduced in Section 1.3.2, ETP is a carboxy-terminal cleavage product of COL6A3. In preclinical models, ETP has been identified as an obesity-associated adipokine which is abundantly generated and released from adipocytes. ETP levels are significantly increased in both DIO and genetic obese mouse models as well as human tumor tissues[167, 171, 244]. Mechanistically, ETP is involved in the regulation of metabolic homeostasis[244] and promotes malignant tumor progression[167, 171], suggesting it could be a key factor linking obesity and cancer progression.

Building on the previous studies regarding the increased ETP levels in tumor tissues, we first examined the circulating levels of ETP in breast cancer patients. Elevated plasma ETP level was observed in breast cancer patients compared with healthy individuals. However, in a study by Willumsen *et al*., the PRO-C6 (known as ETP)

level was not significantly changed in serum of breast cancer patients compared with healthy individuals[245], which was likely caused by the small sample size (13 healthy controls and 8 breast cancer patients) and a notable variances between patients. Importantly, a trend of elevated level of PRO-C6 was observed in late-stage cancer patients in their study[245], suggesting a predictive potential of ETP. In line with previous observations in animal models, we demonstrate that human breast tumor tissues express high levels of *COL6A3*. Moreover, the high *COL6A3* expression is associated with poor survival in breast cancer patients. As the parental molecule of ETP, COL6A3 may reflect the ETP level *in vivo*, however, direct measures in larger patient cohorts are still required to investigate the diagnostic and predictive potential of ETP in breast cancer patients.

To translate previous findings from mouse models to clinical applications, we generated human recombinant ETP and a bank of novel antibodies targeting human ETP. By performing a series of *in vitro* assays, we verified that ETP acts as an EMT stimulator on human cell lines and functionally promotes cancer cell migration and invasion. It also works on human endothelial cells to enhance tube formation capacity, reflecting the pro-angiogenic potential in tumor progression. All the above-mentioned observations comply with the functions of ETP observed in murine mammary tumor development[167, 171]. Additionally, in two human cancer cell line-derived xenograft mouse models, ETP neutralizing antibody effectively inhibits tumor growth and cisplatin resistance.

Several recent studies have reported the measurements of serum ETP concentration in humans. The increased ETP levels are frequently associated with severe pathological states and metabolic diseases[245-247]. Additionally, its parental molecule COL6A3 also contributes to numerous metabolic dysfunctions in obese state, such as reduced lipid clearance rate and impaired insulin signaling[248]. Due to the limitation of mouse models, our study did not investigate the effectiveness of ETP antibody in an obese cancer mouse model. However, the reported results combined with existing evidence suggest a clinical application potential of the ETP antibody in obese breast cancer patients.

Limitations of study

In paper I, we suggested a C/EBPB-regulated cell-autonomous mechanism promoting cancer initiation in obese environment. Given the complexity of obese environment, the phenotype is unlikely regulated by a single mechanism. Although we examined tumor proliferation and metabolic behaviors throughout the entire study, it cannot exclude the contribution of previous reported mechanisms, such as inflammation and altered adipokines, to this connection.

The obese model in this study was established by exposing animals to an extreme high fat content diet (60% kcal energy from fat) - a model has been widely used in cancer-obesity research[208]. However, such high fat content cannot represent the majority of human diets, including obesogenic Western diet (30-40% kcal energy from fat) [249]. For this reason, the diet with lower fat content should be considered in obese mouse model. In addition to the fat content, fat sources and other components of diet may independently affect tumor outcomes[250]. With this in mind, genetic obese models, such as *ob/ob* mouse, could serve as a control to exclude the effects of dietary compositions.

According to the results from a high-coverage sequencing of 360 known breast cancer related genes and a barcode tracing analysis, we did not identify any notable genetic alterations under the challenge of obese environment. Here, a whole-genome sequencing analysis would have provided more comprehensive information about the cancer genetic alterations in an obese environment.

Our study demonstrated the essential roles of the C/EBPB target genes, *LCN2* and *CLDN1*, in obesity-induced tumor formation capacity. In addition, the pathway analysis pointed out C/EBPB targets contributing ECM remodeling. However, we did not systematically study the functions of LCN2 and CLDN1 in ECM remodeling. Further functional assays are required to investigate how those genes participate in ECM remodeling to promote the cancer stem-like phenotype in an obese condition.

In addition, the expression and cellular distribution of C/EBPB was not changed in obese settings, suggesting that PA regulated C/EBPB activation at an epigenetic level. Of note, we did not focus on the upstream regulators of C/EBPB. Particularly, how PA regulates chromatin remodeling and which factors decide the preferences of PA-regulated chromatin regions are still unknown.

In paper II, we present the human ETP sequence that contains two predicted MMP14 cleavage sites. However, our study did not provide direct evidence to address which MMPs specifically cleave ETP from COL6A3. By preforming the survival analysis, we showed that the expression of *COL6A3* is associated with poor survival in breast cancer patients. As the parental molecule of ETP, the association suggests the diagnostic and predictive roles of ETP in breast cancer patients' survival. However, further studies are still needed to directly prove the association between ETP and patients' survival.

Last, a tube formation assay was performed in *in vitro* HUVEC cells and the results reflected a pro-angiogenic role of ETP. However, a systematic measurement of blood vessel formation and capillary density in tumor samples could have been performed for a direct support of ETP-induced angiogenesis.

6. Conclusion

Emerging evidence have demonstrated that cancer cells exposed to obese environment display altered phenotypes. Our discoveries extend on the current knowledge of how breast cancer cells adapt to obese environments and highlight the translational potential of these findings.

In paper I, we showed that obesity promoted a cancer stem-like phenotype in both animal models and human breast cancer patients. The long-term exposure to an obese environment induced widespread chromatin remodeling in breast cancer cells. Specifically, enrichment of the C/EBPB binding motifs in the opened chromatin regions suggested the regulator roles of C/EBPB in this process. According to the epigenetic profiles and functional studies, we identified essential roles of the C/EBPB LAP2 isoform and its target genes in obesity-induced cancer stemness.

In paper II, we investigate the biological functions of COL6A3 cleaved product ETP in human cell lines and preclinical models. ETP-induced pro-fibrotic, chemoattractant and pro-angiogenic properties have been widely studied in murine cells. In our study, we confirmed these ETP-induced properties in human cells. Accordingly, we generated a panel of neutralizing human ETP antibodies and validated their effectiveness in human cell line-derived *in vivo* models. The antibody displayed potent tumor inhibition and anti-cisplatin resistance effects in these models, suggesting its clinical potential in obese breast cancer patient treatment.

In summary, the present studies provide a novel insight into the link between obesity and breast cancer stemness as well as build up the translational potential of existing mechanisms. Our findings could ultimately be exploited to design the personalized treatments for obese breast cancer patients.

7. Future perspectives

Over the past decades, emerging studies have uncovered the relationship between obesity and cancer risk. Most of our knowledge come from epidemiological studies, *in vitro* cell lines as well as tumor materials from human or preclinical animal models. Given the complexity of obese environment, the studies performed on single cell lines or bulk materials normally cannot recapitulate the complexity of a tumor microenvironment - especially the cell-stroma and cell-cell interactions. Advanced sequencing technologies and imaging-based single cell multiomics approaches provide opportunities to address this problem. By preforming single cell level analysis, several recent studies have proposed in-depth mechanisms to explain the altered cancer development in obese states. The study by Ringel *et al*. preformed single-cell RNA-seq to investigate the tumor-immune cell interactions in an obese tumor microenvironment. They revealed the obesity-induced distinct metabolic adaptations in cancer cells and immune cells[190]. By integrating single-cell and spatial transcriptomics analysis, Wu *et al*. presented the high-resolution cellular landscape of human breast tumors, which help us better understand breast cancer inter/intracellular heterogeneity, cancer stromal-immune niche and cancer cell differentiation lineages[251]. In our studies, the inclusion of single cell level and spatial analysis would have helped us draw a detailed picture showing cell-cell interactions in obese environment. Particularly, it would be better understood how C/EBPB targets shape the ECM to maintain cancer stem-like properties and how stromal cells affect cancer cell fate decision in an obese environment. Additionally, these approaches may provide valuable information about the origin of ETP and the ETP-mediated crosstalk between adipocytes and cancer cells.

In addition to spatial heterogeneity, both obesity and cancer development are dynamic states. The temporal alterations add another level of complexity into understanding the connection between obesity and cancer progression. The high-resolution spatial analysis on tumor tissues could capture the dynamic differences in a limited degree. Given the huge time spans of cancer development, from initiation to advanced late stages, the issue would still exist. Several studies have tried to collect samples from specific time points or measure certain parameters throughout the tumor development

to solve this problem[187, 252]. In paper I, since increased tumor initiation was our main cancer phenotype, most materials were collected from early-stage tumors. Still, the comparison between obese/non-obese tumors from different time points would help us better illustrate the dynamic adaptation of cancer cells in an obese environment. Importantly, in paper II, this kind of information will directly answer how early the elevated ETP levels can be detected during tumor progression, which could strongly support the diagnostic potential of ETP during early tumor development.

References

1. WHO. *Obesity and overweight*. [cited 2021 27/09]; Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.
2. Gonzalez-Muniesa, P., et al., *Obesity.* Nat Rev Dis Primers, 2017. **3**: p. 17034.
3. Romieu, I., et al., *Energy balance and obesity: what are the main drivers?* Cancer Causes Control, 2017. **28**(3): p. 247-258.
4. Rothman, K.J., *BMI-related errors in the measurement of obesity.* Int J Obes (Lond), 2008. **32 Suppl 3**: p. S56-9.
5. Kushner, R.F. and D.H. Ryan, *Assessment and lifestyle management of patients with obesity: clinical recommendations from systematic reviews.* JAMA, 2014. **312**(9): p. 943-52.
6. Organization, W.H., *Waist Circumference and Waist-hip Ratio: Report of a WHO Expert Consultation, Geneva, 8-11 December 2008*. 2011: World Health Organization.
7. Cornier, M.A., et al., *Prevention of overweight/obesity as a strategy to optimize cardiovascular health.* Circulation, 2011. **124**(7): p. 840-50.
8. Goodwin, P.J. and V. Stambolic, *Impact of the obesity epidemic on cancer.* Annu Rev Med, 2015. **66**: p. 281-96.
9. Cusi, K., *Role of obesity and lipotoxicity in the development of nonalcoholic steatohepatitis: pathophysiology and clinical implications.* Gastroenterology, 2012. **142**(4): p. 711-725 e6.
10. Petersen, M.C. and G.I. Shulman, *Roles of Diacylglycerols and Ceramides in Hepatic Insulin Resistance.* Trends Pharmacol Sci, 2017. **38**(7): p. 649-665.
11. Adams, J.M., 2nd, et al., *Ceramide content is increased in skeletal muscle from obese insulin-resistant humans.* Diabetes, 2004. **53**(1): p. 25-31.
12. Robertson, R.P., et al., *Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection.* Diabetes, 2003. **52**(3): p. 581-7.
13. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA Cancer J Clin, 2021. **71**(3): p. 209-249.
14. Ginsburg, O., et al., *The global burden of women's cancers: a grand challenge in global health.* Lancet, 2017. **389**(10071): p. 847-860.
15. Malhotra, G.K., et al., *Histological, molecular and functional subtypes of breast cancers.* Cancer Biol Ther, 2010. **10**(10): p. 955-60.
16. Sinn, H.P. and H. Kreipe, *A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition.* Breast Care (Basel), 2013. **8**(2): p. 149-54.
17. Weigelt, B., F.C. Geyer, and J.S. Reis-Filho, *Histological types of breast cancer: how special are they?* Mol Oncol, 2010. **4**(3): p. 192-208.
18. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
19. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10869-74.
20. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8418-23.
21. Parker, J.S., et al., *Supervised risk predictor of breast cancer based on intrinsic subtypes.* J Clin Oncol, 2009. **27**(8): p. 1160-7.
22. Prat, A., et al., *PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer.* Breast Cancer Res Treat, 2012. **135**(1): p. 301-6.
23. Cheang, M.C., et al., *Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer.* J Natl Cancer Inst, 2009. **101**(10): p. 736-50.
24. Goldhirsch, A., et al., *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, 2011. **22**(8): p. 1736-47.
25. Harbeck, N., et al., *Breast cancer.* Nat Rev Dis Primers, 2019. **5**(1): p. 66.
26. Foulkes, W.D., I.E. Smith, and J.S. Reis-Filho, *Triple-negative breast cancer.* N Engl J Med, 2010. **363**(20): p. 1938-48.
27. Macias, H. and L. Hinck, *Mammary gland development.* Wiley Interdiscip Rev Dev Biol, 2012. **1**(4): p. 533-57.
28. Visvader, J.E. and J. Stingl, *Mammary stem cells and the differentiation hierarchy: current status and perspectives.* Genes Dev, 2014. **28**(11): p. 1143-58.

29. Visvader, J.E., *Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis.* Genes Dev, 2009. **23**(22): p. 2563-77.
30. Rios, A.C., et al., *In situ identification of bipotent stem cells in the mammary gland.* Nature, 2014. **506**(7488): p. 322-7.
31. Li, Y., W.P. Hively, and H.E. Varmus, *Use of MMTV-Wnt-1 transgenic mice for studying the genetic basis of breast cancer.* Oncogene, 2000. **19**(8): p. 1002-9.
32. Chu, E.Y., et al., *Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis.* Development, 2004. **131**(19): p. 4819-29.
33. Fu, N.Y., et al., *Stem Cells and the Differentiation Hierarchy in Mammary Gland Development.* Physiol Rev, 2020. **100**(2): p. 489-523.
34. Watson, C.J. and W.T. Khaled, *Mammary development in the embryo and adult: a journey of morphogenesis and commitment.* Development, 2008. **135**(6): p. 995-1003.
35. Bloushtain-Qimron, N., et al., *Cell type-specific DNA methylation patterns in the human breast.* Proc Natl Acad Sci U S A, 2008. **105**(37): p. 14076-81.
36. Holliday, H., et al., *Epigenomics of mammary gland development.* Breast Cancer Res, 2018. **20**(1): p. 100.
37. Pal, B., et al., *Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2.* Cell Rep, 2013. **3**(2): p. 411-26.
38. Polyak, K., *Breast cancer: origins and evolution.* J Clin Invest, 2007. **117**(11): p. 3155-63.
39. Prat, A., et al., *Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer.* Breast Cancer Res, 2010. **12**(5): p. R68.
40. Lim, E., et al., *Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers.* Nat Med, 2009. **15**(8): p. 907-13.
41. Shiovitz, S. and L.A. Korde, *Genetics of breast cancer: a topic in evolution.* Ann Oncol, 2015. **26**(7): p. 1291-9.
42. King, M.C., et al., *Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2.* Science, 2003. **302**(5645): p. 643-6.
43. Force, U.S.P.S.T., et al., *Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer: US Preventive Services Task Force Recommendation Statement.* JAMA, 2019. **322**(7): p. 652-665.
44. Liu, S., et al., *BRCA1 regulates human mammary stem/progenitor cell fate.* Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1680-5.
45. Molyneux, G., et al., *BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells.* Cell Stem Cell, 2010. **7**(3): p. 403-17.
46. Consortium, I.T.P.-C.A.o.W.G., *Pan-cancer analysis of whole genomes.* Nature, 2020. **578**(7793): p. 82-93.
47. Feng, Y., et al., *Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis.* Genes Dis, 2018. **5**(2): p. 77-106.
48. Zhan, T., N. Rindtorff, and M. Boutros, *Wnt signaling in cancer.* Oncogene, 2017. **36**(11): p. 1461-1473.
49. Liu, B.Y., et al., *The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4158-63.
50. Khorasanizadeh, S., *The nucleosome: from genomic organization to genomic regulation.* Cell, 2004. **116**(2): p. 259-72.
51. Holliday, R., *The inheritance of epigenetic defects.* Science, 1987. **238**(4824): p. 163-70.
52. Esteller, M., *Epigenetics in cancer.* N Engl J Med, 2008. **358**(11): p. 1148-59.
53. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. **31**(1): p. 27-36.
54. Flavahan, W.A., E. Gaskell, and B.E. Bernstein, *Epigenetic plasticity and the hallmarks of cancer.* Science, 2017. **357**(6348).
55. Feinberg, A.P. and B. Tycko, *The history of cancer epigenetics.* Nat Rev Cancer, 2004. **4**(2): p. 143-53.
56. Ehrlich, M., *DNA hypermethylation in disease: mechanisms and clinical relevance.* Epigenetics, 2019. **14**(12): p. 1141-1163.
57. Esteller, M., et al., *Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors.* J Natl Cancer Inst, 2000. **92**(7): p. 564-9.
58. Glodzik, D., et al., *Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers.* Nat Commun, 2020. **11**(1): p. 3747.
59. Lonning, P.E., et al., *White Blood Cell BRCA1 Promoter Methylation Status and Ovarian Cancer Risk.* Ann Intern Med, 2018. **168**(5): p. 326-334.

60. Eden, A., et al., *Chromosomal instability and tumors promoted by DNA hypomethylation.* Science, 2003. **300**(5618): p. 455.
61. Reid, M.A., Z. Dai, and J.W. Locasale, *The impact of cellular metabolism on chromatin dynamics and epigenetics.* Nat Cell Biol, 2017. **19**(11): p. 1298-1306.
62. Kinnaird, A., et al., *Metabolic control of epigenetics in cancer.* Nat Rev Cancer, 2016. **16**(11): p. 694-707.
63. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
64. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
65. Elsheikh, S.E., et al., *Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome.* Cancer Res, 2009. **69**(9): p. 3802-9.
66. Zhao, Y. and B.A. Garcia, *Comprehensive Catalog of Currently Documented Histone Modifications.* Cold Spring Harb Perspect Biol, 2015. **7**(9): p. a025064.
67. Currais, A., et al., *Elevating acetyl-CoA levels reduces aspects of brain aging.* Elife, 2019. **8**.
68. Takahashi, H., et al., *Nucleocytosolic acetyl-coenzyme a synthetase is required for histone acetylation and global transcription.* Mol Cell, 2006. **23**(2): p. 207-17.
69. Li, X., et al., *Regulation of chromatin and gene expression by metabolic enzymes and metabolites.* Nat Rev Mol Cell Biol, 2018. **19**(9): p. 563-578.
70. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome - biological and translational implications.* Nat Rev Cancer, 2011. **11**(10): p. 726-34.
71. Eckschlager, T., et al., *Histone Deacetylase Inhibitors as Anticancer Drugs.* Int J Mol Sci, 2017. **18**(7).
72. Wang, G.G., C.D. Allis, and P. Chi, *Chromatin remodeling and cancer, Part II: ATP-dependent chromatin remodeling.* Trends Mol Med, 2007. **13**(9): p. 373-80.
73. Clapier, C.R., et al., *Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes.* Nat Rev Mol Cell Biol, 2017. **18**(7): p. 407-422.
74. Bracken, A.P., G.L. Brien, and C.P. Verrijzer, *Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer.* Genes Dev, 2019. **33**(15-16): p. 936-959.
75. Wilson, B.G. and C.W. Roberts, *SWI/SNF nucleosome remodellers and cancer.* Nat Rev Cancer, 2011. **11**(7): p. 481-92.
76. Lai, A.Y. and P.A. Wade, *Cancer biology and NuRD: a multifaceted chromatin remodelling complex.* Nat Rev Cancer, 2011. **11**(8): p. 588-96.
77. Xu, G., et al., *ARID1A determines luminal identity and therapeutic response in estrogen-receptor-positive breast cancer.* Nat Genet, 2020. **52**(2): p. 198-207.
78. Jang, K.S., et al., *MTA1 overexpression correlates significantly with tumor grade and angiogenesis in human breast cancers.* Cancer Sci, 2006. **97**(5): p. 374-9.
79. Mazumdar, A., et al., *Transcriptional repression of oestrogen receptor by metastasis-associated protein 1 corepressor.* Nat Cell Biol, 2001. **3**(1): p. 30-7.
80. Si, W., et al., *Dysfunction of the Reciprocal Feedback Loop between GATA3- and ZEB2-Nucleated Repression Programs Contributes to Breast Cancer Metastasis.* Cancer Cell, 2015. **27**(6): p. 822-36.
81. Fujita, N., et al., *MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer.* Cell, 2003. **113**(2): p. 207-19.
82. Education, N. *transcription factor / transcription factors*. [cited 2021 27/09]; Available from: https://www.nature.com/scitable/definition/transcription-factor-167/.
83. Yang, L., et al., *Targeting cancer stem cell pathways for cancer therapy.* Signal Transduct Target Ther, 2020. **5**(1): p. 8.
84. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
85. Ponti, D., et al., *Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties.* Cancer Res, 2005. **65**(13): p. 5506-11.
86. Rodriguez-Pinilla, S.M., et al., *Sox2: a possible driver of the basal-like phenotype in sporadic breast cancer.* Mod Pathol, 2007. **20**(4): p. 474-81.
87. Descombes, P. and U. Schibler, *A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA.* Cell, 1991. **67**(3): p. 569-79.
88. Zahnow, C.A., *CCAAT/enhancer-binding protein beta: its role in breast cancer and associations with receptor tyrosine kinases.* Expert Rev Mol Med, 2009. **11**: p. e12.

89. Wethmar, K., et al., *C/EBPbetaDeltauORF mice--a genetic model for uORF-mediated translational control in mammals.* Genes Dev, 2010. **24**(1): p. 15-20.
90. Grimm, S.L. and J.M. Rosen, *The role of C/EBPbeta in mammary gland development and breast cancer.* J Mammary Gland Biol Neoplasia, 2003. **8**(2): p. 191-204.
91. Robinson, G.W., et al., *The C/EBPbeta transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland.* Genes Dev, 1998. **12**(12): p. 1907-16.
92. Grimm, S.L., et al., *Disruption of steroid and prolactin receptor patterning in the mammary gland correlates with a block in lobuloalveolar development.* Mol Endocrinol, 2002. **16**(12): p. 2675-91.
93. LaMarca, H.L., et al., *CCAAT/enhancer binding protein beta regulates stem cell activity and specifies luminal cell fate in the mammary gland.* Stem Cells, 2010. **28**(3): p. 535-44.
94. Vegesna, V., et al., *C/EBP-beta, C/EBP-delta, PU.1, AML1 genes: mutational analysis in 381 samples of hematopoietic and solid malignancies.* Leuk Res, 2002. **26**(5): p. 451-7.
95. Li, W., et al., *Aerobic Glycolysis Controls Myeloid-Derived Suppressor Cells and Tumor Immunity via a Specific CEBPB Isoform in Triple-Negative Breast Cancer.* Cell Metab, 2018. **28**(1): p. 87-103 e6.
96. Wang, S., et al., *C/EBPbeta regulates the JAK/STAT signaling pathway in triple-negative breast cancer.* FEBS Open Bio, 2021. **11**(4): p. 1250-1258.
97. Bundy, L.M. and L. Sealy, *CCAAT/enhancer binding protein beta (C/EBPbeta)-2 transforms normal mammary epithelial cells and induces epithelial to mesenchymal transition in culture.* Oncogene, 2003. **22**(6): p. 869-83.
98. Damiano, L., et al., *Oncogenic targeting of BRM drives malignancy through C/EBPbeta-dependent induction of alpha5 integrin.* Oncogene, 2014. **33**(19): p. 2441-53.
99. Jozwik, K.M. and J.S. Carroll, *Pioneer factors in hormone-dependent cancers.* Nat Rev Cancer, 2012. **12**(6): p. 381-5.
100. Cirillo, L.A. and K.S. Zaret, *An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA.* Mol Cell, 1999. **4**(6): p. 961-9.
101. Zaret, K.S. and J.S. Carroll, *Pioneer transcription factors: establishing competence for gene expression.* Genes Dev, 2011. **25**(21): p. 2227-41.
102. Swinstead, E.E., et al., *Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective: Multiple transcription factors can effect chromatin pioneer functions through dynamic interactions with ATP-dependent chromatin remodeling factors.* Bioessays, 2016. **38**(11): p. 1150-1157.
103. Hoogenkamp, M., et al., *Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program.* Blood, 2009. **114**(2): p. 299-309.
104. Hurtado, A., et al., *FOXA1 is a key determinant of estrogen receptor function and endocrine response.* Nat Genet, 2011. **43**(1): p. 27-33.
105. Serandour, A.A., et al., *Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers.* Genome Res, 2011. **21**(4): p. 555-65.
106. Eeckhoute, J., et al., *Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer.* Cancer Res, 2007. **67**(13): p. 6477-83.
107. Polyak, K., *Heterogeneity in breast cancer.* J Clin Invest, 2011. **121**(10): p. 3786-8.
108. Yates, L.R., et al., *Subclonal diversification of primary breast cancer revealed by multiregion sequencing.* Nat Med, 2015. **21**(7): p. 751-9.
109. Dagogo-Jack, I. and A.T. Shaw, *Tumour heterogeneity and resistance to cancer therapies.* Nat Rev Clin Oncol, 2018. **15**(2): p. 81-94.
110. Ellsworth, R.E., et al., *Molecular heterogeneity in breast cancer: State of the science and implications for patient care.* Semin Cell Dev Biol, 2017. **64**: p. 65-72.
111. Bell, C.C. and O. Gilan, *Principles and mechanisms of non-genetic resistance in cancer.* Br J Cancer, 2020. **122**(4): p. 465-472.
112. Huang, S., *Tumor progression: chance and necessity in Darwinian and Lamarckian somatic (mutationless) evolution.* Prog Biophys Mol Biol, 2012. **110**(1): p. 69-86.
113. Patch, A.M., et al., *Whole-genome characterization of chemoresistant ovarian cancer.* Nature, 2015. **521**(7553): p. 489-94.
114. Yip, S., et al., *MSH6 mutations arise in glioblastomas during temozolomide therapy and mediate temozolomide resistance.* Clin Cancer Res, 2009. **15**(14): p. 4622-9.
115. Sokol, E.S., et al., *Loss of function of NF1 is a mechanism of acquired resistance to endocrine therapy in lobular breast cancer.* Ann Oncol, 2019. **30**(1): p. 115-123.

116. Shu, S., et al., *Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer.* Nature, 2016. **529**(7586): p. 413-417.
117. Su, Y., et al., *Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance.* Proc Natl Acad Sci U S A, 2017. **114**(52): p. 13679-13684.
118. Pisco, A.O., et al., *Non-Darwinian dynamics in therapy-induced cancer drug resistance.* Nat Commun, 2013. **4**: p. 2467.
119. Xiao, T., et al., *Estrogen-regulated feedback loop limits the efficacy of estrogen receptor-targeted breast cancer therapy.* Proc Natl Acad Sci U S A, 2018. **115**(31): p. 7869-7878.
120. Rathert, P., et al., *Transcriptional plasticity promotes primary and acquired resistance to BET inhibition.* Nature, 2015. **525**(7570): p. 543-547.
121. Sun, C., et al., *Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma.* Nature, 2014. **508**(7494): p. 118-22.
122. Sharma, S.V., et al., *A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations.* Cell, 2010. **141**(1): p. 69-80.
123. Kim, C., et al., *Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing.* Cell, 2018. **173**(4): p. 879-893 e13.
124. Waks, A.G. and E.P. Winer, *Breast Cancer Treatment: A Review.* JAMA, 2019. **321**(3): p. 288-300.
125. Gianni, L., et al., *5-year analysis of neoadjuvant pertuzumab and trastuzumab in patients with locally advanced, inflammatory, or early-stage HER2-positive breast cancer (NeoSphere): a multicentre, open-label, phase 2 randomised trial.* Lancet Oncol, 2016. **17**(6): p. 791-800.
126. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence.* Clin Cancer Res, 2007. **13**(15 Pt 1): p. 4429-34.
127. Chaffer, C.L. and R.A. Weinberg, *A perspective on cancer cell metastasis.* Science, 2011. **331**(6024): p. 1559-64.
128. Tong, C.W.S., et al., *Recent Advances in the Treatment of Breast Cancer.* Front Oncol, 2018. **8**: p. 227.
129. Finn, R.S., et al., *Palbociclib and Letrozole in Advanced Breast Cancer.* N Engl J Med, 2016. **375**(20): p. 1925-1936.
130. Cristofanilli, M., et al., *Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial.* Lancet Oncol, 2016. **17**(4): p. 425-439.
131. Adams, S., et al., *Current Landscape of Immunotherapy in Breast Cancer: A Review.* JAMA Oncol, 2019. **5**(8): p. 1205-1214.
132. Fuentes-Antras, J., et al., *Adoptive Cell Therapy in Breast Cancer: A Current Perspective of Next-Generation Medicine.* Front Oncol, 2020. **10**: p. 605633.
133. Lauby-Secretan, B., et al., *Body Fatness and Cancer--Viewpoint of the IARC Working Group.* N Engl J Med, 2016. **375**(8): p. 794-8.
134. Calle, E.E. and R. Kaaks, *Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms.* Nat Rev Cancer, 2004. **4**(8): p. 579-91.
135. Calle, E.E., et al., *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults.* N Engl J Med, 2003. **348**(17): p. 1625-38.
136. van den Brandt, P.A., et al., *Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk.* Am J Epidemiol, 2000. **152**(6): p. 514-27.
137. Munsell, M.F., et al., *Body mass index and breast cancer risk according to postmenopausal estrogen-progestin use and hormone receptor status.* Epidemiol Rev, 2014. **36**: p. 114-36.
138. Renehan, A.G., et al., *Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies.* Lancet, 2008. **371**(9612): p. 569-78.
139. Sun, L., et al., *Body mass index and prognosis of breast cancer: An analysis by menstruation status when breast cancer diagnosis.* Medicine (Baltimore), 2018. **97**(26): p. e11220.
140. Hirko, K.A., et al., *Erythrocyte membrane fatty acids and breast cancer risk: a prospective analysis in the nurses' health study II.* Int J Cancer, 2018. **142**(6): p. 1116-1129.
141. Saadatian-Elahi, M., et al., *Biomarkers of dietary fatty acid intake and the risk of breast cancer: a meta-analysis.* Int J Cancer, 2004. **111**(4): p. 584-91.
142. Friedman, J.M., *Leptin and the endocrine control of energy balance.* Nat Metab, 2019. **1**(8): p. 754-764.
143. Considine, R.V., et al., *Serum immunoreactive-leptin concentrations in normal-weight and obese humans.* N Engl J Med, 1996. **334**(5): p. 292-5.

144. Frederich, R.C., et al., *Leptin levels reflect body lipid content in mice: evidence for diet-induced resistance to leptin action.* Nat Med, 1995. **1**(12): p. 1311-4.
145. Wu, M.H., et al., *Circulating levels of leptin, adiposity and breast cancer risk.* Br J Cancer, 2009. **100**(4): p. 578-82.
146. Khabaz, M.N., et al., *Immunohistochemical staining of leptin is associated with grade, stage, lymph node involvement, recurrence, and hormone receptor phenotypes in breast cancer.* BMC Womens Health, 2017. **17**(1): p. 105.
147. Garofalo, C., et al., *Increased expression of leptin and the leptin receptor as a marker of breast cancer progression: possible role of obesity-related stimuli.* Clin Cancer Res, 2006. **12**(5): p. 1447-53.
148. Hu, X., et al., *Leptin--a growth factor in normal and malignant breast cells and for normal mammary gland development.* J Natl Cancer Inst, 2002. **94**(22): p. 1704-11.
149. Yin, N., et al., *Molecular mechanisms involved in the growth stimulation of breast cancer cells by leptin.* Cancer Res, 2004. **64**(16): p. 5870-5.
150. Chang, C.C., et al., *Leptin-STAT3-G9a Signaling Promotes Obesity-Mediated Breast Cancer Progression.* Cancer Res, 2015. **75**(11): p. 2375-2386.
151. Bowers, L.W., et al., *Leptin Signaling Mediates Obesity-Associated CSC Enrichment and EMT in Preclinical TNBC Models.* Mol Cancer Res, 2018. **16**(5): p. 869-879.
152. Zyromski, N.J., et al., *Obesity potentiates the growth and dissemination of pancreatic cancer.* Surgery, 2009. **146**(2): p. 258-63.
153. Park, E.J., et al., *Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression.* Cell, 2010. **140**(2): p. 197-208.
154. Yoshimoto, S., et al., *Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome.* Nature, 2013. **499**(7456): p. 97-101.
155. Matsubara, M., S. Maruoka, and S. Katayose, *Inverse relationship between plasma adiponectin and leptin concentrations in normal-weight and obese women.* Eur J Endocrinol, 2002. **147**(2): p. 173-80.
156. Miyoshi, Y., et al., *Association of serum adiponectin levels with breast cancer risk.* Clin Cancer Res, 2003. **9**(15): p. 5699-704.
157. Kubota, N., et al., *Adiponectin stimulates AMP-activated protein kinase in the hypothalamus and increases food intake.* Cell Metab, 2007. **6**(1): p. 55-68.
158. Dieudonne, M.N., et al., *Adiponectin mediates antiproliferative and apoptotic responses in human MCF7 breast cancer cells.* Biochem Biophys Res Commun, 2006. **345**(1): p. 271-9.
159. Dos Santos, E., et al., *Adiponectin mediates an antiproliferative response in human MDA-MB 231 breast cancer cells.* Oncol Rep, 2008. **20**(4): p. 971-7.
160. Taliaferro-Smith, L., et al., *LKB1 is required for adiponectin-mediated modulation of AMPK-S6K axis and inhibition of migration and invasion of breast cancer cells.* Oncogene, 2009. **28**(29): p. 2621-33.
161. Wang, Y., et al., *Adiponectin modulates the glycogen synthase kinase-3beta/beta-catenin signaling pathway and attenuates mammary tumorigenesis of MDA-MB-231 cells in nude mice.* Cancer Res, 2006. **66**(23): p. 11462-70.
162. Gao, Y., et al., *Adipocytes promote breast tumorigenesis through TAZ-dependent secretion of Resistin.* Proc Natl Acad Sci U S A, 2020. **117**(52): p. 33295-33304.
163. Qureshi, R., et al., *The Major Pre- and Postmenopausal Estrogens Play Opposing Roles in Obesity-Driven Mammary Inflammation and Breast Cancer Development.* Cell Metab, 2020. **31**(6): p. 1154-1172 e9.
164. Kim, S.H., J. Turnbull, and S. Guimond, *Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor.* J Endocrinol, 2011. **209**(2): p. 139-51.
165. Ling, L., et al., *Obesity-associated Adipose Stromal Cells Promote Breast Cancer Invasion Through Direct Cell Contact and ECM Remodeling.* Adv Funct Mater, 2020. **30**(48).
166. Seo, B.R., et al., *Obesity-dependent changes in interstitial ECM mechanics promote breast tumorigenesis.* Sci Transl Med, 2015. **7**(301): p. 301ra130.
167. Iyengar, P., et al., *Adipocyte-derived collagen VI affects early mammary tumor progression in vivo, demonstrating a critical interaction in the tumor/stroma microenvironment.* J Clin Invest, 2005. **115**(5): p. 1163-76.
168. Baldock, C., et al., *The supramolecular organization of collagen VI microfibrils.* J Mol Biol, 2003. **330**(2): p. 297-307.
169. Cescon, M., et al., *Collagen VI at a glance.* J Cell Sci, 2015. **128**(19): p. 3525-31.

170. Wishart, A.L., et al., *Decellularized extracellular matrix scaffolds identify full-length collagen VI as a driver of breast cancer cell invasion in obesity and metastasis.* Sci Adv, 2020. **6**(43).
171. Park, J. and P.E. Scherer, *Adipocyte-derived endotrophin promotes malignant tumor progression.* J Clin Invest, 2012. **122**(11): p. 4243-56.
172. Bu, D., et al., *Human endotrophin as a driver of malignant tumor growth.* JCI Insight, 2019. **5**.
173. Iyengar, N.M., C.A. Hudis, and A.J. Dannenberg, *Obesity and cancer: local and systemic mechanisms.* Annu Rev Med, 2015. **66**: p. 297-309.
174. Vaysse, C., et al., *Inflammation of mammary adipose tissue occurs in overweight and obese patients exhibiting early-stage breast cancer.* NPJ Breast Cancer, 2017. **3**: p. 19.
175. Hermano, E., et al., *Heparanase Accelerates Obesity-Associated Breast Cancer Progression.* Cancer Res, 2019. **79**(20): p. 5342-5354.
176. Nagahashi, M., et al., *Targeting the SphK1/S1P/S1PR1 Axis That Links Obesity, Chronic Inflammation, and Breast Cancer Metastasis.* Cancer Res, 2018. **78**(7): p. 1713-1725.
177. Wogsland, C.E., et al., *High-dimensional immunotyping of tumors grown in obese and non-obese mice.* Dis Model Mech, 2021. **14**(4).
178. Kolb, R., et al., *Obesity-associated NLRC4 inflammasome activation drives breast cancer progression.* Nat Commun, 2016. **7**: p. 13007.
179. Incio, J., et al., *PlGF/VEGFR-1 Signaling Promotes Macrophage Polarization and Accelerated Tumor Progression in Obesity.* Clin Cancer Res, 2016. **22**(12): p. 2993-3004.
180. Tiwari, P., et al., *Metabolically activated adipose tissue macrophages link obesity to triple-negative breast cancer.* J Exp Med, 2019. **216**(6): p. 1345-1358.
181. Font-Burgada, J., B. Sun, and M. Karin, *Obesity and Cancer: The Oil that Feeds the Flame.* Cell Metab, 2016. **23**(1): p. 48-62.
182. Karpe, F., J.R. Dickmann, and K.N. Frayn, *Fatty acids, obesity, and insulin resistance: time for a reevaluation.* Diabetes, 2011. **60**(10): p. 2441-9.
183. Nielsen, S., et al., *Splanchnic lipolysis in human obesity.* J Clin Invest, 2004. **113**(11): p. 1582-8.
184. Madak-Erdogan, Z., et al., *Free Fatty Acids Rewire Cancer Metabolism in Obesity-Associated Breast Cancer via Estrogen Receptor and mTOR Signaling.* Cancer Res, 2019. **79**(10): p. 2494-2510.
185. Ladanyi, A., et al., *Adipocyte-induced CD36 expression drives ovarian cancer progression and metastasis.* Oncogene, 2018. **37**(17): p. 2285-2301.
186. Pascual, G., et al., *Targeting metastasis-initiating cells through the fatty acid receptor CD36.* Nature, 2017. **541**(7635): p. 41-45.
187. Hao, J., et al., *Circulating Adipose Fatty Acid Binding Protein Is a New Link Underlying Obesity-Associated Breast/Mammary Tumor Development.* Cell Metab, 2018. **28**(5): p. 689-705 e5.
188. Nieman, K.M., et al., *Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth.* Nat Med, 2011. **17**(11): p. 1498-503.
189. Reinfeld, B.I., et al., *Cell-programmed nutrient partitioning in the tumour microenvironment.* Nature, 2021. **593**(7858): p. 282-288.
190. Ringel, A.E., et al., *Obesity Shapes Metabolism in the Tumor Microenvironment to Suppress Anti-Tumor Immunity.* Cell, 2020. **183**(7): p. 1848-1866 e26.
191. Fischer, J.P., et al., *Impact of obesity on outcomes in breast reconstruction: analysis of 15,937 patients from the ACS-NSQIP datasets.* J Am Coll Surg, 2013. **217**(4): p. 656-64.
192. DiSipio, T., et al., *Incidence of unilateral arm lymphoedema after breast cancer: a systematic review and meta-analysis.* Lancet Oncol, 2013. **14**(6): p. 500-15.
193. Lyman, G.H., D.C. Dale, and J. Crawford, *Incidence and predictors of low dose-intensity in adjuvant breast cancer chemotherapy: a nationwide study of community practices.* J Clin Oncol, 2003. **21**(24): p. 4524-31.
194. Griggs, J.J., et al., *Racial disparity in the dose and dose intensity of breast cancer adjuvant chemotherapy.* Breast Cancer Res Treat, 2003. **81**(1): p. 21-31.
195. Griggs, J.J., et al., *Effect of patient socioeconomic status and body mass index on the quality of breast cancer adjuvant chemotherapy.* J Clin Oncol, 2007. **25**(3): p. 277-84.
196. Griggs, J.J., M.E. Sorbero, and G.H. Lyman, *Undertreatment of obese women receiving breast cancer chemotherapy.* Arch Intern Med, 2005. **165**(11): p. 1267-73.
197. Lyman, G.H. and A. Sparreboom, *Chemotherapy dosing in overweight and obese patients with cancer.* Nat Rev Clin Oncol, 2013. **10**(8): p. 451-9.

198. Chrisanthar, R., et al., *Predictive and prognostic impact of TP53 mutations and MDM2 promoter genotype in primary breast cancer patients treated with epirubicin or paclitaxel.* PLoS One, 2011. **6**(4): p. e19249.
199. Vargo-Gogola, T. and J.M. Rosen, *Modelling breast cancer: one size does not fit all.* Nat Rev Cancer, 2007. **7**(9): p. 659-72.
200. Holliday, D.L. and V. Speirs, *Choosing the right cell line for breast cancer research.* Breast Cancer Res, 2011. **13**(4): p. 215.
201. Neve, R.M., et al., *A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.* Cancer Cell, 2006. **10**(6): p. 515-27.
202. Fillmore, C.M. and C. Kuperwasser, *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**(2): p. R25.
203. Zhang, Y., et al., *Positional cloning of the mouse obese gene and its human homologue.* Nature, 1994. **372**(6505): p. 425-32.
204. Zheng, Q., et al., *Leptin deficiency suppresses MMTV-Wnt-1 mammary tumor growth in obese mice and abrogates tumor initiating cell survival.* Endocr Relat Cancer, 2011. **18**(4): p. 491-503.
205. Cleary, M.P., et al., *Genetically obese MMTV-TGF-alpha/Lep(ob)Lep(ob) female mice do not develop mammary tumors.* Breast Cancer Res Treat, 2003. **77**(3): p. 205-15.
206. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.
207. Buettner, R., J. Scholmerich, and L.C. Bollheimer, *High-fat diets: modeling the metabolic disorders of human obesity in rodents.* Obesity (Silver Spring), 2007. **15**(4): p. 798-808.
208. Liu, X.Z., L. Pedersen, and N. Halberg, *Cellular mechanisms linking cancers to obesity.* Cell Stress, 2021. **5**(5): p. 55-72.
209. Olson, L.K., et al., *Pubertal exposure to high fat diet causes mouse strain-dependent alterations in mammary gland development and estrogen responsiveness.* Int J Obes (Lond), 2010. **34**(9): p. 1415-26.
210. Lamas, B., et al., *Dietary fat without body weight gain increases in vivo MCF-7 human breast cancer cell growth and decreases natural killer cell cytotoxicity.* Mol Carcinog, 2015. **54**(1): p. 58-71.
211. Kleinert, M., et al., *Animal models of obesity and diabetes mellitus.* Nat Rev Endocrinol, 2018. **14**(3): p. 140-162.
212. Gengenbacher, N., M. Singhal, and H.G. Augustin, *Preclinical mouse solid tumour models: status quo, challenges and perspectives.* Nat Rev Cancer, 2017. **17**(12): p. 751-765.
213. Dogan, S., et al., *Effects of high-fat diet and/or body weight on mammary tumor leptin and apoptosis signaling pathways in MMTV-TGF-alpha mice.* Breast Cancer Res, 2007. **9**(6): p. R91.
214. Boulange, C.L., et al., *Early metabolic adaptation in C57BL/6 mice resistant to high fat diet induced weight gain involves an activation of mitochondrial oxidative pathways.* J Proteome Res, 2013. **12**(4): p. 1956-68.
215. Chen, J. and X.F. Huang, *High fat diet-induced obesity increases the formation of colon polyps induced by azoxymethane in mice.* Ann Transl Med, 2015. **3**(6): p. 79.
216. Ma, S. and Y. Zhang, *Profiling chromatin regulatory landscape: insights into the development of ChIP-seq and ATAC-seq.* Molecular Biomedicine, 2020. **1**(1): p. 9.
217. Minnoye, L., et al., *Chromatin accessibility profiling methods.* Nature Reviews Methods Primers, 2021. **1**(1): p. 10.
218. Klemm, S.L., Z. Shipony, and W.J. Greenleaf, *Chromatin accessibility and the regulatory epigenome.* Nature Reviews Genetics, 2019. **20**(4): p. 207-220.
219. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nature Methods, 2013. **10**(12): p. 1213-1218.
220. Berest, I., et al., *Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF.* Cell Rep, 2019. **29**(10): p. 3147-3159 e12.
221. Skene, P.J. and S. Henikoff, *An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites.* Elife, 2017. **6**.
222. Skene, P.J., J.G. Henikoff, and S. Henikoff, *Targeted in situ genome-wide profiling with high efficiency for low cell numbers.* Nature Protocols, 2018. **13**(5): p. 1006-1019.
223. Mihaylova, M.M., D.M. Sabatini, and O.H. Yilmaz, *Dietary and metabolic control of stem cell function in physiology and cancer.* Cell Stem Cell, 2014. **14**(3): p. 292-305.

224. Beyaz, S., et al., *High-fat diet enhances stemness and tumorigenicity of intestinal progenitors.* Nature, 2016. **531**(7592): p. 53-8.
225. Bhardwaj, P. and K.A. Brown, *Obese Adipose Tissue as a Driver of Breast Cancer Growth and Development: Update and Emerging Evidence.* Front Oncol, 2021. **11**: p. 638918.
226. Wlodarczyk, M. and G. Nowicka, *Obesity, DNA Damage, and Development of Obesity-Related Diseases.* Int J Mol Sci, 2019. **20**(5).
227. Chung, K.M., et al., *Endocrine-Exocrine Signaling Drives Obesity-Associated Pancreatic Ductal Adenocarcinoma.* Cell, 2020. **181**(4): p. 832-847 e18.
228. Leung, A., et al., *Open chromatin profiling in mice livers reveals unique chromatin variations induced by high fat diet.* J Biol Chem, 2014. **289**(34): p. 23557-67.
229. Qin, Y., et al., *An obesity-associated gut microbiome reprograms the intestinal epigenome and leads to altered colonic gene expression.* Genome Biol, 2018. **19**(1): p. 7.
230. Du, J., et al., *Chromatin variation associated with liver metabolism is mediated by transposable elements.* Epigenetics Chromatin, 2016. **9**: p. 28.
231. McDonnell, E., et al., *Lipids Reprogram Metabolism to Become a Major Carbon Source for Histone Acetylation.* Cell Rep, 2016. **17**(6): p. 1463-1472.
232. Ishdorj, G., et al., *Lysophosphatidic acid protects cancer cells from histone deacetylase (HDAC) inhibitor-induced apoptosis through activation of HDAC.* J Biol Chem, 2008. **283**(24): p. 16818-29.
233. Naba, A., et al., *Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters.* Elife, 2014. **3**: p. e01308.
234. Oudin, M.J., et al., *Tumor Cell-Driven Extracellular Matrix Remodeling Drives Haptotaxis during Metastatic Progression.* Cancer Discov, 2016. **6**(5): p. 516-31.
235. Yang, J., et al., *Lipocalin 2 promotes breast cancer progression.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3913-8.
236. Mahati, S., et al., *TMPRSS4 promotes cancer stem cell traits by regulating CLDN1 in hepatocellular carcinoma.* Biochem Biophys Res Commun, 2017. **490**(3): p. 906-912.
237. Yan, Q.W., et al., *The adipokine lipocalin 2 is regulated by obesity and promotes insulin resistance.* Diabetes, 2007. **56**(10): p. 2533-40.
238. Yang, J., et al., *Lipocalin 2 is a novel regulator of angiogenesis in human breast cancer.* FASEB J, 2013. **27**(1): p. 45-50.
239. Berger, T., et al., *Disruption of the Lcn2 gene in mice suppresses primary mammary tumor formation but does not decrease lung metastasis.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 2995-3000.
240. Yan, L., et al., *The high molecular weight urinary matrix metalloproteinase (MMP) activity is a complex of gelatinase B/MMP-9 and neutrophil gelatinase-associated lipocalin (NGAL). Modulation of MMP-9 activity by NGAL.* J Biol Chem, 2001. **276**(40): p. 37258-65.
241. Dhawan, P., et al., *Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer.* J Clin Invest, 2005. **115**(7): p. 1765-76.
242. Suh, Y., et al., *Claudin-1 induces epithelial-mesenchymal transition through activation of the c-Abl-ERK signaling pathway in human liver cells.* Oncogene, 2013. **32**(41): p. 4873-82.
243. Zhou, B., et al., *Claudin-18-mediated YAP activity regulates lung stem and progenitor cell homeostasis and tumorigenesis.* J Clin Invest, 2018. **128**(3): p. 970-984.
244. Sun, K., et al., *Endotrophin triggers adipose tissue fibrosis and metabolic dysfunction.* Nat Commun, 2014. **5**: p. 3485.
245. Willumsen, N., C. Bager, and M.A. Karsdal, *Matrix Metalloprotease Generated Fragments of Type VI Collagen Have Serum Biomarker Potential in Cancer - A Proof of Concept Study.* Transl Oncol, 2019. **12**(5): p. 693-698.
246. Rasmussen, D.G.K., et al., *Higher Collagen VI Formation Is Associated With All-Cause Mortality in Patients With Type 2 Diabetes and Microalbuminuria.* Diabetes Care, 2018. **41**(7): p. 1493-1500.
247. Rasmussen, D.G.K., et al., *Urinary endotrophin predicts disease progression in patients with chronic kidney disease.* Sci Rep, 2017. **7**(1): p. 17328.
248. Khan, T., et al., *Metabolic dysregulation and adipose tissue fibrosis: role of collagen VI.* Mol Cell Biol, 2009. **29**(6): p. 1575-91.
249. Lai, M., P.C. Chandrasekera, and N.D. Barnard, *You are what you eat, or are you? The challenges of translating high-fat-fed rodents to human obesity and diabetes.* Nutr Diabetes, 2014. **4**: p. e135.
250. Speakman, J.R., *Use of high-fat diets to study rodent obesity as a model of human obesity.* Int J Obes (Lond), 2019. **43**(8): p. 1491-1492.

251. Wu, S.Z., et al., *A single-cell and spatially resolved atlas of human breast cancers.* Nature Genetics, 2021. **53**(9): p. 1334-1347.
252. Fan, X., et al., *High-Fat Diet Alters the Expression of Reference Genes in Male Mice.* Front Nutr, 2020. **7**: p. 589771.

I

C/EBPB-dependent Adaptation to Palmitic Acid Promotes Tumor Formation in Hormone Receptor Negative Breast Cancer

Xiao-Zheng Liu[1,7], Anastasiia Rulina[1,7], Man Hung Choi[2,3,7], Line Pedersen[1], Johanna Lepland[1], Sina T Takle[1], Noelly Madeleine[1], Stacey D'mello Peters[1], Cara Ellen Wogsland[1], Sturla Magnus Grøndal[1], James B Lorens[1], Hani Goodarzi[4], Per E Lønning[5,6], Stian Knappskog[5,6], Anders Molven[2,3], Nils Halberg[1,*]

[1]Department of Biomedicine, University of Bergen, N-5020 Bergen, Norway

[2]Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, N-5020 Bergen, Norway

[3]Department of Pathology, Haukeland University Hospital, N-5021 Bergen, Norway

[4]Department of Biophysics and Biochemistry, University of California San Francisco, San Francisco, CA 94158, USA

[5]Department of Clinical Science, Faculty of Medicine, University of Bergen, N-5020 Bergen, Norway

[6]Department of Oncology, Haukeland University Hospital, N-5021 Bergen, Norway

[7]These authors contributed equally

*Correspondence:

Nils Halberg

Department of Biomedicine

University of Bergen

Jonas Lies vei 91

5020 Bergen, Norway

Phone: +47 5558 6442

Email: nils.halberg@uib.no

Abstract

Epidemiological studies have established a positive association between obesity and the incidence of postmenopausal breast cancer. Moreover, it is known that obesity promotes stem cell-like properties of breast cancer cells. However, the cancer cell-autonomous mechanisms underlying this correlation are not well defined. Here we demonstrate that obesity-associated tumor formation is driven by cellular adaptation rather than expansion of pre-existing clones within the cancer cell population. While there is no correlation with specific mutations, cellular adaptation to obesity is governed by palmitic acid (PA) and leads to enhanced tumor formation capacity of breast cancer cells. This process is governed epigenetically through increased chromatin occupancy of the transcription factor CCAAT/enhancer-binding protein beta (C/EBPB). Obesity-induced epigenetic activation of C/EBPB regulates cancer stem-like properties by modulating the expression of key downstream regulators including *CLDN1* and *LCN2*. Collectively, our findings demonstrate that obesity drives cellular adaptation to PA drives tumor initiation in the obese setting through activation of a C/EBPB dependent transcriptional network.

Key words:

C/EBPB, obesity, breast cancer, palmitic acid, tumor initiation

Significance:

Cellular adaptation to obesity-induced palmitic acid drives tumor initiation through activation of a C/EBPB-dependent transcriptional network. This highlights a cell-autonomous mechanistic connection between obesity and postmenopausal hormone receptor negative breast cancer.

Introduction

Breast cancer is the most frequently diagnosed cancer and a leading cause of cancer-related death among women. Risk factors for breast cancer consist of non-modifiable

factors, such as age, genetic predisposition and reproductive history, and modifiable factors such as obesity and alcohol consumption and tobacco smoking. As an independent risk factor, postmenopausal (PM) obesity accounts for up to 20% higher risk of developing breast cancer, and every 5-unit increase in BMI is associated with a 12% increase in breast cancer risk[1]. Whereas obesity in PM individuals has been consistently linked to higher risk of developing estrogen receptor (ER) positive breast cancer, the effect in ER negative breast cancer has been more debated[2]. In addition to effects on breast cancer risk, meta-analyses have suggested that overweight and obesity are associated with worse overall survival and metastasis-free survival independent of menopause or hormone receptor status[3-5].

Work in mouse models generally recapitulates both obesity-linked tumor initiation and progression[6]. Of these, tumor progression has been most extensively studied and proposed mechanisms include obesity-induced chronic inflammation[7,8], altered insulin signaling[9], deregulation of estrogen[10], rewiring of cancer metabolism[11] and adipokine secretion[12]. Recent insights into obesity-dependent regulation of tumor initiation in breast cancer have highlighted a number of non-cell-autonomous mechanisms including regulation of metabolically activated macrophages[13], leptin[14] and FABP4[12].

Here, we aimed to identify cancer cell-autonomous determinants of obesity-induced PM breast cancer risk. We demonstrate that obesity has adverse effects on patient survival in PM, ER/progesterone receptor (PR) negative breast cancers compared to other subtypes. We show that cellular adaptation to obese environments is phenotypically and mechanistically recapitulated by long-term exposure to high concentrations of palmitic acid (PA) *in vitro*. Both obesity and long-term adaptation to high levels of PA engender cancer cell dedifferentiation towards stem cell-like properties in both human biobank material and mouse models. Mechanistically, we identify epigenetic activation of the CCAAT/enhancer-binding protein beta (C/EBPB) transcription factor as a required regulator of obesity-induced cancer stem-like properties. We further demonstrate that C/EBPB governs cancer stemness through the modulation of CLDN1 and LCN2. Taken together, our findings demonstrate that cellular adaptation to obesity-induced PA is a key driver of tumor initiation in PM/ER$^-$/PR$^-$ breast cancer cells in obesity.

Results

Obesity is associated with increased frequency of stem cell-like cancer cells in PM/ER$^-$/PR$^-$ breast cancer patients and mouse models of breast cancer

To quantitively determine how obesity is linked to increased risk of breast cancer, we orthotopically implanted E0771 and TeLi (Wnt1-driven) cells at limiting dilutions in a C57BL/6J diet-induced obesity model and measured tumor formation. High-fat diet (HFD) feeding resulted in weight gain, and HFD-fed mice displayed multiple hallmarks of obesity-induced comorbidities such as liver steatosis, hyperinsulinemia, hyperglycemia and reduced glucose clearance compared to the regular chow diet-fed mice (Figure S1A-D). Following mammary fat pad implantation of limiting numbers of E0771 and TeLi cells, we demonstrate that high-fat environments consistently promote tumor formation with a 6 to 10-fold enrichment in cancer stem cell frequencies (Figure 1A). This is consistent with a previous report[13] that focused on non-cancer cell-autonomous regulation linking obesity to breast cancer initiation. In contrast, we set out to investigate how the distorted metabolic environment affects cancer cell behavior and thereby gain insights into potential cancer cell-autonomous mechanisms that drive breast cancer initiation in obese environments. To establish the framework for such mechanistic studies, we first sought to identify a group of patients affected by the obese state. To that end, we performed survival analyses of 115 PM (defined by age of >50 years) breast cancer patients using BMI and hormone (estrogen and progesterone) receptor status as variables in a highly controlled in-house dataset[15]. Overweight and obesity (BMI>25) were associated with significantly reduced disease-specific survival rates in hormone receptor negative patients as compared to non-obese patients (Figure 1B). No effects of BMI on disease-specific survival were observed in the hormone receptor positive patients despite having equal BMI distribution as hormone receptor negative patients (Figure S1E, F). Importantly, within the PM/ER$^-$/PR$^-$ patient group there were no differences between the high and low BMI groups with respect to patient age (Figure S1G), tumor size (Figure S1H) or tumor stage (all included patients were stage 3) at the time of diagnosis.

To examine potential cancer cell specific adaptations in the *in vivo* tumor microenvironment of obese and non-obese patients, we obtained tumor tissue microarrays (TMA) from this group of PM/ER$^-$/PR$^-$ patients and immunostained the

cores for the stemness markers CD133[16,17] and Axl[18]. The image analysis platform QuPath[19] was used to segment the images, differentiate between stromal and cancer cells and to quantify cancer cell specific $CD133^{+}$ and Axl^{high} cell frequencies. $PM/ER^{-}/PR^{-}$ breast cancer patients with a BMI above 25 displayed both higher $CD133^{+}$ and Axl^{high} cancer cell frequencies as compared to the normal BMI patients (Figure 1C, D). This suggests that adaptation to obese environment leads to an enrichment in cancer stemness in both mice models and breast cancer patients.

To investigate how the obese environment affects cancer cells phenotypes, we hypothesized that the cellular adaptations induced by obesity are maintained *ex vivo*. We therefore dissociated tumors formed in the obesogenic and non-obesogenic environments and sought to identify deregulated cellular traits. While cellular proliferation was unaffected (Figure 1E), the ability to form tumorspheres was significantly enhanced following adaptation to the obesogenic environment (Figure 1F). After isolation from tumors formed in obese mice, *ex vivo* E0771 cells displayed metabolic rewiring rendering the cells more reliant on PA oxidation and less reliant on glucose oxidation (Figure 1G). Both enhanced tumorsphere formation capacity[20-22] and metabolic reprogramming[23,24] has been linked to stem cell behavior in breast cancer and are thus consistent with cancer cell-autonomous effects of obesity.

Long-term adaptation to PA phenocopy obesity-induced stem-cell features

We next wondered if the obese environment selects for a pre-existing clone or governs generalized adaptation. To this end we tagged individual E0771 cells with a high complexity DNA barcode library (ClonTracer)[25] to track the fate of individual cancer cells as tumors formed in obese and non-obese environments. Generally, we observed a high variability of unique barcodes in tumors evolving in lean as well as in obese mice. The three replicates presented different abundances of barcodes, and the enrichment of certain subset of barcodes was not presented in all replicates (FigureS2A). We then compared the barcode distribution of relative barcode size between the tumors derived from chow and HFD-fed mice and demonstrate that the overall barcode distributions were unaffected by the different diets, suggesting that exposure to an obese environment did not select for a pre-existing clone within the subpopulations (Figure 2A).

Obesity leads to the production of reactive oxygen species in adipose tissue[26]. Given the abundant adipose tissue in the mammary gland and the association between reactive oxygen species and mutagenesis[27], we then asked if the obese phenotype enriches for specific mutations. We therefore performed high-coverage sequencing of 360 known cancer genes[28] in tumor samples collected from the PM/ER$^-$/PR$^-$ patient group at the time of diagnosis. Based on the analysis we were not able to detect any mutations correlating to obesity across this panel (Supplementary Table S1). The combined results from the *in vivo* barcode studies and patient analysis suggest that obesity governs cellular adaptation independent of obesity-dependent genetic changes.

To understand what drives such cellular adaptation we reasoned that PA could play a central role as i) the circulating abundance of PA strongly elevated in obese individuals reaching levels that are toxic to cancer cells and, thus, could feasibly provide a new selection pressure[29-31], ii) PA has been reported to be epidemiologically associated with a higher risk of developing PM breast cancer[32,33] and iii) we (Figure 1) and others demonstrate that cancer cell PA metabolism is altered by obesity[7]. To assess how breast cancer cells adapt to PA exposure, we cultured hormone receptor negative breast cancer cell lines to increasing PA concentrations over a period of 2 months. Human (MDA-MB-231 and HCC1806) and mouse (E0771) breast cancer cells consistently adapted to acquire resistance to PA-induced apoptosis to enable persistent growth even in the high PA environment (Figure 2B; Figure S2B). For adapted MDA-MB-231 (MDAapa) and HCC1806 (HCC1806apa), acquired resistance was accompanied by a reduction in growth rate, whereas adapted E0771 (E0771apa) cells maintained its growth rate even after adaptation to high levels of PA (Figure 2C; Figure S2C). Importantly, the final PA concentration is similar to the serum levels of PA in obese individuals[29].

To obtain insights into the cellular dynamics of cellular adaptation to PA, we performed a single cell mass cytometry analysis of MDA-MB-231 and HCC1806 parental (MDApar and HCC1806par) and adapted cells, using an antibody panel targeting 27 markers of cellular differentiation states and signaling pathways (Supplementary Table S2). Dimensionality reduction and visualization were performed based on the differential marker expression profiles and cellular densities were depicted in the tSNE maps. This analysis revealed that PA adaptation governed a clear shift within the

cancer cell subpopulations (Figure 2D and Figure S2D). Prominently, cellular subpopulations enriched by adaptation to PA where characterized by increased expression of the cancer stem cell markers CD44[34], CD133 and Axl, in both HCC1806 (Figure 2E) and MDA-MB-231 cells (Figure S2E). This supports that long-term adaptation to PA phenocopies adaptation to the obese environment by inducing cellular dedifferentiation towards a cancer stem cell-like state. Increased frequency of $CD44^{high}/CD133^{+}$ cell populations were validated using flow cytometry (Figure S2F, G).

These findings suggested that PA-adapted cells would have greater tumor initiation capacity in the obese setting. To test this, we implanted E0771 parental (E0771par) and adapted cells at limiting dilutions in the mammary fat pad of obese and non-obese mice and scored tumor formation rates. Interestingly, *in vitro* adaptation to PA enriched for stemness properties that confer increased tumor formation capacity *in vivo* in obese, but not lean mice (Figure 2F). Further, adaptation to PA in MDA-MB-231, HCC1806 and E0771 cells phenocopied enhanced tumorsphere formation capacity and metabolic reprogramming as demonstrated in the *ex vivo* E0771 model (Figure S2H-M). These findings reveal that *in vitro* adaptation to high concentration of PA phenocopies key obesity-induced tumor initiation phenotypes.

To ascertain how such adaptation resembles what is observed in obese breast cancer patients, we compared the transcriptional alterations observed during *in vitro* cellular adaptations to PA to the transcriptional changes induced by obesity in PM hormone negative breast cancers patients. To this end, we applied iPAGE, an information-theoretic framework[35], to query how genes induced or repressed in obesity were changed upon adaptation to PA in the *in vitro* model. For this analysis, genes were first ordered based on their expression changes between MDA-MB-231 parental and adapted cells and were subsequently divided into 10 equally populated bins. We then assessed the distribution of obesity-associated genes across these bins. As shown in Figure 2G, we observed a significant depletion/enrichment pattern (MI=0.006 bits, z-score=21.14). We specifically noted a significant overlap between genes that were induced by the obesogenic state in patients and those up-regulated through *in vitro* adaptation to PA (Figure 2G). This shared reprogramming of the gene expression landscape suggested that the *in vitro* long-term adaptation to high abundancies of PA provides clinically relevant information about the molecular drivers of obesity-induced hormone receptor negative breast cancers. Combined, this suggest that the cellular

reprogramming leading to enhanced tumor initiation in obese patients can be governed by long-term adaptation to PA.

Adaptation to obese environments induces open chromatin linked with C/EBPB occupancy

Deregulation of metabolic intermediates has recently been tightly linked to epigenetic remodeling and cell fates[36]. To obtain mechanistic insights into obesity-induced cellular adaptations, we therefore next assessed chromatin accessibility by ATAC sequencing (ATACseq) of E0771 cancer cells collected *ex vivo* after adaptation to lean or obese environments *in vivo*. This demonstrated that exposure to an obese environment causes epigenetic remodeling in already malignant cells (329 gain peaks and 1158 loss peaks; Figure 3A, B). To ascertain if such remodeling could be related to cancer cell dedifferentiation, we obtained ATACseq data from isolated murine fetal mammary stem cells, basal cells, luminal progenitors and mature luminal cells (GSE116386). We then performed a principal component analysis of these developmental stages and the E0771 tumors isolated from obese and lean mice. This revealed that adaptation to the obese environment drives significant epigenetic remodeling towards the fetal mammary stem cell state. Compared to the lean state, tumor adaptation to the obese setting resulted in 11.24 (PC1) and 6.01 (PC2) standard deviations closer to the mean principal components of the fetal mammary stem cell state (Figure 3C).

To identify the functional consequence of such epigenetic remodeling, we next aggregated changes in chromatin accessibility near putative transcription factor binding motifs to infer differential motif activity and occupancy[37]. To determine potentially common regulators, we included both obesity-induced (*ex vivo* E0771 derived from HFD compared to chow mice) and PA-specific adaption (MDAapa compared to MDApar), which similarly caused chromatin remodeling (Figure S3A, B). Interestingly, this analysis identified the C/EBPB and C/EBPA transcription factors as having the strongest association with the more accessible chromatin in both obesity-induced and PA-adapted cells (Figure 3D). As homologs of the C/EBP family transcription factors, C/EBPA and C/EBPB bind to the similar 8-mer canonical TTGCGCAA motif, which is difficult to resolve with motif enrichment algorithms.

However, as C/EBPB is robustly expressed and C/EBPA expression is very low in MDAapa cells, we focused on C/EBPB for downstream studies (719-fold difference; Figure 3E).

C/EBPB has been implicated in determination of cell fate in a variety of tissues, including mammary gland[38]. To examine relevance of epigenetic regulation of C/EBPB accessibility in early mammary gland developmental processes, we examined published single-nuclei ATACseq analysis of murine mammary cells at different developmental stages (GSE125523). Pseudotime and transcription factor motif analysis revealed that the C/EBPB motif was highly enriched in open chromatin in fetal mammary stem cells and other progenitor cells along the luminal lineage but became inactive as the cells became terminally mature luminal cells (Figure S3C-E). This suggested that epigenetic activation of C/EBPB may be important for maintenance of mammary stem/progenitor cell fates.

To independently validate enhanced chromatin accessibility for C/EBPB in the obese setting, we performed protein-DNA mapping (Cut&Run sequencing) against activating (H3K4me1) and repressive (H3K27me3) histone marks[39] in MDApar and MDAapa cells, across the same regions (-1 kb - +1 kb relative to C/EBPB motifs) as assessed by ATACseq (Figure 3F). Cut&Run uses micrococcal nuclease tethered to DNA-bound proteins to generate short DNA cleavage fragments and thus enables base-resolution digital footprints that reflect precise protein-DNA binding sites[40]. Consistent with the ATACseq analysis, the active mark was increased and the repressive mark was repressed in the in the MDAapa compared to the MDApar cells (Figure 3G, H). Collectively, the ATACseq and Cut&Run analysis implicate epigenetic activation of C/EBPB transcriptional activity as a major driving factor of tumor initiating capacity in obese breast cancer.

C/EBPB promotes tumor stemness specifically in obese environments

We next asked if C/EBPB is functionally related to obesity-induced cancer stemness. C/EBPB is encoded by an intron-less gene that is expressed in three isoforms; LAP1, LAP2 and LIP by alternative use of transcription start sites[41,42]. Functional depletion of C/EBPB by independent short hairpin RNAs (shRNAs) in E0771apa cells (Figure 4A)

led to a significant reduction in tumorsphere formation capacity (Figure 4B) without affecting tumor proliferation (Figure 4C). C/EBPB depletion also resulted in decreased reliance on PA oxidation and reduced use of glucose for oxidation (Figure 4D, E) demonstrating that C/EBPB is functionally required for key obesity-induced phenotypes. Importantly, upon transplantation into the mammary fat pad, depletion of C/EBPB significantly delayed the onset of tumor formation in the obese setting, while the knockdown had no effect in the non-obese setting (Figure 4F). All together, these experiments support a model wherein C/EBPB is associated with transcriptionally active chromatin and is required for the cancer stem-like phenotype in obesity.

Both LAP1 and LAP2 C/EBPB isoforms contain a dimerization and a transcriptional regulation domain and function as dimers[41]. LIP lacks the DNA binding domain and has been suggested to function as a competitive inhibitor of LAP1/2[41]. As the protein levels of C/EBPB isoforms and cellular localization did not differ between adapted and parental cells (Figure S4A-D), we reasoned that obesity- and PA-dependent epigenetic remodeling is required to confer stem like properties. To test this, we overexpressed either LAP2 (containing a conservative ATG to ATC mutation that eliminate the LIP translation start site) or LIP in both parental and PA-adapted cells (Figure 4G). Interestingly, ectopic overexpression of the LAP2 isoform C/EBPB in adapted MDA-MB-231 and HCC1806 conferred increased tumorsphere formation capacity, increased frequency of $CD44^{high}/CD133^{+}$ cell populations and metabolic rewiring without increasing cellular proliferations rates (Figure 4H-K; Figure S4E-F). In contrast, equal ectopic overexpression of C/EBPB in parental cells did not affect tumorsphere formation (Figure 4L, M). Collectively, these findings suggest that epigenetically controlled accessibility of the C/EBPB isoform LAP2 is a key driver of cancer stem cell-like properties in the obese setting.

Differential C/EBPB occupancy regulates extracellular matrix organization

Having shown that C/EBPB is required and sufficient for tumorsphere formation capacity, we next applied Cut&Run to confirm its genome-wide occupancy and to identify its putative downstream transcriptional targets. Although the total number of upregulated and downregulated C/EBPB peaks were similar in MDAapa relative to MDApar, we observed that a substantial number of these peaks were uniquely present

in MDAapa (5618 unique peaks, Figure 5A, red) and in MDApar (3718 unique peaks, Figure 5B, blue). *De novo* motif discovery analysis revealed that the C/EBP family motif was significantly enriched only in the unique peaks in MDAapa, but not those in MDApar (Figure 5A, B). As an independent means to confirm C/EBPB binding in MDAapa irrespective of peak-calling algorithm, we enumerated the ends of every Cut&Run fragment (≤ 120 bp) for each base of the genome and detected significant footprints *de novo* based on the footprint occupancy score[43]. As expected, motif enrichment analysis identified C/EBPB as the most significantly enriched motif in Cut&Run footprints (Figure S5A, B). Also, increased C/EBPB binding coincided with the chromatin regions which had increased accessibility in MDAapa cells as compared to MDApar cells (Figure 5C). These findings confirm increased C/EBPB occupancy in its canonical binding sites in MDAapa, whereas the observed peaks in MDApar may represent nonspecific DNA binding of C/EBPB during its search of accessible motif sites[44].

To determine putative genes regulated by C/EBPB epigenetic remodeling, we focused on genes whose expression increased and where there were distal or proximal gains in C/EBPB occupancy and chromatin accessibility in PA adaptation. In addition, we considered high-confidence enhancer-gene associations identified cross-platform in GeneHancer[45] (e.g. *LCN2*; Figure 5D). Pathway analysis of these regions revealed a significant enrichment in processes involved in ECM organization (Figure 5E), suggesting a link between C/EBPB-dependent ECM remodeling and cancer tumor formation.

To integrate our data derived from the *in vivo* E0771 obesity model and the *in vitro* PA adaptation system with the clinical setting, we subsequently focused on the subset of the putative C/EBPB target genes whose expression was significantly elevated in the obese as compared to the lean PM/ER$^-$/PR$^-$ patients. This analysis identified nine genes, namely, *SERPINB2*, *LCN2*, *SERPINB7*, *NELL2*, *MMP9*, *CLDN1*, *LYPD6B*, *CRISPLD1* and *CHST4* (Figure 5F). Interestingly, all of these nine genes had elevated expression in E0771 cells analyzed *ex vivo* after having been grown in obese and non-obese mice (Figure 5G). In sum, these data supported a model wherein obesity induced C/EBPB chromatin binding, activating a transcriptional network involved in ECM regulatory processes.

***CLDN1 and LCN2 are* required for C/EBPB dependent stem cell-like capabilities.**

To determine the functional importance of the nine genes in C/EPBP-dependent cancer stemness, we next assessed the levels of the nine genes in cells where C/EBPB was overexpressed. Ectopic overexpression of the LAP2 isoform of C/EBPB in MDAapa and HCC1806apa cells led consistent induction of two genes, *LCN2* and *CLDN1* (Figure 6A, B: Figure S6A, B), which paralleled the differential expression patterns observed in cells adapted to obese and non-obese environment (Figure 5G). Ectopic expression of LIP did not affect expression levels across the potential C/EBPB targets genes (Figure 6A, B). We therefore functionally tested the role of LCN2 and CLDN1 in tumorsphere formation assays and found that depletion of both genes reduced tumorsphere formation capacity without affecting proliferation rates (Figure 6C-F, Figure S6C, D) – thus phenocopying C/EBPB depletion. Further, both LCN2 and CLDN1 were epistatically required for LAP2 induced tumorsphere formation capacity (Figure 6G, H and Figure S6E, F). This suggested that CLDN1 and LCN2 were the main downstream mediators of C/EBPB induced stemness. To validate these findings *in vivo*, we implanted control, CLDN1 and LCN2 depleted cells into the mammary fat pad of obese and lean mice and assessed tumor formation. While high fat feeding resulted in accelerated tumor formation rates, this was prevented by depletion of either CLDN1 or LCN2 (Figure 6I, J). Supportive of an obesity-specific effect, depletion of either CLDN1 or LCN2 did not affect tumor take rate in the lean mice (Figure 6I, J). Combined, these results suggest that CLDN1 and LCN2 are the downstream mediators of C/EBPB induced tumor formation capacity in the obese setting.

Discussion

Obesity is a complex pathological condition that conceivably affects the formation and development of cancers through multiple avenues. Here we have demonstrated that cancer cell exposed to the obese environments specifically adapts to high levels of PA to drive enhanced tumor formation capacity in PM/ER$^-$/PR$^-$ breast cancer. Our findings further suggest that this is mediated through a general cellular adaptation process

rather than expansion of a pre-existing cellular subpopulation. We find that obesity-induced adaptation to PA governs dedifferentiation of cancer cells towards a tumor stem cell-like phenotype leading to augmented tumor formation capacity. Clinically, this manifests in a higher cancer cell frequency of $CD133^{+}$ and Axl^{high} cancer stem cells and shorter disease-specific survival in obese and overweight $PM/ER^{-}/PR^{-}$ breast cancer patient compared to normal weight patients. This is corroborated epidemiologically by the association of obesity with higher cancer risk[46] and poor prognosis[4] of $PM/ER^{-}/PR^{-}$ breast cancer patients[47].

Our studies did not identify any specific genetic mutations correlating to obesity in the tumors from PM/ER-/PR- breast patients. Such lack of a genetic link between obesity and cancer formation were supported in a mutated *Kras*-dependent pancreatic ductal adenocarcinoma model. Here, obesity, as induced through the *ob/ob* mutation, led to enhanced tumorigenicity independently of the acquisition of new driver mutations[48]. In contrast to a genetic link, we identified a critical link between adaptation to the obese environment and genome-wide changes in chromatin accessibility. This is supported by recently observations that high fat feeding leads to alterations in chromatin interactions to drive adaptive networks in the liver[49]. These interactions might reflect diet-induced alterations in metabolic intermediates that are intimately connected to epigenetic control of gene transcription[50,51]. Interestingly, lipid-derived acetyl-CoA has been suggested to be the source of up to 90% of acetylation modifications of certain histone lysine's[52]. Our work and the recent work by Ringel et al.,[7], both describe obesity-dependent changes in lipid handling in cancer cells. This suggests that alteration to lipid derived acetyl-CoA could potently affect the chromatin landscape of cancer cells and thus link obesity to tumor formation and progression.

Importantly, we demonstrate that the obesity-dependent epigenetic remodeling is specific, rendering chromatin regions containing the binding motif for C/EBPB more accessible and thereby activating a C/EBPB dependent transcriptional network. Through complementary sets of *in vitro* and *in vivo* experiments, we show that C/EBPB is required for obesity-induced tumor formation. Conversely, ectopic overexpression of C/EBPB enhanced the frequency of cancer stem cells. Previous reports observed that C/EBPB is required for stem cell maintenance in the developing breast using whole body knockout mice[38] and that expression of the LAP2 isoform of C/EBPB can

transform a non-cancerous cell line MCF10A[53], lending further support to the functional role for C/EBPB-dependent tumor formation.

Our unbiased Cut&Run analysis of direct C/EBPB target genes suggested that C/EBPB regulates tumor initiation features through regulation of the surrounding ECM. Cancer cell-autonomous regulation of the ECM is intrinsically linked to cancer stemness through manipulation of mechanical properties and signaling molecules[54,55]. Consistent with our findings, obesity-induced alterations in the ECM mechanics have been reported to support tumorigenesis[56]. Interestingly, a total of nine C/EBPB target genes were also induced in obese PM/ER$^-$/PR$^-$ breast cancer patients. Of these nine genes, depletion *CLDN1* and *LCN2* phenocopied *C/EBPB* knockdown and were epistatically required for C/EBPB induced tumorsphere formation capacity *in vitro* and tumor formation rate *in vivo* suggesting that these engender the downstream effects of C/EBPB (Figure 7A).

LCN2 is induced in adipose tissue of obese individuals[57] and were previously described to reduce inflammation and fibrosis and in an obesity-driven pancreatic ductal adenocarcinoma model[58]. In breast cancer, LCN2 has been linked to cellular differentiation through modulation of the epithelial to mesenchymal transition[59]. As a tight junction protein, CLDN1 is expressed in several types of human cancers[60-62]. In breast cancers, CLDN1's expression was significantly associated with the basal-like subtype of breast cancers[61]. Accumulating evidence has demonstrated that CLDN1 induces EMT to lead metastatic behavior in colon[60] and liver cancer[62].As such, both CLDN1 and LCN2 have been suggested to be involved in cancer dedifferentiation, future work is needed to establish the mechanistic basis of their actions – especially in the context of obese environments.

Aberrant lipid metabolism is a hallmark of deregulated cancer metabolism[63]. It has been widely reported that cancer cells augment their *de novo* lipid biosynthesis for energy production, synthesis of new membranes, to regulate membrane structures that coordinate signal transduction, and for the biosynthesis of lipid signaling molecules such as phosphatidylinositol-3,4,5-trisphosphate[64]. In addition, cancer cells can stimulate the release of fatty acids from surrounding adipocytes to provide energy for tumor growth[65]. In support of a link between fatty acids and stemness, is the observation that slow-cycling metastasis-initiating cells are dependent on the lipid

uptake protein CD36[66]. While we did not observe any direct involvement of CD36 in our studies of obesity-induced breast cancer, both studies describe a critical role for fatty acid metabolism in cancer stemness. Specifically, our study expands on the importance of fatty acids, by demonstrating that obesity-induced PA concentrations drives cellular adaptation of the cancer cell suggesting that PA might exert a critical regulatory role outside of its role in energetics in during tumor formation.

Combined, our analysis of cellular adaptations to obese environments has revealed changes of cellular phenotypes, driven by the combined modulation of C/EBPB transcriptional activity. In the context of personalized medicine, this suggest that obese cancer patients might benefit from specific targeted therapies rather than generic treatment regiments.

Methods

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Axl (application – TMA staining)	RnD Systems	AF154
C/EBPB	Santa Cruz	sc-7962
CD133 (application – TMA staining)	Miltenyi Biotec	130-090-422
Beta-actin	Invitrogen	PA1-183
E-cadherin-158Gd, Extracellular	Fluidigm	3158021A
Cleaved caspase 3-142Nd, Intracellular	Cell Signalling technology	Clone SA1E
CD44-173Yb, Extracellular	Fluidigm	3150018B
CD133-160Gd, Extra- and intracellular, conjugated in the lab	Miltenyi Biotec	130-090-422

Axl-168Er, Extracellular	BGB/creative biolabs	HPAB-0110-LS
CD133-APC	Invitrogen	17-1331-81
CD44-FITC	BioLegend	338803
pEGFR-151Eu, Intracellular, conjugated in the lab	abcam	ab32430
PCreb-176Yb, Intracellular	Fluidigm	3176005A
PAkt-152Sm, Intracellular	Fluidigm	3156002A
P38-156Gd, Intracellular	Fluidigm	3156002A
N-cadherin-143Nd, Extracellular	Fluidigm	3143016B
Keratin7-164Dy, Extracellular	BD	ab9021
EGFR-170Er, Extra- and intracellular	Fluidigm	3170009B
YAP-167Er, Intracellular, conjugated in the lab	Santa Cruz	sc-271134
Vimentin-154Sm, Intracellular	Fluidigm	3154014A
TGFβ-163Dy, Extra- and intracellular	Fluidigm	3163010B
PStat5-147Sm, Intracellular, conjugated in the lab	BD	562077
PStat3-145Nd, Intracellular, conjugated in the lab	BD	624084
PStat1-153Eu, Intracellular	Fluidigm	3153005A
PSHP2-141Pr, Intracellular	Fluidigm	3141002A

PS6-172Yb, Intracellular	Fluidigm	3172008A
PRb-150Nd, Intracellular	Fluidigm	3150013A
PNFKB-166Er, Intracellular	Fluidigm	3166006A
PMAPKAPK2-159Tb, Intracellular	Fluidigm	3159010A
pHistone H3-175Lu, Intracellular	Fluidigm	3175012A
pErk1/2-171Yb, Intracellular, conjugated in the lab	BD	624084
AF647 goat anti-mouse	Life Technologies	A21238
IRDye® 800CW Donkey anti-Rabbit IgG (H + L), 0.1 mg	Leicor	[P/N 926-32213], 0.1 mg
IRDye® 680RD Goat anti-Mouse IgG (H + L), 0.1 mg	Leicor	[P/N 925-68070], 0.1 mg
Mouse IgG isotype	Merck-Millipore	12-371
Bacterial and Virus Strains		
MAX Efficiency™ DH5α™ Competent Cells	Thermofisher Scientific	18258012
Biological Samples		
TMAs of breast cancer tissues	Haukeland University Hospital	[15]
Chemicals, Peptides, and Recombinant Proteins		
Penicillin/streptomycin	Sigma	P-0781
Fetal bovine serum	Sigma	F-7524

DMEM	Sigma	D5671
RPMI1640	Sigma	R8758
BSA, fatty-acids free	Sigma	A7030
Palmitic Acid	Sigma	P5585
Phusion polymerase	NEB	M0530S
QIAquick Gel Extraction Kit	QIAGEN	28704
QIAGEN Plasmid Plus Maxi Kit	QIAGEN	12965
QIAquick PCR Purification Kit	QIAGEN	28104
E.Z.N.A. Tissue DNA Kit	Omega Bio-tek	D3396
Opti-MEM	Thermo Fisher	31985070
Lipofectamine 2000	Invitrogen	11668019
Polybrene Infection / Transfection Reagent	Sigma	TR-1003-G (1 ML)
Puromycin dihydrochloride from Streptomyces alboniger	Sigma	P8833-100MG
Annexin V, Alexa Fluor™ 488 conjugate	Thermo Fisher	A13201
Propidium Iodide	Sigma	P4864
Trypsin	Sigma	T4049
BSA	Sigma	A9647
Accutase	Sigma	A6964
DAKO EnVision+ System- HRP Labelled Polymer Anti-Rabbit	Dako	K4003

Bodipy reagent	Thermo Fisher	D3922
DPBS	Gibco	14040-133
DAPI	Sigma	D9542
ProLong™ Diamond Antifade Mountant	Invitrogen	P36970
NuPAGE™ MOPS SDS Running Buffer (20X)	Invitrogen	NP000102
HEPES solution	Sigma	H0887-100ML
Tween 20	Sigma	P9616-100ML
PBS	Thermo Fischer Scientific	14040133
Tris/EDTA buffer, pH 9	Dako	S2367
Dako Real Peroxidase Blocking solution	Dako	S2023
Protein Block, Serum-free	Dako	X0909
Antibody Diluent with Background Reducing Components	Dako	S3022
Dako Wash Buffer	Dako	S3006
DAB+	Dako	K3468
Hematoxylin	Dako	S3301
Pertex	Histolab	00801
TrypLE Express	Gibco	12604-021
DNase I	Sigma	DN25
Cell-ID™ 20-Plex Pd Barcoding Kit	Fluidigm	201060

Maxpar® 10X Barcode Perm Buffer	Fluidigm	201057
Maxpar® Cell Staining Buffer	Fluidigm	201068
Cell-ID™ Intercalator-Ir	Fluidigm	201192A
Maxpar® Water	Fluidigm	201069
NP-40	New England Biolabs	B2704
5% digitonin	Invitrogen	BN2006
AMPure XP beads	Beckman	A63880
BioMag® Plus Concanavalin A	Bangs Laboratories	BP531
Spermidine	Sigma	S2501
Roche Complete Protease Inhibitor, EDTA-free	Roche	05 892 791 001
PhosSTOP™	Roche	04 906 837 001
Protein-A/G-MNase	Epicypher	15-1016
EDTA	Sigma	03690
EGTA	Boston BioProducts	BM723
RNase A	Sigma	R4642
GlycoBlue	ThermoFisher	AM9515
Matrigel	Corning	356231
L-Glutamine	Sigma	G-7513
Poly-L-Lysine	Sigma	P4832

Paraformaldehyde Aqueous Solution (PFA)	Electron Microscopy Sciences	15710
L-carnitine	Sigma	C-0283
D-[^{14}C(u)]-Glucose	Perkin Elmer	NEC042A001MC
[^{14}C]-Palmitic Acid	PerkinElmer	NEC075H25OUC
Etomoxir	Sigma	E-1905-5M
NaF	Sigma	S-6776
$NaVO_4$	Aldrich	450243
Blotting-Grade Blocker	BioRad	1706404
Tris-HCl	Sigma	T2194-1L
NaCl	Sigma	S5150
$MgCl_2$	Sigma	8266
CaCl2	Sigma	C4830
Lipofectamine™ RNAiMAX Transfection Reagent	Invitrogen	13778075
Lipofectamine™ 2000 Transfection Reagent	Invitrogen	11668019
B-27™ Supplement (50X), minus vitamin A	Gibco	12587010
Recombinant Human FGF-basic (154 a.a.)	PeproTech	100-18B
hEGF	Sigma	E9644
mEGF	R&D Systems	2028-EG

DMEM/F12	Sigma	D8062
Critical Commercial Assays		
PI/RNase staining kit	BD Pharmigen	550825
Senescence β-Galactosidase Staining Kit	Cell Signalling	9860S
Total RNA purification Kit	NORGEN Biotek	37500
SuperScript® III First-Strand Synthesis Kit	ThermoFisher Scientific	18080-051
LightCycler® 480 SYBR Green I Master Mix	Roche	04887352001
MACH3 mouse probe	Biocare Medical	BC-M3M530H
RNA Clean & concentrator with DnaseI kit	Biosite	R1013
Nextera DNA Library Prep kit	Illumina	FC-121-1030
DNA Clean and Concentrator-5 kit	Zymo	D4014
NEBNext® Ultra™ II DNA Library Prep Kit	New Englands Biolabs	E7645
MicroScintPS	PerkinElmer	6013631
Deposited Data		
Sequencing data (RNA-seq, ATACseq and Cut&Run)	EMBL-EBI	PRJEB 39793
Mass cytometry data	FLOW Repository	FR-FCM-Z2TK
Experimental Models: Cell Lines		
MDA-MB-231	ATCC	RRID:CVCL_0062

HCC1806	ATCC	PRID:CVCL_1258
E0771	CH3 BioSystems	SKU: 94A001
TeLi	on-site	N/A
HEK293T	ATCC	PRID:CVCL_0063
Experimental Models: Organisms/Strains		
C57BL/6J	The Jackson Laboratory	000664
Oligonucleotides		
Mouse *Actin* Forward: TACCACAGGCATTGTGATGG Reverse: TTTGATGTCACGCACGATTT Application: qPCR	IDT	N/A
Mouse *C/ebpb* Forward: GGTTTCGGGACTTGATGCA Reverse: CAACAACCCCGCAGGAAC Application: qPCR	IDT	N/A
Mouse *Serpinb2* Forward: TCCCAAACCTGCTACCCGAA	IDT	N/A

Reverse: TGCGAGTTCACACGGAAAGG Application: qPCR		
Mouse *Nell2* Forward: TTGGTGTGGACCCCTCCCTA Reverse: ACTTGGCGCACTCCATCTGT Application: qPCR	IDT	N/A
Mouse *Cldn1* Forward: TCTACGAGGGACTGTGGATG Reverse: TCAGATTCAGCAAGGAGTCG Application: qPCR	IDT	N/A
Mouse *Lcn2* Forward: TGCCACTCCATCTTTCCTGTT Reverse: GGGAGTGCTGGCCAAATAAG Application: qPCR	IDT	N/A
Mouse *Mmp9* Forward: CTTCTGGCGTGTGAGTTTCC	IDT	N/A

Reverse: ACTGCACGGTTGAAGCAAAGA Application: qPCR		
Mouse *Serpinb7* Forward: CTTCACTGCCCTGACCCTAATC Reverse: TGCAGTGCCTTGTCAATCTGA Application: qPCR	IDT	N/A
Human *HPRT* Forward: CCTGACCAAGGAAAGCAAAG Reverse: GACCAGTCAACAGGGGACAT Application: qPCR	IDT	N/A
Human *C/EBPB* Forward: TCGCAGGTCAAGAGCAAGG Reverse: TACTCGTCGCTGTGCTTGTC Application: qPCR	IDT	N/A
Human *RUNX1* Forward:	IDT	N/A

CTGCTCCGTGCTGCCTAC Reverse: AGCCATCACAGTGACCAGAGT Application: qPCR		
Human *C/EBPA* Forward: GGAGCTGAGATCCCGACA Reverse: TTCTAAGGACAGGCGTGGAG Application: qPCR	IDT	N/A
Human *SERPINB2* Forward: CATGGAGCATCTCGTCCAC Reverse: ACTGCATTGGCTCCCACTT Application: qPCR	IDT	N/A
Human *NELL2* Forward: TAAGGGTATAATGCAAGATGTCCAATT Reverse: AGATCTGGGCACTGAGCAATAAA Application: qPCR	IDT	N/A
Human *CLADN1* Forward:	IDT	N/A

GAAGTGCTTGGAAGACGATG Reverse: GAGCCTGACCAAATTCGTAC Application: qPCR		
Human *SERPINB7* Forward: CACTGGTGACTTGACCCTTCCT Reverse: GGTGAGACACATGGTGGTAGAATG Application: qPCR	IDT	N/A
Human *CHST4* Forward: TGGCCATCTTGGCTCTATTC Reverse: CTGCTTGAAGGTCATCCACA Application: qPCR	IDT	N/A
Human *MMP9* Forward: GTTCGACGTGAAGGCGCAG Reverse: TAGTGTGGTGTCTCACGAAGG Application: qPCR	IDT	N/A

Human *LCN2* Forward: TCACCCTCTACGGGAGAACC Reverse: GGGACAGGGAAGACGATGTG Application: qPCR	IDT	N/A
Human *LYPD6B* Forward: TGCAAACCTTTTCACTGTTCCA Reverse: GAGAGCGTGACAGAGGAGCAG Application: qPCR	IDT	N/A
Human *CRISPLD1* Forward: TGCCCAAGAGTATACTGTCCT Reverse: GATTTCGAACCACTCCAGCA Application: qPCR	IDT	N/A
LCN2 siRNA	IDT	hs.Ri.LCN2.13.1, hs.Ri.LCN2.13.3
CLDN1 siRNA	IDT	hs.Ri.CLDN1.13.1, hs.Ri.CLDN1.13.2, hs.Ri.CLDN1.13.3

Negative Control DsiRNA	IDT	51-01-14-04
Recombinant DNA		
m*C/ebpb*-shRNA-1	Sigma	SHCLNG-NM_005194, TRCN0000364533
m*C/ebpb*-shRNA-2	Sigma	SHCLNG-NM_009883, TRCN0000231409
h*C/EBPB*-shRNA-1	Sigma	SHCLNG-NM_005194, TRCN0000364533
h*C/EBPB*-shRNA-2	Sigma	SHCLNG-NM_005194, TRCN0000007443
pBABE-puro LAP2	Addgene	15712
pBABE-puro LIP	Addgene	15713
pBABE-puro	Addgene	1764
Software and Algorithms		
GraphPad Prism	GraphPad	Version 8.4.1
FlowJo	BD Biosciences	Version 10.7.0
R		Version 3.5.0
Cytobank		[67]
ClonTracer	Python package	Version 1.2 [25]
Other		
Rat and Mouse No.1 Maintenance	Special Diet Services	RM1 (P) 801151
Rodent Diet With 60 kcal% Fat	Research Diets	D12492

CONTACT FOR REAGENT FOR AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nils Halberg (nils.halberg@uib.no). This study did not generate new unique reagents. The sequencing data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB 39793. The mass cytometry data has been deposited in the FLOW Repository under repository ID FR-FCM-Z2TK.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Breast Cancer Patient Cohort

This study enrolled a total of 223 patients with primary stage III breast cancers. Out of these 115 patients were PM patients (defined by age > 50 years). Recruitment period was between November 24, 1997 and December 16, 2003. The median age was 51 years (range 25–70). Patiens' BMI, age, hormone status at the time of diagnosis, as well as patient survival times (overall survival and disease specific survival) were documented. The study was approved by the regional committees for medical and health research of Western Norway (REK-Vest; approval number 273/96-82.96). More details about the study cohort can be found in the following report[15].

Animal Models

All animal experiments were approved by the Norwegian Animal Research Authority and conducted according to the European Convention for the Protection of Vertebrates Used for Scientific Purposes, Norway. The Animal Care and Use Programs at University of Bergen are accredited by AAALAC international. The laboratory animal facility at University of Bergen was used for the housing and care of all mice. C57BL/6J mice were obtained from Jackson Laboratories and bred on site. Female mice were kept in IVC-II cages (SealsafeÒ IVC Blue Line 1284L, Tecniplast, Buguggiate, Italy); 5-6 mice were housed together and maintained under standard housing conditions at 21°C ± 0.5°C, 55% ± 5% humidity, and 12h artificial light-dark cycle (150 lux). Mice

were provided with standard rodent chow (Special Diet Services, RM1 801151, Scanbur BK, Oslo Norway) and water *ab libitium*.

To mimic both obese and non-obese environments, 6 weeks old female littermates were randomly assigned to chow and HFD groups and fed either standard chow diet (7.5% kcal from fat, 17.5% from proteins and 75% from carbohydrates, Special Diet Services RM1, 801151) or high fat containing diets (60% kcal from fat, 20% from protein and 20% from carbohydrates, Research Diets, D12492) for 10 weeks prior to tumor cell implantations. Body weight was monitored every week. The respective diets were maintained throughout the experiment.

Cell Lines and Culture

MDA-MB-231 (TNBC, human), HCC1806 (TNBC, human) and HEK293T cell lines were purchased from the American Type Culture Collection (ATCC). E0771 (TNBC, mouse) cell line was purchased from the CH3 BioSystems. TeLi (basal breast cancer, mouse) cells were originally derived from a tumor formed in MMTV-Wnt1 transgenic mouse and then propagated *in vivo* for four generations through mammary fat pad injections before being passaged *in vitro*. Tumors were dissociated using Mouse tumor dissociation kit (Miltenyi Biotec, 130-096-730) according to manufacturer's instructions. Dissociated tumor cells were cultured *in vitro* for two months to obtain pure tumor cells. The *in vivo* passaged MMTV-Wnt cells were kindly provided by Stein-Ove Døskeland, University of Bergen. MDA-MB-231, E0771 and TeLi cells were cultured at 37°C, 5% CO_2 in high-glucose DMEM (Sigma, D5671) supplemented with 10% FBS (Sigma, F-7524), 100U/mL penicillin and 100 µg/mL streptomycin (Sigma, P-0781) and 4 mM L-glutamine (Sigma, G-7513). HCC1806 cells were cultured in RPMI1640 (Sigma, R8758) supplemented with 10% FBS and 100U/mL penicillin/ and 100 µg/mL streptomycin.

For cell line authentication, MDA-MB-231 cells were harvested for genomic DNA extraction using Genomic DNA isolation kit (Norgen Biotek, 24700). Isolated genomic DNA was analyzed by Eurofins Genomics laboratory and the cell line authenticated based on genetic fingerprinting and short tandem repeat (STR) profiling.

Patient Tissue Microarray and Transcriptomic Analysis

Tissue Microarray

Tissue specimens were from the human breast cancer patient cohort described above[15]. At the time of diagnosis, each patient from the study cohort had an incisional tumor biopsy. All tissue samples were fixed in formaldehyde for paraffin embedding in the operating theatre immediately on removal. Paraffin embedded tissues were subject to tissue microarray (TMA) construction. From each tumor, 4 cores of 1.2 mm diameter from tumor rich areas were punched out using Manual Tissue Arrayer Punchers (MP10; Beecher Instruments). The patient cores were embedded into ten 8 x 10 array blocks plus 1 to 2 liver control cores for orientation. Microtome sectioned slides were stored at 4°C until ready for use.

Immunohistochemical staining for CD133 was done as described previously[17]. In short, slides were dried at 58°C over two days and deparaffinization was performed using xylene, rehydrated with ethanol and dH_2O. Target retrieval was done in Tris/EDTA buffer, pH 9 (Dako, S2367) in a microwave for 25 min. Slides with buffers were cooled down at room temperature for 15 min, followed by rinsing with cold dH_2O. Samples were then blocked in the Peroxidase Blocking solution (Dako REAL, S2023) for 8 min, rinsed with water and then blocked in a serum-free protein block buffer for 8 min (Dako, X0909). Primary CD133 antibody (Miltenyi Biotec, 130-090-422) was diluted 1:25 in Antibody Diluent with Background Reducing Components (Dako, S3022). 200 µL of antibody solution was put on each slide to cover all TMA specimens and incubated overnight at 4°C. The following day, slides were washed twice with Dako Wash Buffer (S3006). Primary antibody detection was performed using MACH3 mouse probe (Biocare Medical) followed by MACH3 HRP polymer (Biocare Medical, BC-M3M530H) and the signal was developed with diamino-benzidine DAB+ (Dako, K3468). Finally, the slides were counterstained with hematoxylin (Dako, S3301), dehydrated in alcohol solutions and xylene, and mounted in Pertex Mount Agent (Histolab, 00801).

Immunohistochemical staining for Axl was done as described previously[68]. In short, slides were dried at 58°C over two days and deparaffinization was performed using xylene, rehydrated with ethanol and dH_2O. Target retrieval was done in Tris/EDTA buffer, pH 6 (Dako, S2367) in a microwave for 25 min. Slides with buffers were cooled

down at room temperature for 15 min, followed by rinsing with cold dH_2O. Samples were then blocked in the Peroxidase Blocking solution (Dako REAL, S2023) for 15 min, rinsed with water and then blocked in a serum-free protein block buffer for 15 min (Dako, X0909). Primary Axl antibody (RnD Systems, AF154) was diluted 1:400 in Antibody Diluent with Background Reducing Components (Dako, S3022). 200 µL of antibody solution was put on each slide to cover all TMA specimens and incubated overnight at 4°C. The following day, secondary Rabbit anti-Goat IgG control antibodies were added on the slides in 1:400 dilution in DAKO Antibody Diluent for 30 min at room temperature. Further, slides were washed twice with Dako Wash Buffer (S3006). Secondary antibody detection was performed using DAKO EnVision (DAKO, K4003) followed by the signal development by diamino-benzidine DAB+ (Dako, K3468). Finally, the slides were counterstained with hematoxylin (Dako, S3301), dehydrated in alcohol solutions and xylene, and mounted in Pertex Mount Agent (Histolab, 00801).

Transcriptomics

mRNA expression levels were extracted from previously reported microarray analyses[69]. In brief, these analyses were performed on a Human HT-12-v4 BeadChip (Illumina) after labeling (Ambion; Aros Applied Biotechnology). Illumina BeadArray Reader (Illumina) and the Bead Scan Software (Illumina) were used to scan BeadChips. Expression signals from the beads were normalized and further processed as previously described[70]. The data set was re-annotated using illuminaHumanv4.db from AnnotationDbi package, built under Bioconductor 3.3 in R[71], to select only probes with "perfect" annotation[72]. The probes represented 21043 identified and unique genes.

Sequencing of 360 cancer related genes

Targeted sequencing of 360 cancer genes, was performed and described previously[28]. In brief, native, genomic DNA from tumor was fragmented and subjected to Illumina DNA sequencing library preparation. Libraries were then hybridized to custom RNA baits according to the Agilent SureSelect protocol. Paired-end, 75bp sequence reads were generated. Sequencing coverage for the targeted regions (average per bp) within

each sample was >120x for all samples (mean 439x). Supplemental Table S1 lists the included 360 genes.

Proliferation assay

Cell proliferation assay was determined by high-content imaging using the IncuCyte Zoom (Essen Bioscience) according to the manufacturer's instructions. In all experiments, cells were seeded into a 96-well culture plate and for each well four fields were imaged under 10x magnification every 2 h. The IncuCyte Zoom (v2018A) software was used to calculate confluency values.

Glucose and insulin measurements

For glucose and insulin measurements, mice were fasted overnight (9 hours) with free access to water. Blood glucose concentrations were determined using Accu-Check Aviva glucometer (Roche). For insulin measurements, blood was collected from the tail using EDTA coated capillary tubes (Fisher Scientific, 11383994), stored on ice before centrifuged at 2000 g, 4°C for 10 min. Plasma insulin concentrations was determined in duplicates using the Ultra Sensitive Mouse Insulin ELISA Kit (Crystal Chem, 90080) following the manufactures instructions for wide range measures.

Glucose tolerance test

For glucose tolerance test, mice fed a HFD or chow-diet for 10 weeks were fasted overnight (15 hours) with free access to water. Glucose (2.5 g/kg) was administered by gavage, and blood glucose concentrations were determined by using Accu-Check Aviva glucometer (Roche).

Mammary Fat Pad Implantations

E0771 or TeLi cells were prepared in PBS and mixed 1:1 by volume with Matrigel (Corning, 356231) and orthotopically implanted into the fourth inguinal mammary fat

pad of chow and HFD-fed mice in a total volume of 50µL. Tumor diameters (width and length) were measured 2-3 times per week with caliper. Tumor volumes were calculated using formula Tumor volume (mm^3) = Width x $Length^2$ x π/6. Tumors were considered established when the volumes were larger than $50mm^3$.

Cellular Adaptation to PA

Cells were seeded on 10 cm culture dishes so that the confluency at the starting day of adaptation was 80-90%. To start adaptation, all media was removed and replaced by growth media supplemented with palmitic acid (PA) (Sigma, P5585) and 1% fatty acid free BSA as a carrier (Sigma, A7030). Adaptation was done using gradual increase in PA concentration MDA-MB-231 (50 µM, 200 µM and 400 µM, HCC1806 (200 µM and 400 µM), E0771 (200 µM, 400 µM, 500 µM) ensuring around 50% of cell death at each step. Parental cells were cultured in parallel using growth media supplemented only with 1% fatty acid free BSA (Sigma, A7030).

For PA supplemented media, PA was first dissolved in absolute ethanol to obtain a 50mM stock. To prepare the working concentrations, certain volumes of PA stock were added into 1%BSA growth media and incubated at 37°C for 1 hour. PA stock was stored at 4°C and used for no longer than 2 weeks.

Generation of knockdown and overexpressing cell lines

Short hairpin RNAs (shRNA) for target genes and scramble (shCtrl) were purchased from Sigma. pBabe-puro plasmids containing human C/EBPB LAP2 and LIP isoforms were from Addgene (Cat.# 15712 and 15713).

For production of virus, HEK293T cells were seeded onto 10 cm plates to reach 80% confluency on the following day. For retroviral overexpression, 12µg of Gag/Pol plasmid, 6µg of VSVG plasmid and 12 µg of pBabe-puro plasmid containing C/EBPB isoforms were respectively co-transfected into the HEK293T cells using 60 µL Lipofectamine 2000 according to manufacturer's protocol. For lentiviral-mediated depletion of target genes, cells were transfected with 12µg Gag/Pol plasmid, 6µg envelope plasmid and 12µg shRNA containing plasmid (pLKO).

6 hours following transfection, the media was replaced with fresh media. The virus was harvested 48hours post transfection by spinning the collected culture media for 5 mins at 1200 rpm and then filtered through a 0.22 μm filter to completely remove cell debris. The virus was then stored at -20°C for several days or at -80°C for several months.

To infect target cells, 5mL of the appropriate virus was used to infect a subconfluent 10 cm cell culture dish in the presence of 10 μg/mL of polybrene overnight. 48 hours after infection, puromycin was added to select for successfully infected cells: 4 μg/mL for TeLi, 2 μg/mL for MDA-MB-231 and E0771 and 1.33 μg/mL for HCC1806 cells. Uninfected control cells were processed the same way to determine the endpoint of selection. Typically, selection took 2-3 days for all cell lines. After the end of selection cells were released from puromycin for at least 1 day before starting experiments.

Tumorsphere Formation Assay

Assay was performed as previously described[13,22,73]. Cells were harvested using Accutase and re-suspended in PBS. After counting, indicated number of cells (1000 cells/well for E0771 and HCC1806 cell lines, 3000 cells/well for MDA-MB-231 cell lines) were seeded into ultra-low attachment 6-well plates (Corning) in the stem cell media (DMEM/F12 with 20 ng/mL EGF, 20 ng/mL bFGF, 1x B27 supplement). Tumorspheres were quantified after 5 days (for E0771 cell line) or 7-10 days incubation (for HCC1806 and MDA-MB-231 cell lines). Tumorspheres were counted when the size is larger than 60 μm for HCC1806 and MDA-MB-231 cells or 100μm for E0771 cells. For tumorsphere propagation assay, tumorspheres were harvested and trypsinized by using Accutase to obtain single cells. Cells were counted and seeded (500 cells/well for E0771 cells and 1000 cells/well for HCC1806 cells) into ultra-low attachment 6-well plates in the stem cell media. Tumorspheres were imaged and quantified after 5-7 days.

Apoptosis

Analysis of apoptosis was performed using Alexa Fluor™ 488 conjugate Annexin V (Thermo Fisher, A13201) and propidium iodide (PI) according to the manufacturer's instructions. Shortly, cells and their culture media were harvested and washed once in cold PBS. Cells were then resuspended in Annexin binding buffer (10 mM HEPES, 140 mM NaCl, and 2.5 mM $CaCl_2$, pH 7.4) in a concentration of 1×10^6 cells/mL. To each 100 µL of cell suspension 5 µL of the Annexin V and 2 µL PI (at final concentration 2 µg/mL) was added. Cells were incubated in the dark at room temperature for 15 min. After the incubation period, 400 µL of Annexin binding buffer was added and cells were analyzed by flow cytometry (BD LSR Fortessa).

Flow cytometry analysis

For immunostaining for flow cytometry, cells were collected using Accutase (Sigma, A6964) and washed once in PBS. 1×10^6 cells per sample were stained with 0.6 µL of APC conjugated CD133 antibody (Invitrogen, 17-1331-81) and 1 µL of FITC conjugated CD44 antibody (BioLegend, 338803) in 100 µL of 1% BSA supplemented PBS solution and incubated in dark for 20 mins at room temperature. After incubation, cells were washed once with 5 mL of PBS/1% BSA and analyzed on flow cytometry (BD LSR Fortessa). To gate the $CD44^{high}/CD133^{+}$ cell population, the median fluorescent intensity (MFI) of CD44-FITC was measured on control replicates (termed parental cells in Figure S2F-G and overexpression control cells in Figure S4E-F). The average value of CD44-FITC MFI was used to gate $CD44^{high}$ cells, and $CD133^{+}$ cells were gated according to the negative staining samples.

Immunofluorescent analysis

Cells were seeded in 24-well plates on Poly-L-lysin treated cover slips at 75 000 cells per well one day before the staining. On the day of the analysis, culture media was removed and 4% paraformaldehyde (PFA) in Distilled-PBS (DPBS) was added to fix cells for 20 minutes. Then PFA was removed and cells were permeabilized in 0.4% Tween/DPBS for 10 min at RT. This was followed by 3 washes in DPBS. Blocking was performed in 3%BSA/0.2% Tween/DPBS for 90 min. Slides were shortly washed in staining media containing DPBS/0.2% Tween/1.5% BSA. Then slides were covered

by 500 µL of staining media with C/EBPB antibodies (1:100 dilution) and incubated overnight at 4°C. Next day, slides were washed in DPBS 3 x 5 min and incubated with secondary antibodies (1:500 dilution) for 2 hours. This was followed by 5 min wash in DPBS, then 5 min incubation with DAPI (1:500 in DPBS) and then another wash in DPBS. Further, slides were rinsed in distilled water and mounted with ProLong™ Diamond Antifade Mountant. Slides were dried overnight and imaged using Leica SP5 with 63x magnification.Image quantification was performed using Fiji software. The nucleus and whole cell were demarcated based on DAPI and bright field, respectively. % nuclear C/EBPB were calculated by diving the nuclear signal by whole cell signal multiplied by 100.

Fatty Acid and Glucose Oxidation Assay

Fatty acid and glucose oxidation were assessed by providing ^{14}C-labled palmitic acid or glucose to the cells, with subsequent capture of the released $^{14}CO_2$; a technique previously described[74]. In brief, cells were plated into 96-well tissue culture plates (MDA-MB-231, 45000 cells/well; HCC1806, 45000 cells/well; dissociated E0771, 25000 cells/well) in corresponding growth medium and incubated overnight to allow proper attachment. Radiolabeled [1-^{14}C] palmitic acid (1 µCi/mL) and D-[^{14}C(U)] glucose (1 µCi/mL) were given in PBS supplemented with 10mM HEPES and 1mM L-carnitine. Respective amounts of non-radiolabeled substrate were added to obtain final concentrations of D-glucose (5 mM) and BSA-conjugated palmitic acid (100 µM). Etomoxir (40 µM) was added to certain wells during palmitic acid oxidation, to monitor the non-mitochondrial CO_2 production. An UniFilter®-96w GF/B microplate was activated for capture of CO_2 by the addition of 1M NaOH (25 µL/well) and sealed to the top of the 96-well tissue culture plates and incubated for the indicated period of time at 37°C. Subsequently, 30 µL scintillation liquid (MicroScint PS PerkinElmer) was added to the filters and the filter plate was sealed with a TopSealA (PerkinElmer). Radioactivity was measured using MicroBeta2 Microplate Counter (PerkinElmer). Protein measurement was performed for data normalization. The cells were washed twice with PBS, lysed by 0.1 M NaOH, and protein was measured using Pierce® BCA Protein Assay Kit (Thermo Fisher Scientific, 23225).

RNA extraction, RT-PCR, and qPCR

Total RNA was extracted with a Total RNA purification Kit (NORGEN Biotek, 37500) according to the manufacturer's protocol. cDNA was synthesized from 1 µg total RNA template with oligo-dT primers using a SuperScript® III First-Strand Synthesis kit (ThermoFisher Scientific, 18080-051) according to the manufacturer's protocol. qPCR was carried out in quadruplicates with a LightCycler® 480 SYBR Green I Master Mix (Roche, 04887352001) using a LightCycler® 480 Instrument II (Roche, 05015243001). The results were calculated by ΔΔCt method using human HPRT (h*HPRT*) for human genes and mouse actin (m*Actin*) for mouse genes. Primer sequences are listed in the key resources table.

Transfection of siRNA duplexes

One day before transfection cells were plated on T25 flasks at the density 250 000 cells/flask. After overnight incubation, cells were transfected using Lipofectamine® RNAiMAX according to the manufacturer's protocol with some modifications: we used 3 µL of Lipofectamine per flask and final concentration of siRNAs was 20 nM. After 48 hours incubation cells were harvested and seeded for tumorsphere formation assay.

Western blotting

Cells were lysed in RIPA lysis buffer (Thermo Scientific, 89901) complemented with protease inhibitor cocktail (cOmplete ULTRA Tablets, MINI, EDTA-free, EASYpack, 05892 791001) and phosphatase inhibitor cocktail (PhosStop, 04906837001). After quantification with the Pierce® BCA Protein Assay Kit, equal amounts of protein (typically 20-50 µg of protein per lane) were separated by electrophoresis on a NuPAGE 10% Bis-Tris Gel (Invitrogen, NP0315BOX) in NuPAGE™ MOPS SDS Running Buffer (20X, Invitrogen, NP000102) and then transferred to an activated Immobilon-P PVDF Membrane (Merck Millipore Ltd, IPVH00010 PORE SIZE: 0.45 µm). The membranes were blocked using 5% nonfat dry milk in PBS/0.1% Tween20 for 1h at RT, incubated with indicated primary antibodies for overnight at 4°C. This step was followed by an incubation with secondary IRDye-conjugated antibodies

(Leicor, P/N 925-68070, P/N 926-32213). Detection and quantification were performed on Amersham Typhoon Gel and Blot Imaging Systems. A list of antibodies is given in the key resources table.

RNA sequencing

MDA-MB-231 parental and selected cells were plated at 1×10^5 cells/mL into 6-well plates in corresponding medium. After three days, cells were harvested, and RNA extraction was performed with a Total RNA purification Kit according to the manufacturer's protocol. Potential DNA contaminations were removed by applying the RNA Clean & concentrator with DNaseI kit (Zymo, R1013). RNA sequencing libraries were prepared at the Genomic Core Facility at University of Bergen using Illumina TruSeq Stranded mRNA sample preparation kit according to the manufacturer's instructions and sequenced on the same lane on a HiSeq 4000 sequencer with paired-end 75bp reads.

ATACseq library construction

ATACseq libraries were constructed as previously described[75]. In brief, $5x10^4$ cells were washed once with ice-cold PBS and pelleted by centrifugation. Cells were lysed in 50 µL RSB buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl and 3 mM $MgCl_2$) containing 0.1% NP-40, 0.1% Tween-20 and 0.01% digitonin, and incubated on ice for 3 minutes for permeabilization. After incubation, samples were washed in 1 mL RSB containing 0.1% Tween-20 and pelleted at 500 g for 10 minutes at 4°C. Samples were then resuspended on ice in 50 µL transposition reaction mix containing 2.5 µL Tn5 transposase, 1x TD buffer (both Illumina FC-121-1030), 1x PBS, 0.1% Tween-20 and 0.01% digitonin, and incubated at 37°C for 30 minutes with agitation. Tagmented DNA was purified using Zymo DNA Clean and Concentrator-5 kit (Zymo D4014). The resulting DNA was amplified for 12-13 cycles. The libraries were purified with AMPure XP beads (Beckman A63880), quality-checked on Bioanalyzer (Agilent) and 75 bp paired-end sequenced on Illumina HiSeq 4000 at Genomic Core Facility at University of Bergen.

Cut&Run and library construction

Cut&Run was performed as described with minor modifications[40]. Briefly, $5x10^5$ cells were washed and bound to concanavalin A-coated magnetic beads (Bangs Laboratories, BP531). The cells were then permeabilized with Wash Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine and 1x Roche Complete Protease Inhibitor, EDTA-free) containing 0.025% digitonin (Digitonin Buffer) and 2 mM EDTA and incubated with primary antibody (anti-C/EBPB or IgG isotype control) overnight at 4°C. The cell-bead slurry was washed twice with Digitonin Buffer and incubated with 1x Protein-A/G-MNase (pAG-MNase; Epicypher) in Digitonin Buffer for 10 minutes at room temperature. The slurry was washed twice with Digitonin Buffer and incubated in Digitonin Buffer containing 2 mM $CaCl_2$ for 2 hours at 4°C to activate pAG-MNase digestion. The digestion was stopped by addition of 2x Stop Buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/mL RNase A, 50 µg/mL GlycoBlue and 300 pg/mL in-house MNase-digested yeast spike-in chromatin) and the sample was incubated for 10 minutes at 37°C to release chromatin to the supernatant and degrade RNA. The supernatant was recovered, and DNA was isolated through phenol-chloroform extraction and ethanol precipitation. Libraries were constructed to enrich for sub-nucleosomal fragments using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina as described (NEB, E7645S). The libraries were size-selected and purified with AMPure XP beads, quality-checked on Tapestation (Agilent) and 100 bp or 75 bp paired-end sequenced on MiSeq or HiSeq 4000 at Genomic Core Facility at University of Bergen.

ClonTracer barcoding of cancer cell lines and in vivo implantation

The ClonTracer barcoding library was obtained from Addgene (#67267). The library was electrotransformed as described[25], expanded, extracted and pooled together. For viral production, HEK293T cells were transfected with 12 µg Gag/Pol plasmid, 6 µg envelope plasmid and 12 µg ClonTracer library. Six million E0771 cells were infected by lentiviral ClonTracer barcodes at a multiplicity of infection (MOI) of around 0.1 and infected cells were selected with puromycin. After selection, infected cells were pooled

and expended in vitro. At the day of injection, cells were harvested and counted. 10 000 cells were suspended with 50% (v/v) matrigel and injected into each chow or HFD-fed mice and three replicates were set up for each condition. And five million cells were washed with PBS and the cell pellet was stored at -80°C as a pre-injection control for further process. At 18 days post-injection, the mice were sacrificed and the snap frozen tumors were stored at -80°C for further process.

Barcode amplification and sequencing

For each sample, the frozen tumor was crushed and the tumor pieces from different areas (tumor core, intermediate and peripheral layers) were weighed for DNA extraction. Genomic DNA was extracted from around 130mg frozen tumor tissues cell with a Tissue DNA Kit (E.Z.N.A). PCR was used to amplify the barcode sequence for NGS, and PCR primer sequences information is listed in supplementary information (Supplementary Table S3). For each PCR reaction, 2 µg of genomic DNA was used as a template and eight parallel PCR reactions were set up to ensure the sampling of sufficient template coverage. PCR products were cleaned up by PCR Purification Kit (QIAGEN) and further purified with AMPure XP beads (Beckman A63880). After purification and quality control, the samples were 75 bp paired-end sequenced on MiSeq Genomic Core Facility at University of Bergen.

Mass Cytometry

Cells were plated in 10 cm plates in triplicates to reach a confluency of 80% after 48 hours. For the analysis, cells were collected using TrypLE Express (Gibco, 12604-021). $1x10^6$ cells per condition were included. Cells were resuspended in cell culture media and treated with 0.25 µM Cisplatin for 5 min at RT. Further, cells were fixed in 1 mL of 1.6% PFA (Electron Microscopy Sciences, 15710) in PBS for 10 min at RT. Cells were pelleted by centrifugation for 5 min at 900g and the pellets were stored at -80°C until staining with heavy metal tagged antibodies. On the day of staining, samples were thawed on ice, resuspended in 500 µL of DPBS (Gibco, 14040-133) and incubated for 10 min at RT in DPBS/DNases (Sigma, DN25) solution. Next, cells

were washed in D-WASH solution (DPBS + 1% FA-free BSA + 0.02% NaN3 + 0.25 mg/mL DNase) and barcoded (Fluidigm, 201060) according to the manufacturer's protocol. Cells were then washed twice in the Maxpar Cell Staining Buffer (Fluidigm, 201068), all samples were combined and labeled with surface antibody cocktail (Figure S2A, extracellular) for 30 min at RT. Further, cells were pelleted by centrifugation and incubated in 4 mL of DPBS/DNase solution for 10 min at RT. After this step cells were washed in PBS-EDTA and fixed in 2% PFA/PBS (filtered through a 0.22 μm filter) for 30 min RT, followed by wash in Cell Staining Buffer and permeabilization in cold methanol (-20°C) for 10 min. After incubation, cells were washed in once PBS, once in D-WASH and labeled with intracellular antibody cocktail (Supplementary Table S2, intracellular) for 30 min at RT. This was followed by incubation of cells in D-WASH for 10 min at RT and double wash in D-WASH. Then cells were incubated in 2% PFA/PBS with Cell-ID™ Intercalator-Ir (Fludigm, 201192B) at 4°C overnight. The samples were spun down the following day and incubated in D-WASH for 10 min at RT, washed once in PBS/EDTA, 3 times in Maxpar Water (Fluidigm, 201069), resuspended in EQ Four Element Calibration Beads (Fluidigm, 201078) diluted 1:9 in water, and acquired on the Helios - Mass Cytometer.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mass cytometry Data Analysis

FCS-files were normalized, concatenated and debarcoded in R using CATALYST[76]. Samples from the different conditions were subsampled (85000 for HCC1806 cell lines and 100000 per sample MDA-MB-231 cell lines) prior to analysis. Dimensionality reduction with tSNE, density plotting and pseudo coloring were performed in Cytobank[67].

Survival Analysis

Patients were stratified into two groups by BMI 25. Disease-Specific survival (DSS) Kaplan-Meier curves were generated using GraphPad Prism software and statistical significance was calculated using Log-rank (Mantel-Cox) test.

Mutual Information

Mutual information was calculated as described in Goodarzi 2009[35].

Tissue Microarray Analysis

The ten CD133-stained and ten Axl-stained TMA slides were scanned with an Aperio Scanscope CS Slide Scanner. The breast cancer cores were 1.2 mm in diameter with up to 4 cores per patient. Full analysis was performed on valid cores for patients 50 years and older with ER and PR negative status. Cores with too few cells, poor quality, excessive tearing, or folding were not considered valid and were omitted from analysis.

QuPath (version 0.2.0-m5) was used to dearray the TMAs, segment cells, and classify cell types. The following detection steps and parameters were applied to all TMA slides. Simple tissue detection was used to find the approximate tissue borders within each dearrayed TMA core. For CD133-stained cores, a threshold of 229 (default 127), requested pixel size of 1 μm (default 20 μm), and checking the box for Expand boundaries were found to be the most important parameter setting changes for accurate tissue detection.

Watershed cell detection was used to create cell masks within the detected tissue of each valid core. The watershed parameters were optimized to detect large weakly hematoxylin-stained cancer cells, to minimize false positive cell detection from areas of high background signal, and to reduce the creation of cell masks that spanned multiple cells. The watershed parameter changes deemed most important for accurate cell mask creation were: nucleus background radius of 10 μm (default 8 μm), nucleus minimum area of 24 μm^2 (default 10 μm^2), nucleus maximum area of 230 μm^2 (default 400 μm^2), intensity parameters for threshold and max background both set to 0.07, and exclusion of DAB staining (as was recommended for membrane staining markers). Additionally, the cell expansion was set to 10 μm, 5 μm larger than the default setting, in order to capture the CD133 membrane staining on the large cancer cells.

Annotation objects were drawn around easily defined areas that contained primarily cancer cells, non-cancer cells, or platelets/RBCs and labeled as the classes tumor,

stroma, or ignore, respectively. Platelets/RBCs were ignored because they appeared brown even before staining and show up as falsely positive for CD133. 9039 cells from the annotation objects drawn across 5 of the 10 slides were used to train the random forest (trees) classifier in QuPath. DAB specific measurements were excluded from the classifier selected features. The intensity feature used to identify CD133 positive cells was Cell: DAB OD max at a threshold of 0.45. With these parameters, the detection classifier created seven classification groups of cells: total (base) tumor cells, total stroma cells, $CD133^+$ tumor cells, $CD133^-$ tumor cells, $CD133^+$ stroma cells, $CD133^-$ stroma cells, and ignored cells. Cell masks from cores with partial low quality due to folding or poor imaging were removed to prevent false positive cells. All cores were visually inspected for false positive cancer cell masks and false positive masks were removed. Mean $CD133^+$ cancer cell percentage was calculated for each patient for all valid tumor cores by QuPath and exported to MS Excel. Patients with greater than 2% CD133 positive cancer cells were considered to have CD133 positive tumors.

For Axl signal analysis, a few parameters were modified from CD133-staining analysis due to the difference of staining pattern. A threshold of 228 was used for simple tissue detection. The modified watershed parameters were: nucleus maximum area of 400 μm^2, max background intensity set to 2 and no exclusion of DAB. The total number of training objects used to train the random forest classifier were 7230 from across all the 10 slides. The ninetieth percentile of the Cell: DAB OD max intensity feature for all cells in 8 of 10 TMAs were used as a threshold to identify Axl^{high} cells. This resulted in Cell: DAB OD max threshold set to 1.46492. With these parameters, we identified the detection classifier created seven classification groups of cells: total (base) tumor cells, total stromal cells, Axl^{high} tumor cells, $Axl^{-/low}$ tumor cells, Axl^{high} stroma cells, $Axl^{-/low}$ stroma cells and ignored cells. The number of detected cells in each cell group for each patient core were exported from QuPath. In MS Excel, this data was used to find the percentage of Axl^{high} cells of all tumor cells present for each ER^-PR^- postmenopausal patient. An outlier value deviating more than 2 times the standard deviation in the Axl^{high} BMI ≤ 25 group was excluded prior to statistics performed. Statistical analysis was performed in GraphPad Prism.

Student's t-test

Statistical analysis of flow cytometry data was performed using Student's t-test on GraphPad Prism 8 software.

Limiting dilution analysis

The frequency of tumor initiating cells was calculated using the Extreme Limiting Dilution Analysis (ELDA) (http://bioinf.wehi.edu.au/software/elda/index.html)[77].

RNA sequencing data analysis

Sequenced reads were quality checked with FastQC and aligned to the UCSC hg19 reference genome with Hisat2. Aligned reads were counted and summarized for the annotated genes using featureCounts. Differential gene expression analysis was performed by DESeq2. For visualization, read counts were normalized and regularized log transformed (rlog) for cross-sample comparison or were converted to fragments per kilobase of transcript per million mapped reads (FPKM) for within-sample comparison.

ATACseq data analysis

ATACseq reads were quality-checked with FastQC[78] before and after adapter trimming with Trimmomatic[79]. The trimmed reads were aligned to the UCSC hg19 or mm10 reference genome using Bowtie2[80] with the parameters --phred33 --end-to-end --very-sensitive -X 2000. Reads were then removed if they were mapped to the mitochondria and non-assembled contigs, had a mapping quality score below 10 and were PCR duplicates. Read start sites were adjusted for Tn5 insertion by offsetting +stand by +4 bp and -strand by -5 bp as previously described[67]. For peak calling, MACS2[81] was used with the parameters -q 0.01 --nomodel. Peaks residing in the ENCODE blacklisted regions were removed for further downstream analysis. deepTools[82] was used to generate 1x normalized bigwig files for visualization.

Peaks unique to or shared across conditions were identified using the occupancy mode in DiffBind[83]. Subsequently differential analysis was performed on these peaks

with default settings, and annotated genome-wide with respect to the closest transcription start site with ChIPseeker[84]. To infer differential transcription factor binding motif activity between conditions, diffTF[37] was used. Input transcription factor binding sites for 640 human transcription factors were generated as described using the HOCOMOCO database and PWMscan (cutoff p-value - 0.00001, background base composition - 0.29;0.21;0.21;0.29).

To infer the differentiation state, chromVar was applied to the consensus peakset in E0771 *ex vivo* cells and extended to published data[85] on mammary gland development to generate a matrix of average transcription factor binding motif activity of each cell type. The data matrix was then mean-centered and used for principal component analysis (PCA). The difference from the fetal mammary stem cell stage was quantified by the differences in z-scores between E0771 HFD and Chow from the mean PC1 and PC2 of fetal mammary stem cell stage. To visualize the motif activity of specific transcription factors along the mammary gland developmental trajectory, snATACseq data on mammary gland development was analyzed as originally described[86] and presented in pseudotime.

Cut&Run data analysis

Cut&Run reads were quality-checked with FastQC before and after adapter trimming with Trimmomatic. The trimmed reads were separately aligned to the UCSC hg19 and sacCer3 reference genomes using Bowtie2 with the parameters --local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700 and --local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700 --no-overlap --no-dovetail, respectively. Reads were then removed if they were mapped to the mitochondria and non-assembled contigs and had a mapping quality score below 10. Mapped reads were converted to paired-end BED files containing coordinates for the termini of each read pair and the fragment length, and calibrated to the yeast spike-in using spike_in_calibration.csh (https://github.com/Henikoff/Cut-and-Run/) in bedgraph formats for visualization. Peaks were called with SEACR[87] with respect to the IgG control using the non and stringent mode. Peaks overlapping with the ENCODE blacklisted regions were removed for further downstream analysis.

To identify enriched motif sequences protected by transcription factor binding independent of the peak calling algorithm, pA/G-MNase cutting footprints were detected. Ends of all CUT&RUN fragments ≤ 120 bp were enumerated to determine the precise single base pair cut sites and sorted. Footprints were detected using Footprint Occupancy Score (FOS)[43]. Significant footprints with FOS ≤ 1 were analyzed for enriched motif sequences with HOMER using the position weight matrices (PWMs) from the HOCOMOCO database[88].

Peaks unique to or shared across conditions were identified using the occupancy mode in DiffBind to generate a consensus peakset. Raw counts of this peakset across samples were input to DESeq2 with the inverse of the spike-in calibration factors as sizeFactors to perform differential analysis. Differential peaks were annotated respect to the closest transcription start site with ChIPseeker. HOMER was used for *de novo* motif analysis on differential peaks unique to each condition, and the identified motifs were compared to the HOCOMOCO database for best matches.

ClonTracer Barcode Analysis

Barcode-composition analysis was carried out by using the python package clonTracer v1.2[25]. As previously described, only barcodes passing all the quality filters and seen at least twice are considered for the analysis. To estimate the barcode distribution in each group, we first excluded the unique barcodes which were not presented in the pre-injection control cell sample from tumor samples and further a threshold of minimum ten reads per unique barcode was set up to identify the barcodes in each tumor. After calculating the fractions of barcode in each tumor sample, we pooled the data for all tumors in each group. The overall distributions of relative barcode size for chow and HFD groups were plotted in GraphPad Prism.

Acknowledgements

We thank Erik Løkkevik, Bjørn Østenstad, Steinar Lundgren, Terje Risberg, and Ingvil Mjaaland for providing clinical samples. We thank the genomic score facility (GSF) at

the University of Bergen, which is a part of the NorSeq consortium, provided services on RNAseq, ATACseq and Cut&Run. GSF is supported by grants from the Research Council of Norway (245979/F50) and the Trond Mohn Foundation (BFS2016-genom). The flow cytometry and mass cytometry were performed at the Flow Cytometry Core Facility, Department of Clinical Science, University of Bergen. Helios Mass Cytometer was supported by the Trond Mohn Foundation. We thank Ingeborg Winge from Department of Pathology, Haukeland University Hospital, Bergen, for provided training and help with TMA Immunohistochemistry. Hani Goodarzi is supported by R01CA240984 and R01GM123977. NH was funded by a Starter Grant from the Trond Mohn Foundation, the Norwegian Research Council (275250) and the Norwegian Cancer Society (212734-2019; National Group of Expertise on Pancreatic Cancer).

Authors Contributions

Conceptualization, N.H; Methodology, N.H., X.L., A.R., S.M.G., M.H.C., S.T.T and L.P.; Software, S.M.G., C.E.W., X.L., M.H.C. Validation: T.L., A.R., M.H.C. Formal Analysis, C.E.W., N.M., P.E.L., S.K., H.G., S.D.P., S. T.T., X.L., A.R., M.H.C.; Investigation, A.R., X.L., M.H.C.; Resources, N.H., S.D.P., S.M.G., J.L., S.K., P.E.L., S.K., A.M.; Writing – Original Draft, N.H., X.L., A.R.; Visualization, N.H., A.R., X.L., M.H.C., C.E.W.; Supervision, N.H.; Funding Acquisition, N.H.

Declaration of interests

The authors declare no competing interests.

Data availability

Sequencing data (RNA-seq, ATACseq and Cut&Run) for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB 39793. The mass cytometry data has been deposited in the FLOW Repository under repository ID FR-FCM-Z2TK.

REFERENCES

1 Munsell, M. F., Sprague, B. L., Berry, D. A., Chisholm, G. & Trentham-Dietz, A. Body mass index and breast cancer risk according to postmenopausal estrogen-progestin use and hormone receptor status. *Epidemiol Rev* **36**, 114-136, doi:10.1093/epirev/mxt010 (2014).

2 Jiralerspong, S. & Goodwin, P. J. Obesity and Breast Cancer Prognosis: Evidence, Challenges, and Opportunities. *J Clin Oncol* **34**, 4203-4216, doi:10.1200/JCO.2016.68.4480 (2016).

3 Chan, D. S. *et al.* Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. *Ann Oncol* **25**, 1901-1914, doi:10.1093/annonc/mdu042 (2014).

4 Niraula, S., Ocana, A., Ennis, M. & Goodwin, P. J. Body size and breast cancer prognosis in relation to hormone receptor and menopausal status: a meta-analysis. *Breast Cancer Res Treat* **134**, 769-781, doi:10.1007/s10549-012-2073-x (2012).

5 Picon-Ruiz, M., Morata-Tarifa, C., Valle-Goffin, J. J., Friedman, E. R. & Slingerland, J. M. Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention. *CA Cancer J Clin* **67**, 378-397, doi:10.3322/caac.21405 (2017).

6 Liu, X. Z., Pedersen, L. & Halberg, N. Cellular mechanisms linking cancers to obesity. *Cell Stress* **5**, 55-72, doi:10.15698/cst2021.05.248 (2021).

7 Ringel, A. E. *et al.* Obesity Shapes Metabolism in the Tumor Microenvironment to Suppress Anti-Tumor Immunity. *Cell* **183**, 1848-1866 e1826, doi:10.1016/j.cell.2020.11.009 (2020).

8 Wogsland, C. E. *et al.* High-dimensional immunotyping of tumors grown in obese and non-obese mice. *Dis Model Mech* **14**, doi:10.1242/dmm.048977 (2021).

9 Saxena, N. K. *et al.* Bidirectional crosstalk between leptin and insulin-like growth factor-I signaling promotes invasion and migration of breast cancer cells via transactivation of epidermal growth factor receptor. *Cancer Res* **68**, 9712-9722, doi:10.1158/0008-5472.CAN-08-1952 (2008).

10 Qureshi, R. *et al.* The Major Pre- and Postmenopausal Estrogens Play Opposing Roles in Obesity-Driven Mammary Inflammation and Breast Cancer Development. *Cell Metab* **31**, 1154-1172 e1159, doi:10.1016/j.cmet.2020.05.008 (2020).

11 Madak-Erdogan, Z. *et al.* Free Fatty Acids Rewire Cancer Metabolism in Obesity-Associated Breast Cancer via Estrogen Receptor and mTOR Signaling. *Cancer Res* **79**, 2494-2510, doi:10.1158/0008-5472.CAN-18-2849 (2019).

12 Hao, J. *et al.* Circulating Adipose Fatty Acid Binding Protein Is a New Link Underlying Obesity-Associated Breast/Mammary Tumor Development. *Cell Metab* **28**, 689-705 e685, doi:10.1016/j.cmet.2018.07.006 (2018).

13 Tiwari, P. *et al.* Metabolically activated adipose tissue macrophages link obesity to triple-negative breast cancer. *J Exp Med* **216**, 1345-1358, doi:10.1084/jem.20181616 (2019).

14 Bowers, L. W. *et al.* Leptin Signaling Mediates Obesity-Associated CSC Enrichment and EMT in Preclinical TNBC Models. *Mol Cancer Res* **16**, 869-879, doi:10.1158/1541-7786.MCR-17-0508 (2018).

15 Chrisanthar, R. *et al.* Predictive and prognostic impact of TP53 mutations and MDM2 promoter genotype in primary breast cancer patients treated with epirubicin or paclitaxel. *PLoS One* **6**, e19249, doi:10.1371/journal.pone.0019249 (2011).

16 Park, E. K. *et al.* Transcriptional repression of cancer stem cell marker CD133 by tumor suppressor p53. *Cell Death Dis* **6**, e1964, doi:10.1038/cddis.2015.313 (2015).

17 Immervoll, H., Hoem, D., Sakariassen, P. O., Steffensen, O. J. & Molven, A. Expression of the "stem cell marker" CD133 in pancreas and pancreatic ductal adenocarcinomas. *BMC Cancer* **8**, 48, doi:10.1186/1471-2407-8-48 (2008).

18 Asiedu, M. K. *et al.* AXL induces epithelial-to-mesenchymal transition and regulates the function of breast cancer stem cells. *Oncogene* **33**, 1316-1324, doi:10.1038/onc.2013.57 (2014).

19 Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878, doi:10.1038/s41598-017-17204-5 (2017).
20 Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396-401, doi:10.1038/nature03128 (2004).
21 Charafe-Jauffret, E. *et al.* Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer Res* **69**, 1302-1313, doi:10.1158/0008-5472.CAN-08-2741 (2009).
22 Lee, N. H. *et al.* SERPINB2 Is a Novel Indicator of Cancer Stem Cell Tumorigenicity in Multiple Cancer Types. *Cancers (Basel)* **11**, doi:10.3390/cancers11040499 (2019).
23 Dupuy, F. *et al.* PDK1-Dependent Metabolic Reprogramming Dictates Metastatic Potential in Breast Cancer. *Cell Metab* **22**, 577-589, doi:10.1016/j.cmet.2015.08.007 (2015).
24 O'Flanagan, C. H. *et al.* Metabolic reprogramming underlies metastatic potential in an obesity-responsive murine model of metastatic triple negative breast cancer. *NPJ Breast Cancer* **3**, 26, doi:10.1038/s41523-017-0027-5 (2017).
25 Bhang, H. E. *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21**, 440-448, doi:10.1038/nm.3841 (2015).
26 Furukawa, S. *et al.* Increased oxidative stress in obesity and its impact on metabolic syndrome. *J Clin Invest* **114**, 1752-1761, doi:10.1172/JCI21625 (2004).
27 Sabharwal, S. S. & Schumacker, P. T. Mitochondrial ROS in cancer: initiators, amplifiers or an Achilles' heel? *Nat Rev Cancer* **14**, 709-721, doi:10.1038/nrc3803 (2014).
28 Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751-759, doi:10.1038/nm.3886 (2015).
29 Karpe, F., Dickmann, J. R. & Frayn, K. N. Fatty acids, obesity, and insulin resistance: time for a reevaluation. *Diabetes* **60**, 2441-2449, doi:10.2337/db11-0425 (2011).
30 Tremblay, A. J. *et al.* Associations between the fatty acid content of triglyceride, visceral adipose tissue accumulation, and components of the insulin resistance syndrome. *Metabolism* **53**, 310-317, doi:10.1016/j.metabol.2003.10.011 (2004).
31 Korbecki, J. & Bajdak-Rusinek, K. The effect of palmitic acid on inflammatory response in macrophages: an overview of molecular mechanisms. *Inflamm Res* **68**, 915-932, doi:10.1007/s00011-019-01273-5 (2019).
32 Saadatian-Elahi, M. *et al.* Serum fatty acids and risk of breast cancer in a nested case-control study of the New York University Women's Health Study. *Cancer Epidemiol Biomarkers Prev* **11**, 1353-1360 (2002).
33 Saadatian-Elahi, M., Norat, T., Goudable, J. & Riboli, E. Biomarkers of dietary fatty acid intake and the risk of breast cancer: a meta-analysis. *Int J Cancer* **111**, 584-591, doi:10.1002/ijc.20284 (2004).
34 Draffin, J. E., McFarlane, S., Hill, A., Johnston, P. G. & Waugh, D. J. CD44 potentiates the adherence of metastatic prostate and breast cancer cells to bone marrow endothelial cells. *Cancer Res* **64**, 5702-5711, doi:10.1158/0008-5472.CAN-04-0389 (2004).
35 Goodarzi, H., Elemento, O. & Tavazoie, S. Revealing global regulatory perturbations across human cancers. *Mol Cell* **36**, 900-911, doi:10.1016/j.molcel.2009.11.016 (2009).
36 Ly, C. H., Lynch, G. S. & Ryall, J. G. A Metabolic Roadmap for Somatic Stem Cell Fate. *Cell Metab* **31**, 1052-1067, doi:10.1016/j.cmet.2020.04.022 (2020).
37 Berest, I. *et al.* Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Rep* **29**, 3147-3159 e3112, doi:10.1016/j.celrep.2019.10.106 (2019).
38 LaMarca, H. L. *et al.* CCAAT/enhancer binding protein beta regulates stem cell activity and specifies luminal cell fate in the mammary gland. *Stem Cells* **28**, 535-544, doi:10.1002/stem.297 (2010).

39 Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* **7**, a025064, doi:10.1101/cshperspect.a025064 (2015).
40 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, doi:10.7554/eLife.21856 (2017).
41 Descombes, P. & Schibler, U. A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* **67**, 569-579, doi:10.1016/0092-8674(91)90531-3 (1991).
42 Xiong, W., Hsieh, C. C., Kurtz, A. J., Rabek, J. P. & Papaconstantinou, J. Regulation of CCAAT/enhancer-binding protein-beta isoform synthesis by alternative translational initiation at multiple AUG start sites. *Nucleic Acids Res* **29**, 3087-3098, doi:10.1093/nar/29.14.3087 (2001).
43 Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90, doi:10.1038/nature11212 (2012).
44 Suter, D. M. Transcription Factors and DNA Play Hide and Seek. *Trends Cell Biol* **30**, 491-500, doi:10.1016/j.tcb.2020.03.003 (2020).
45 Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, doi:10.1093/database/bax028 (2017).
46 Neuhouser, M. L. *et al.* Overweight, Obesity, and Postmenopausal Invasive Breast Cancer Risk: A Secondary Analysis of the Women's Health Initiative Randomized Clinical Trials. *JAMA Oncol* **1**, 611-621, doi:10.1001/jamaoncol.2015.1546 (2015).
47 Gao, Y. *et al.* Adipocytes promote breast tumorigenesis through TAZ-dependent secretion of Resistin. *Proc Natl Acad Sci U S A* **117**, 33295-33304, doi:10.1073/pnas.2005950117 (2020).
48 Chung, K. M. *et al.* Endocrine-Exocrine Signaling Drives Obesity-Associated Pancreatic Ductal Adenocarcinoma. *Cell* **181**, 832-847 e818, doi:10.1016/j.cell.2020.03.062 (2020).
49 Qin, Y., Grimm, S. A., Roberts, J. D., Chrysovergis, K. & Wade, P. A. Alterations in promoter interaction landscape and transcriptional network underlying metabolic adaptation to diet. *Nat Commun* **11**, 962, doi:10.1038/s41467-020-14796-x (2020).
50 Reid, M. A., Dai, Z. & Locasale, J. W. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat Cell Biol* **19**, 1298-1306, doi:10.1038/ncb3629 (2017).
51 Li, X., Egervari, G., Wang, Y., Berger, S. L. & Lu, Z. Regulation of chromatin and gene expression by metabolic enzymes and metabolites. *Nat Rev Mol Cell Biol* **19**, 563-578, doi:10.1038/s41580-018-0029-7 (2018).
52 McDonnell, E. *et al.* Lipids Reprogram Metabolism to Become a Major Carbon Source for Histone Acetylation. *Cell Rep* **17**, 1463-1472, doi:10.1016/j.celrep.2016.10.012 (2016).
53 Bundy, L. M. & Sealy, L. CCAAT/enhancer binding protein beta (C/EBPbeta)-2 transforms normal mammary epithelial cells and induces epithelial to mesenchymal transition in culture. *Oncogene* **22**, 869-883, doi:10.1038/sj.onc.1206216 (2003).
54 Nallanthighal, S., Heiserman, J. P. & Cheon, D. J. The Role of the Extracellular Matrix in Cancer Stemness. *Front Cell Dev Biol* **7**, 86, doi:10.3389/fcell.2019.00086 (2019).
55 Watt, F. M. & Huck, W. T. Role of the extracellular matrix in regulating stem cell fate. *Nat Rev Mol Cell Biol* **14**, 467-473, doi:10.1038/nrm3620 (2013).
56 Seo, B. R. *et al.* Obesity-dependent changes in interstitial ECM mechanics promote breast tumorigenesis. *Sci Transl Med* **7**, 301ra130, doi:10.1126/scitranslmed.3010467 (2015).
57 Catalan, V. *et al.* Increased adipose tissue expression of lipocalin-2 in obesity is related to inflammation and matrix metalloproteinase-2 and metalloproteinase-9 activities in humans. *J Mol Med (Berl)* **87**, 803-813, doi:10.1007/s00109-009-0486-8 (2009).
58 Gomez-Chou, S. B. *et al.* Lipocalin-2 Promotes Pancreatic Ductal Adenocarcinoma by Regulating Inflammation in the Tumor Microenvironment. *Cancer Res* **77**, 2647-2660, doi:10.1158/0008-5472.CAN-16-1986 (2017).

59 Yang, J. *et al.* Lipocalin 2 promotes breast cancer progression. *Proc Natl Acad Sci U S A* **106**, 3913-3918, doi:10.1073/pnas.0810617106 (2009).
60 Dhawan, P. *et al.* Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer. *J Clin Invest* **115**, 1765-1776, doi:10.1172/JCI24543 (2005).
61 Blanchard, A. A. *et al.* Claudins 1, 3, and 4 protein expression in ER negative breast cancer correlates with markers of the basal phenotype. *Virchows Arch* **454**, 647-656, doi:10.1007/s00428-009-0770-6 (2009).
62 Suh, Y. *et al.* Claudin-1 induces epithelial-mesenchymal transition through activation of the c-Abl-ERK signaling pathway in human liver cells. *Oncogene* **32**, 4873-4882, doi:10.1038/onc.2012.505 (2013).
63 Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat Rev Cancer* **7**, 763-777, doi:10.1038/nrc2222 (2007).
64 DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab* **7**, 11-20, doi:10.1016/j.cmet.2007.10.002 (2008).
65 Nieman, K. M. *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat Med* **17**, 1498-1503, doi:10.1038/nm.2492 (2011).
66 Pascual, G. *et al.* Targeting metastasis-initiating cells through the fatty acid receptor CD36. *Nature* **541**, 41-45, doi:10.1038/nature20791 (2017).
67 Kotecha, N., Krutzik, P. O. & Irish, J. M. Web-based analysis and publication of flow cytometry experiments. *Curr Protoc Cytom* **Chapter 10**, Unit10 17, doi:10.1002/0471142956.cy1017s53 (2010).
68 Ahmed, L. *et al.* Novel anti-human Axl monoclonal antibodies for improved patient biomarker studies. *Diagnostic Pathology* **2**, doi:10.17629/www.diagnosticpathology.eu-2016-2:104 (2016).
69 Poduval, D., Sichmanova, Z., Straume, A. H., Lonning, P. E. & Knappskog, S. The novel microRNAs hsa-miR-nov7 and hsa-miR-nov3 are over-expressed in locally advanced breast cancer. *PLoS One* **15**, e0225357, doi:10.1371/journal.pone.0225357 (2020).
70 Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352, doi:10.1038/nature10983 (2012).
71 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
72 Barbosa-Morais, N. L. *et al.* A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* **38**, e17, doi:10.1093/nar/gkp942 (2010).
73 Johnson, S., Chen, H. & Lo, P. K. In vitro Tumorsphere Formation Assays. *Bio Protoc* **3**, doi:10.21769/bioprotoc.325 (2013).
74 Wensaas, A. J. *et al.* Cell-based multiwell assays for the detection of substrate accumulation and oxidation. *J Lipid Res* **48**, 961-967, doi:10.1194/jlr.D600047-JLR200 (2007).
75 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
76 Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst* **6**, 612-620 e615, doi:10.1016/j.cels.2018.02.010 (2018).
77 Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* **347**, 70-78, doi:10.1016/j.jim.2009.06.008 (2009).
78 Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).

79 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
80 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
81 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
82 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365 (2014).
83 Stark R, B. G. *DiffBind: differential binding analysis of ChIP-Seq peak data.*, <http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf> (2011).
84 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
85 Dravis, C. *et al.* Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell* **34**, 466-482 e466, doi:10.1016/j.ccell.2018.08.001 (2018).
86 Chung, C. Y. *et al.* Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships. *Cell Rep* **29**, 495-510 e496, doi:10.1016/j.celrep.2019.08.089 (2019).
87 Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42, doi:10.1186/s13072-019-0287-4 (2019).
88 Meers, M. P., Janssens, D. H. & Henikoff, S. Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. *Mol Cell* **75**, 562-575 e565, doi:10.1016/j.molcel.2019.05.025 (2019).

Figure 1

A

E0771

Group	Number of tumors / number of injections cell per injection 500 000	10 000	1 000	CSC frequency	p-value
Chow	5/5	9/13	13/24	1/3846	5.39e-11
HFD	5/5	12/12	22/23	1/319	

TeLi

Group	Number of tumors / number of injections cell per injection 500 000	40 000	100	CSC frequency	p-value
Chow	5/5	8/18	0/6	1/67925	2.01e-04
HFD	5/5	14/15	2/5	1/11443	

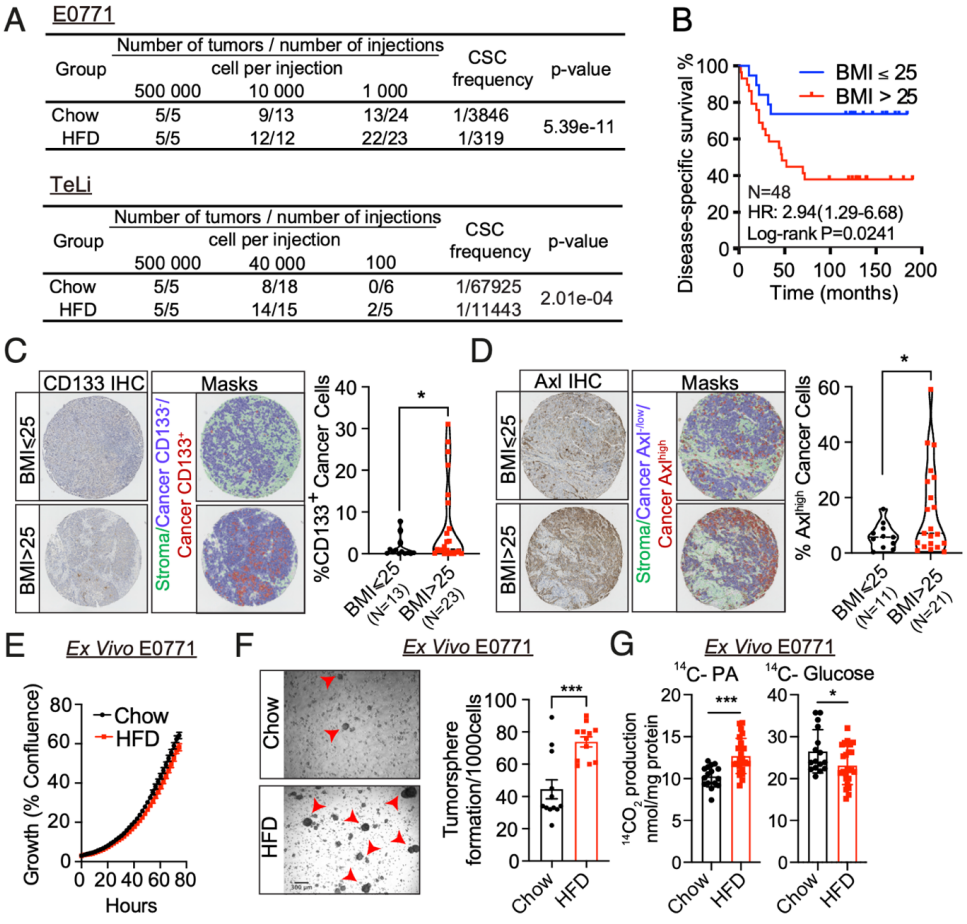


Figure 1. Obesity is associated with increased frequency of stem cell-like cancer cells in PM/ER$^-$/PR$^-$ breast cancer patients and mouse models of breast cancer

(A) Tumor incidence following orthotopic implantation of the indicated number of cells into the fourth mammary fat pads of chow diet and HFD-fed mice. The frequency of cancer stem-like cells was calculated by the extreme limiting dilution analysis[77].

(B) Kaplan-Meier curves show disease specific survival for postmenopausal and hormone receptor negative patients (N=48) with high (Red, BMI > 25) or low (Blue, BMI ≤ 25) BMI. Log-rank (Mantel-Cox) P value is denoted for difference in disease specific survival.

(C-D) Representative tissue microarray and QuPath analysis mask pictures of CD133 (C) and Axl (D) staining in high (BMI > 25, N = 23 for CD133 and N = 21 for Axl) or low (BMI ≤ 25, N = 13 for CD133 and N = 11 for Axl) BMI PM/ER^-/PR^- patients' tumor samples, stroma is marked in green, $CD133^-$ or $Axl^{-/low}$ cancer cells are marked in blue and positive staining cancer cells are marked in red (left panel). Quantification of $CD133^+$ or Axl^{high} staining cancer cells in the tumor samples (right panel).

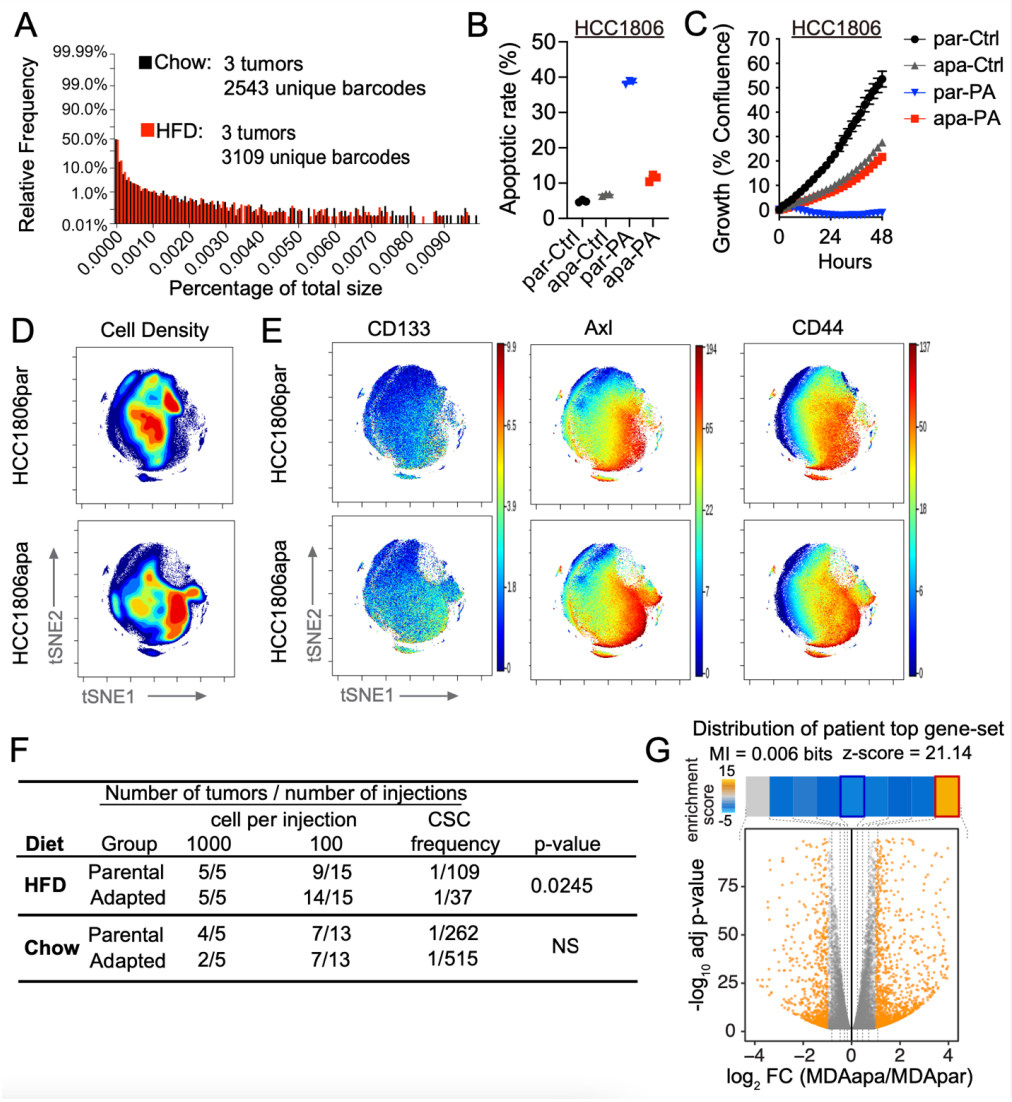
(E) Time-dependent proliferation assay of ex vivo E0771 cells isolated from chow diet or HFD-fed mice. For each time point, data are represented as mean ± SEM of four tumor samples (eight replicates per tumor sample were measured) from each group.

(F) Tumorsphere formation assay of E0771 *ex vivo* cells isolated from chow diet or HFD-fed mice. Representative images of day-5 tumorspheres formation of E0771 *ex vivo* cells and red arrowheads mark the identified tumorspheres (left panel). Quantification of day-5 tumorspheres are represented as mean ± SEM of four tumor samples from each group and three replicates were measured for each sample (right panel).

(G) Fatty acid oxidation on E0771 *ex vivo* cells isolated from two chow diet-fed mice and three HFD-fed mice was shown by cumulative $^{14}CO_2$-production during incubation with radio-labeled [1-^{14}C] palmitic acid (left panel). Glucose oxidation was shown by cumulative $^{14}CO_2$-production during incubation with radio-labeled D-[^{14}C(U)] glucose (right panel). The oxidation data are normalized to cell protein content. For each sample, 8 replicates are measured and data are represented as mean ± SEM of all replicates from each group.

For C-D, unpaired, two-tailed Welch's t-test was used for statistical testing. For F-G, statistical significance determined with unpaired, two-tailed Student's t-test. (*, P value < 0.05; ***, P value < 0.001)

Figure 2



Diet	Group	Number of tumors / number of injections: cell per injection 1000	100	CSC frequency	p-value
HFD	Parental	5/5	9/15	1/109	0.0245
	Adapted	5/5	14/15	1/37	
Chow	Parental	4/5	7/13	1/262	NS
	Adapted	2/5	7/13	1/515	

Figure 2. Long-term adaptation to PA phenocopy obesity-induced stem-cell features

(A) Distributions of relative barcode size for three tumors in each group.

(B) Apoptotic rate of parental (par) and adapted (apa) HCC1806 cells which were treated with 400 µM PA and vehicle (Ctrl) for 48hrs. Apoptosis analysis was performed using Alexa Fluor™ 488 conjugate Annexin V and propidium iodide double staining and the fluorescent signals were measured by flow cytometry. Both early (Annexin V^+/PI^-) and late (Annexin V^+/PI^+) apoptotic cells were included for the apoptotic rate % calculation. Data are represented as mean ± SEM of 3 replicates.

(C) Time-dependent proliferation assay of parental and adapted HCC1806 following 48hrs. Cells were exposed to 400 µM PA and vehicle (Ctrl). Cell growth was determined by high content imaging and represented as % confluence normalized to t=0. For each time point, data are represented as mean ± SEM of 6 replicates.

(D) Representative contour plots of mass cytometry data colored by density of cells showing the changes between parental and adapted HCC1806 cells. Total number of analyzed cells per cell line is equal to 85 000 cells. Color code represents the cell density from low (blue) to high (red).

(E) Representative tSNE plots of single parental and adapted HCC1806 cells colored by expression of CD133, Axl and CD44.

(F) Tumor incidence following orthotopic implantation of the indicated number of parental and adapted E0771 cells into the fourth mammary fat pads of chow diet and HFD-fed mice. The frequency of cancer stem-like cells was calculated by the extreme limiting dilution analysis. (NS, p value > 0.05)

(G) The distribution of genes induced by obesity (obese and overweight compared to non-obese patients) in PM hormone negative breast cancers patients among the gene expression changes observed in PA-adapted cell lines. The ~300 genes were used to perform gene-set enrichment analysis using iPAGE. Shown here is the volcano plot showing gene expression changes in PA-adapted cells relative to their parental line. iPAGE divided the spectrum of log-fold changes into equally populated bins (dotted line) and used mutual information to assess the non-random distribution of the query gene-set among these bins. We have included the mutual information value (MI) and its associated z-score reported by iPAGE. For visualization, the enrichment/depletion of the query gene-set was determined using the hyper-geometric test and the resulting p-value was used to define an enrichment score that is shown as a heatmap across

the expression bins. The obesity-induced genes were significantly enriched in the top-most bin. The red and blue borders in the heatmap denoted statistical significance for enrichment and depletion respectively. The gene expression of parental and adapted cell lines was measured by RNAseq. For mRNA expression of human breast cancer samples were measured by microarray analysis by using the same breast cancer patients cohort in Figure 1B.

Figure 3

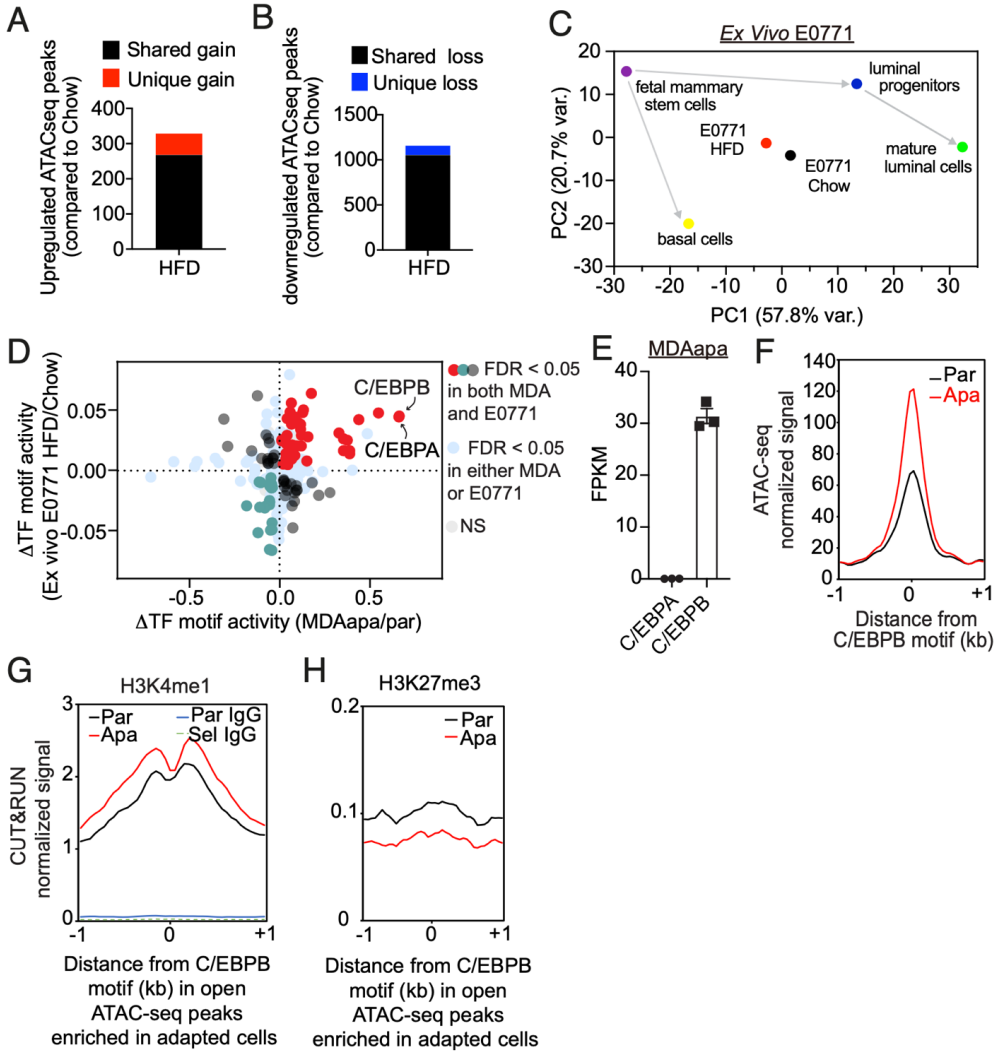


Figure 3. Adaptation to obese environments induces open chromatin linked with C/EBPB occupancy

(A) Total number of significantly upregulated ATACseq peaks in E0771 HFD relative to chow with a false discovery rate (FDR) < 0.05. Unique gain peaks refer to peaks identified only in the HFD condition, whereas shared peaks are peaks called in both conditions.

(B) Total number of significantly downregulated ATACseq peaks in E0771 HFD relative to Chow with a FDR < 0.05. Unique loss peaks refer to peaks identified only in the Chow condition, whereas shared peaks are peaks called in both conditions.

(C) Principal component analysis of the E0771 *ex vivo* cells and different cell lineages along the mammary gland developmental trajectory (GEO: GSE116386) using the average transcription factor motif activity estimated by chromVar.

(D) Overlap of differential transcription factor binding motif activity between MDA-MD-231 (apa/par) and E0771 (HFD/Chow) as determined by diffTF.

(E) Within-sample normalized gene expression of transcription factor homologs C/EBPA and C/EBPB in PA-adapted MDA-MB-231 cells using RNA-seq. FPKM = fragments per kilobase of transcript per million mapped reads.

(F) Metagene representation of the mean ATACseq signal across more accessible C/EBPB motif regions in parental or adapted MDA-MB-231 cells. The mean signal of three adapted or parental MDA-MB-231 biological replicates was determined by averaging signals of 1 kb around the center of C/EBPB DNA-binding motifs.

(G) Metagene representation of the mean H3K4me1 and IgG signals across more accessible C/EBPB motif regions as in (F) in parental or adapted MDA-MB-231 cells. The mean signal of three adapted or parental MDA-MB-231 biological replicates was determined by averaging signals of 1 kb around the center of C/EBPB DNA-binding motifs.

(H) Metagene representation of the mean H3K27me3 signals across more accessible C/EBPB motif regions as in (F) in parental or adapted MDA-MB-231 cells. The mean signal of three adapted or parental MDA-MB-231 biological replicates was determined by averaging signals of 1 kb around the center of C/EBPB DNA-binding motifs.

Figure 4

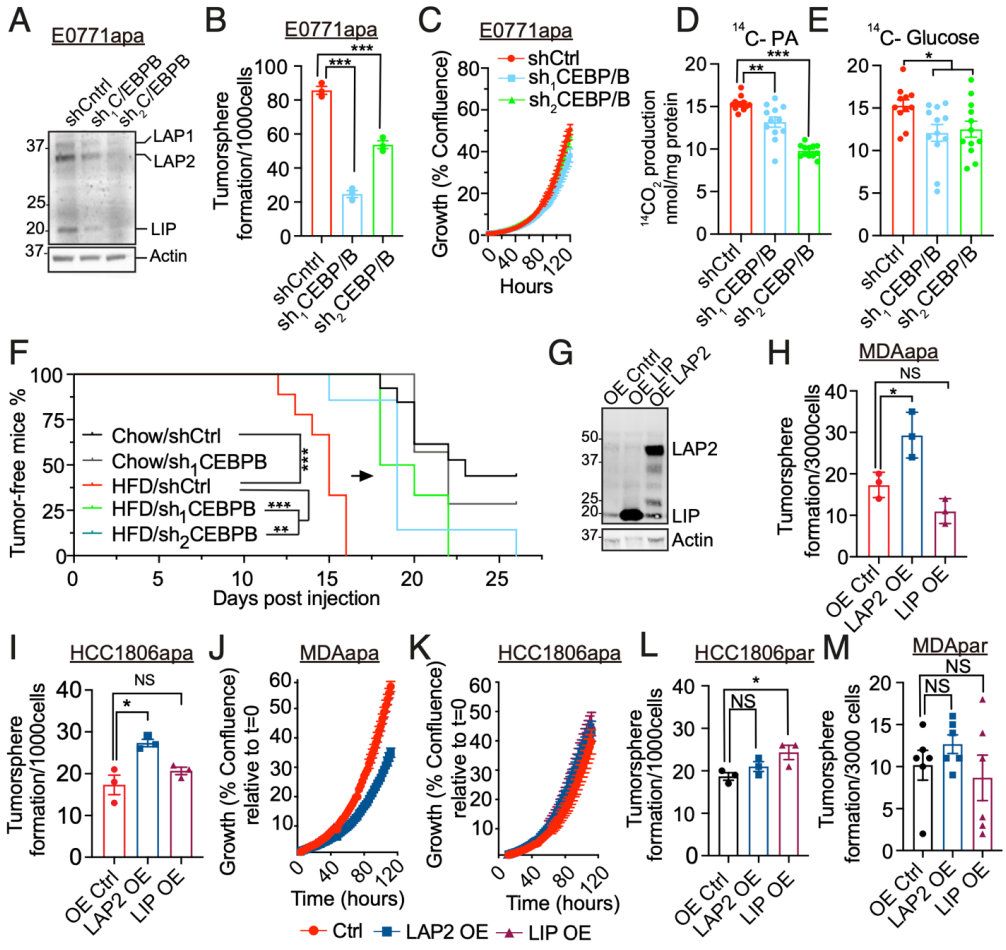


Figure 4. C/EBPB promotes tumor stemness specifically in obese environments

(A) Western blots of C/EBPB and Actin in whole cell lysates extracted from knockdown control (shCtrl) and two independent *C/ebpb* knockdown (sh_1C/EBPB and sh_2C/EBPB) adapted E0771 cells. Actin was used for the normalization. Three C/EBPB isoforms, LAP1, LAP2 and LIP are marked in the blots.

(B) The changes of tumorsphere formation upon stable knockdown of C/ebpb on adapted E0771 cells. The knockdowns were performed by using two independent shRNAs.

(C) Time-dependent proliferation assay of control and C/ebpb knockdown adapted E0771 cells. Cell growth was determined by high content imaging and represented as % confluence normalized to t=0. For each time point, data are represented as mean ± SEM of 8 replicates.

(D) The changes of fatty acid oxidation upon C/ebpb knockdown in adapted E0771 cells. Fatty acid oxidation was shown by cumulative $^{14}CO_2$-production during incubation with radio-labeled [1-^{14}C] palmitic acid. The oxidation data are normalized to cell protein content.

(E) The changes of glucose oxidation upon C/ebpb knockdown in adapted E0771 cells. Glucose oxidation was shown by cumulative $^{14}CO_2$-production during incubation with radio-labeled D-[^{14}C(U)] glucose. The oxidation data are normalized to cell protein content.

(F) Tumor-free survival curves of chow diet and HFD-fed mice orthotopically implanted with E0771 knockdown control and C/ebpb knockdown cells. Tumor volume was measured every 2-3days and tumor formation were recorded when reached a volume of 50 mm^3. The analysis was performed by using the pooled data from two independent experiments (Chow/shCtrl N=13, Chow/sh_1CEBPB N=7, HFD/shCtrl N=9, HFD/sh_1CEBPB N=6, HFD/sh_2CEBPB N=7).

(G) Western blots against C/EBPB in whole cell lysates extracted from control, LAP2 and LIP overexpressed MDA-MB-231 PA-adapted cell line. Actin was used for the normalization.

(H-I) The changes of tumorsphere formation upon the overexpression C/EBPB LAP2 and LIP isoforms on adapted MDA-MB-231 (H) and HCC1806 (I) cells.

(J-K) Time-dependent proliferation assay of control, LAP2 and LIP overexpressed adapted MDA-MB-231 (J) and HCC1806 (K) cells. Cell growth was determined by high content imaging and represented as % confluence normalized to t=0. For each time point, data are represented as mean ± SEM of 8 replicates.

(L-M) The changes of tumorsphere formation upon the overexpression C/EBPB LAP2 and LIP isoforms on parental HCC1806 (L) and MDA-MB-231 (M) cells.

For B, D-E, H-I and L-M, statistical significance determined with unpaired, two-tailed Student's t-test. For F, Log-rank (Mantel-Cox) test was used for statistical testing. (NS, P value > 0.05; *, P value < 0.05; **, P value < 0.01; ***, P value < 0.001)

Figure 5

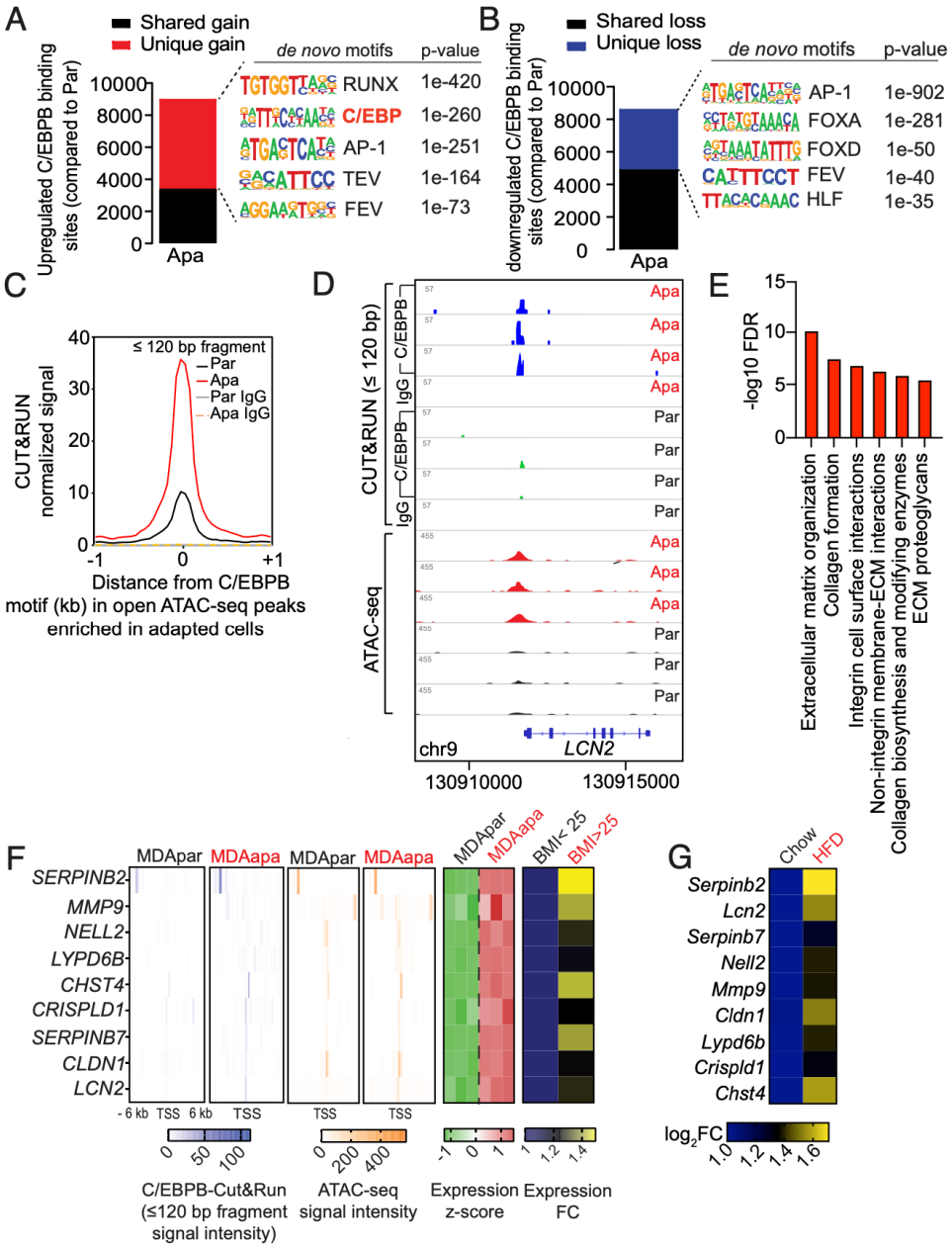

A
Shared gain
Unique gain
de novo motifs
p-value
Upregulated C/EBPB binding sites (compared to Par)
RUNX 1e-420
C/EBP 1e-260
AP-1 1e-251
TEV 1e-164
FEV 1e-73
Apa
B
Shared loss
Unique loss
de novo motifs
p-value
downregulated C/EBPB binding sites (compared to Par)
AP-1 1e-902
FOXA 1e-281
FOXD 1e-50
FEV 1e-40
HLF 1e-35
Apa
C
CUT&RUN normalized signal
≤ 120 bp fragment
Par
Apa
Par IgG
Apa IgG
Distance from C/EBPB motif (kb) in open ATAC-seq peaks enriched in adapted cells
D
CUT&RUN (≤ 120 bp)
C/EBPB
IgG
ATAC-seq
chr9
LCN2
130910000
130915000
E
-log10 FDR
Extracellular matrix organization
Collagen formation
Integrin cell surface interactions
Non-integrin membrane-ECM interactions
Collagen biosynthesis and modifying enzymes
ECM proteoglycans
F
MDApar
MDAapa
BMI< 25
BMI>25
SERPINB2
MMP9
NELL2
LYPD6B
CHST4
CRISPLD1
SERPINB7
CLDN1
LCN2
TSS
C/EBPB-Cut&Run (≤120 bp fragment signal intensity)
ATAC-seq signal intensity
Expression z-score
Expression FC
G
Chow
HFD
Serpinb2
Lcn2
Serpinb7
Nell2
Mmp9
Cldn1
Lypd6b
Crispld1
Chst4
log2FC

Figure 5. Differential C/EBPB occupancy regulates extracellular matrix organization

(A-B) Total number of upregulated (A) and downregulated (B) C/EBPB binding sites in adapted MDA-MB-231 cells relative to the parental with an FDR < 0.05. Unique gain or loss sites refer to binding sites identified only in the adapted or parental condition, whereas shared peaks are peaks called in both conditions. Top 5 significant *de novo* motifs enriched in the unique gain or loss sites were called by HOMER.

(C) Metagene representation of the mean C/EBPB Cut&Run signal (fragment length ≤ 120 bp) across the same chromatin regions as in open ATACseq peak enriched in adapted cells from three biological replicates of adapted or parental MDA-MB-231 cells. Control IgG Cut&Run experiment in adapted and parental cells was included for comparison.

(D) Representative genome browser tracks of normalized C/EBPB and IgG Cut&Run and ATACseq profiles around the LCN2 locus in biological replicates of parental and adapted MDA-MB-231 cells.

(E) Reactome pathway analysis of genes containing gained chromatin accessibility to C/EBPB.

(F) Heatmaps showing average Cut&Run and ATACseq signal intensity centered around the transcription start site (TSS) of the nine putative C/EBPB target genes, and the corresponding mRNA expression of the same genes in three biological replicates of MDApar and MDAapa cells (panels 1-5). Heatmap of expression fold change of the same genes in obese and overweight compared to lean patients was also shown (panel 6).

(G) Heat map showing mRNA expression of potential C/EBPB targets in E0771 cells isolated from chow diet and HFD-fed mice. mRNA expression was measured by RT-qPCR with cells isolated from N=2 chow tumors and N=3 HFD tumors.

Figure 6

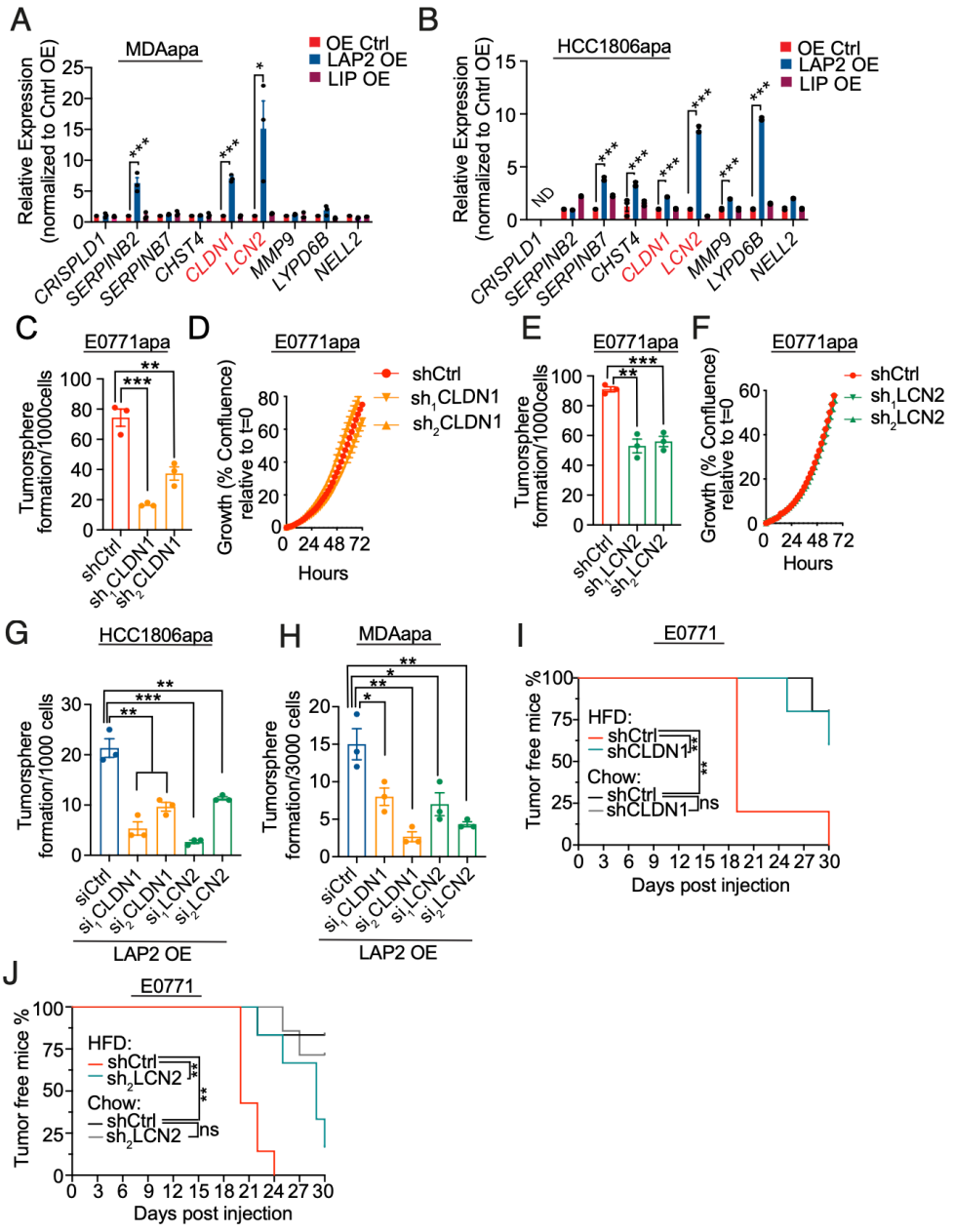

A
MDAapa
OE Ctrl
LAP2 OE
LIP OE
Relative Expression (normalized to Cntrl OE)
CRISPLD1
SERPINB2
SERPINB7
CHST4
CLDN1
LCN2
MMP9
LYPD6B
NELL2
B
HCC1806apa
ND
C
E0771apa
Tumorsphere formation/1000cells
shCtrl
sh1CLDN1
sh2CLDN1
D
Growth (% Confluence relative to t=0
Hours
shCtrl
sh1CLDN1
sh2CLDN1
E
sh1LCN2
sh2LCN2
F
G
HCC1806apa
Tumorsphere formation/1000 cells
siCtrl
si1CLDN1
si2CLDN1
si1LCN2
si2LCN2
LAP2 OE
H
MDAapa
Tumorsphere formation/3000 cells
I
E0771
Tumor free mice %
HFD:
shCtrl
shCLDN1
Chow:
ns
Days post injection
J
shLCN2

Figure 6. *CLDN1 and LCN2 are* required for C/EBPB dependent stem cell-like capabilities

(A-B) RT-qPCR was used to measure changes in the expression of C/EBPB potential target genes upon the overexpression of C/EBPB LIP and LAP2 isoforms on adapted MDA-MB-231 (A) and HCC1806 (B) cells. The expression of target genes is shown as relative fold change over Control OE. Data shown as mean ± SEM of 3 independently repeated experiments.

(C) The changes of tumorsphere formation upon stable knockdown of *Cldn1* on adapted E0771 cells. The knockdowns were performed by using two independent shRNAs.

(D) Time-dependent proliferation assay of control and *Cldn1* knockdown adapted E0771 cells. Cell growth was determined by high content imaging and represented as % confluence normalized to t=0. For each time point, data are represented as mean ± SEM of 8 replicates.

(E) The changes of tumorsphere formation upon stable knockdown of *Lcn2* on adapted E0771 cells. The knockdowns were performed by using two independent shRNAs.

(F) Time-dependent proliferation assay of control and *Lcn2* knockdown adapted E0771 cells. Cell growth was determined by high content imaging and represented as % confluence normalized to t=0. For each time point, data are represented as mean ± SEM of 8 replicates.

(G-H) The changes of tumorsphere formation upon knockdown of *CLDN1* and *LCN2* with siRNAs on the LAP2 overexpressed adapted HCC1806 (G) and MDA-MB-231 (H) cells. The knockdown was performed by using two independent siRNAs for each gene.

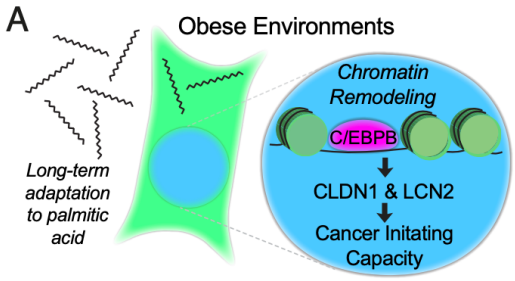
(I) Tumor-free survival curves of chow diet and HFD-fed mice orthotopically implanted with 100 E0771 knockdown control and *Cldn1* knockdown cells (HFD/shCtrl N=5, HFD/shCLDN1 N=5; Chow/shCtrl N=5, Chow/shCLDN1 N=5). Tumor volume was measured every 2-3 days and tumor formation were recorded when reached a volume 50 mm^3.

(J) Tumor-free survival curves of chow diet and HFD-fed mice orthotopically implanted with 100 E0771 knockdown control and *Lcn2* knockdown cells. Tumor volume was

measured every 2-3 days and tumor formation were recorded when reached a volume of 50 mm^3. The analysis was performed by using the pooled data from two independent experiments (HFD/shCtrl N=7, HFD/sh_2LCN2 N=6; Chow/shCtrl N=6, Chow/sh_2LCN2 N=7).

For A-B, C-E and G-H, multiple t tests were performed to assess statistical significance. For I and J, P values were determined with Log-rank (Mantel-Cox) test. (NS, P value > 0.05; *, P value < 0.05; **, P value < 0.01; ***, P value < 0.001)

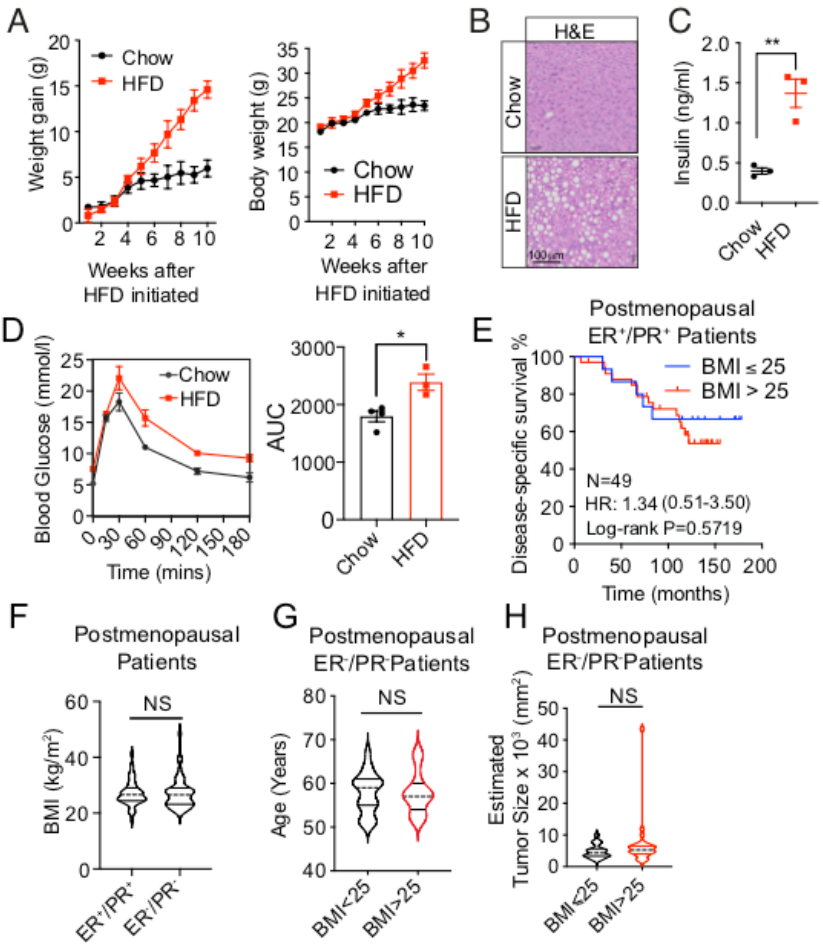
Figure 7



(A) A schematic model of obese environment on breast cancer initiating capacity. Long-term adaptation of breast cancer cells to palmitic acid promotes initiating capacity through increased accessibility of C/EBPB binding motifs, which induces the expression of C/EBPB targets CLDN1 and LCN2

SUPPLEMENTAL INFORMATION

Supplementary Figure 1.



(A) Body weight gain (left panel) and absolute body weight (right panel) of HFD and chow-fed mice before implantation of tumors. Six-weeks old female C57BL/6J mice were started on HFD or standard chow diet (n=4 per group) for ten weeks prior to the tumor implantation. The measurement of animal body weight was started at six weeks of age and recorded weekly. For each time point, data is represented as mean ± SEM of four mice per group (n=4 per group).

(B) Hematoxillin and Eosin (H&E) stained tissue sections of livers from HFD and chow-fed mice. After ten weeks of HFD or chow diet feeding, female C57BL/6J mouse were sacrificed and and livers were harvested. Liver sections were stained using H&E. Histological analysis showed increased liver steatosis in mice from the HFD group compared to mice from the chow group.

(C) Concentration of fasting plasma insulin in HFD and chow-fed mice. Concentrations were determined by ELISA using overnight fasted blood samples collected from female C57BL/6J mice fed an HFD or chow diet for ten weeks (n=3 per group).

(D) Oral glucose tolerance test performed on HFD and chow-fed mice. Blood glucose clearance was determined in mice fed an HFD (n=3) or chow (n=4) diet for ten weeks. Blood glucose concentrations were measured at 0min, 15mins, 30mins, 60mins, 120mins and 180mins following glucose administration by oral gavage. For each time point, data is represented as mean ± SEM. AUC = area under the curve.

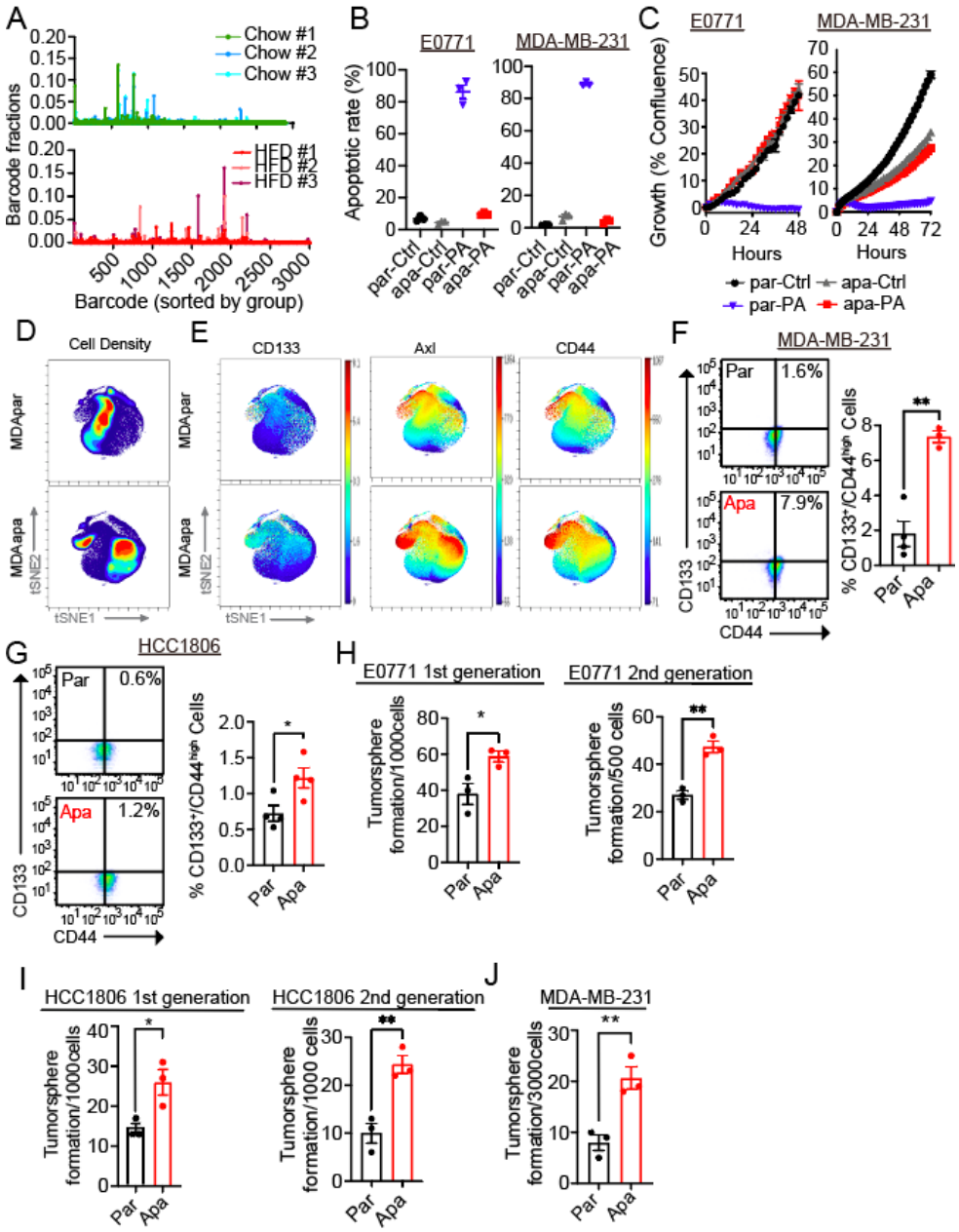
(E) Kaplan-Meier curves display disease specific survival for postmenopausal and ER^+/PR^+ patients (N=49) with high (red, BMI > 25) or low (blue, BMI ≤ 25) BMI. Log-rank (Mantel-Cox) P value is denoted for difference in disease specific survival. The analysis showed no significant difference between the groups.

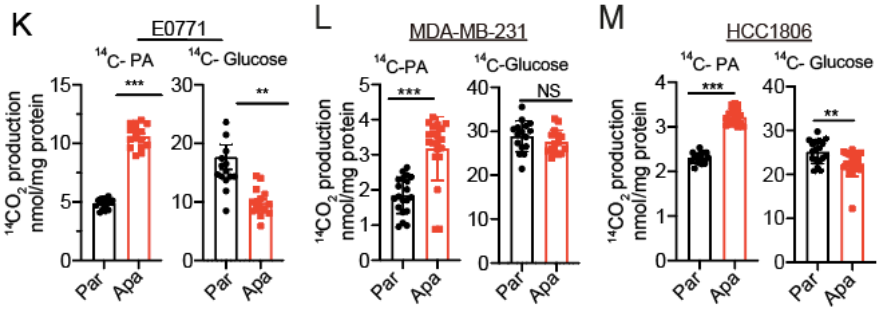
(F) Distribution of BMI in postmenopausal ER+/PR+ and ER-/PR- patients. BMI distribution was similar between the groups.

(G-H) Distribution of postmenopausal ER^-/PR^- patients' age (G) and estimated tumor size (H) in high (BMI > 25) and low (BMI ≤ 25) BMI groups. The estimated tumor size was calculated by multiplying the largest diameter by its perpendicular. The analysis showed no significant difference between the groups.

For C-D, statistical significance determined with unpaired, two-tailed Student's t-test. For F-H Kolmogorov-Smirnov test was used for statistical testing. (NS, P value > 0.05; *, P value < 0.05; **, P value < 0.01).

Supplementary Figure 2.





(A) Barcode distribution of all replicates of tumors derived from chow (upper panel) and HFD (lower panel) mice. The x axis of the histograms is barcode ID which were sorted by group and each bar represents one unique barcode.

(B) Apoptotic rate of parental and adapted E0771 (left panel) and MDA-MB-231 (right panel) cells that were treated with PA (500 µM for E0771 and 400 µM for MDA-MB-231) and vehicle (Ctrl) for 48hrs. Data are represented as mean ± SEM of three replicates.

(C) Time-dependent proliferation assay of parental and adapted E0771 (left panel) and MDA-MB-231 (right panel) cells following 48-72hrs. Cells were exposed to 400 µM (for MDA-MB-231 cells) or 500 µM (for E0771 cells) PA and vehicle (Ctrl). Cell growth was determined by high content imaging and represented as % confluence normalized to t=0. For each time point, data are represented as mean ± SEM of four to eight replicates.

(D) Representative contour plots of mass cytometry data colored by density of cells showing the changes between parental and adapted MDA-MB-231 cells. Total number of analyzed cells per cell line is equal to 100 000 cells. Color code represents the cell density from low (blue) to high (red).

(E) Representative tSNE plots of single parental and adapted MDA-MB-231 cells colored by expression of CD133, Axl and CD44.

(F-G) $CD133^{+}/CD44^{high}$ cells population in parental and adapted MDA-MB-231 (F) and HCC1806 (G) cells. Cells were stained by CD133-APC and CD44-FITC antibodies and measured by flow cytometry. Quantification data is shown as mean ± SEM of four

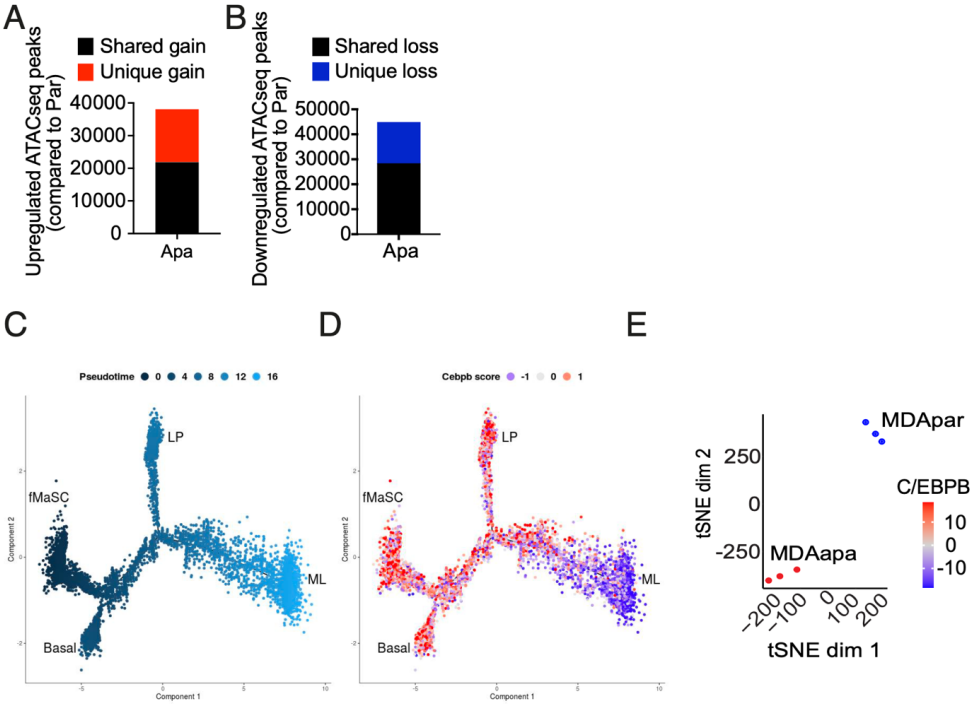
replicates (one outlier in MDA-MB-231apa group was excluded from the quantification by Grubbs' outlier test).

(H-J) Tumorsphere formation assay and serial tumorsphere propagation assay of parental and adapted E0771 (H), HCC1806 (I) and MDA-MB-231 (J) cells. Cells (1000 cells/well for E0771 and HCC1806 cell lines, 3000 cells/well for MDA-MB-231 cell lines) were seeded into ultra-low attachment 6-well plates in the stem cell media and following 5-10days of growth, tumorspheres were imaged and quantified. The serial tumorsphere propagations were performed by dissociating the primary tumrospheres, and following the same method to reseed 500 cells/well for E0771 cells (H right panel) and 1000 cells/well for HCC1806 cells (I right panel). Quantifications of tumorspheres are represented as mean ± SEM of three replicates for each condition (n=3 / condition).

(K-M) Comparison of fatty acid and glucose oxidation assays between parental and adapted cells. Fatty acid oxidation on parental and adapted E0771 (K), MDA-MB-231 (L) and HCC1806 (M) cells was measured by cumulative $^{14}CO_2$-production during incubation with radio-labeled [1-^{14}C] palmitic acid (left panel). Glucose oxidation was shown by cumulative $^{14}CO_2$-production during incubation with radio-labeled D-[^{14}C(U)] glucose (right panel).

For F-M, statistical significance determined with unpaired, two-tailed Student's t-test. (NS, P value > 0.05; *, P value < 0.05; **, P value < 0.01; ***, P value < 0.001).

Supplementary Figure 3.



(A) Total number of significantly upregulated ATACseq peaks in MDAapa relative to MDApar with a FDR < 0.05. Unique gain peaks refer to peaks identified only in the adapted condition, whereas shared peaks are peaks called in both conditions.

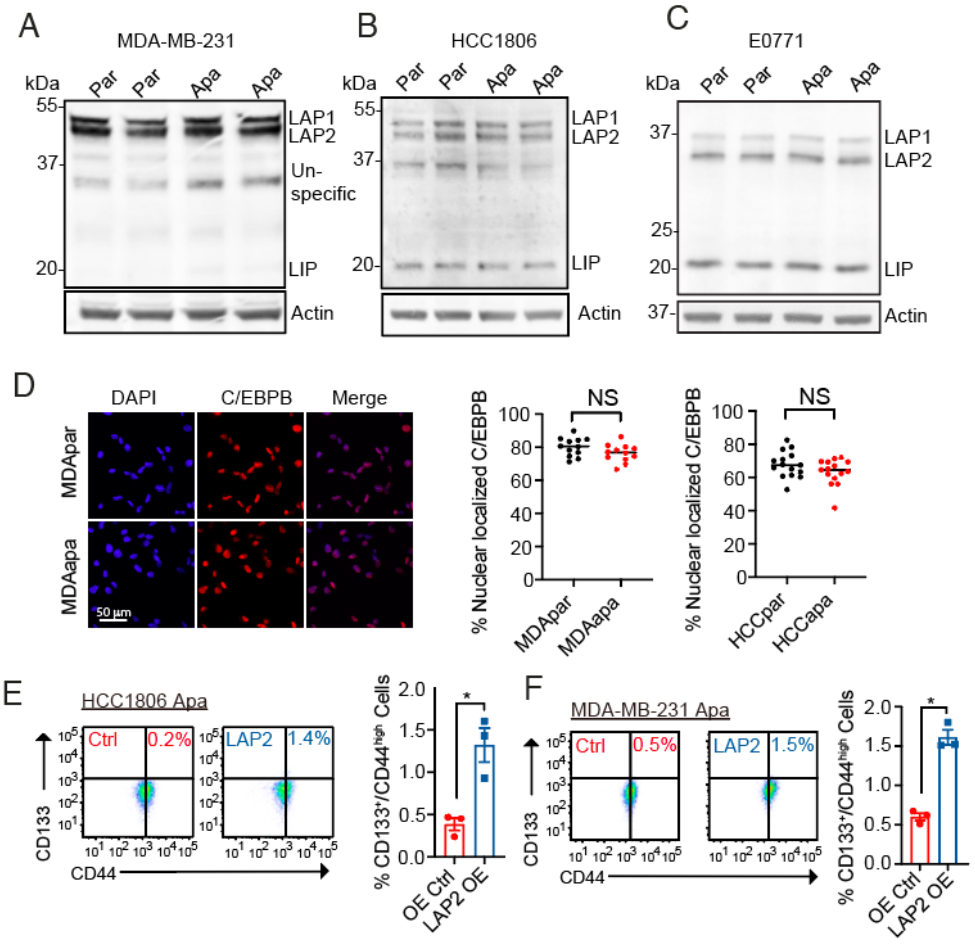
(B) Total number of significantly downregulated ATACseq peaks in MDAapa relative to MDApar with a FDR < 0.05. Unique loss peaks refer to peaks identified only in the parental condition, whereas shared peaks are peaks called in both conditions.

(C) Pseudotime analysis of single-nuclei ATACseq of murine mammary cells at different developmental stages (GSE125523).

(D) Motif enrichment of transcription factors C/ebpb in the open chromatin regions at each individual cell along the mammary gland developmental trajectory is shown. fMaSC, fetal mammary stem cells; basal, adult basal cells; LP, luminal progenitors; and ML, mature luminal cells.

(E) t-SNE clustering of individual MDApar and MDAapa replicates showing differential motif enrichment in transcription factors C/EBPB.

Supplementary Figure 4.



(A-C) Immunoblots of C/EBPB and Actin in parental and adapted MDA-MB-231 (A), HCC1806 (B) and E0771 (C) cell lines. Actin was used for the normalization.

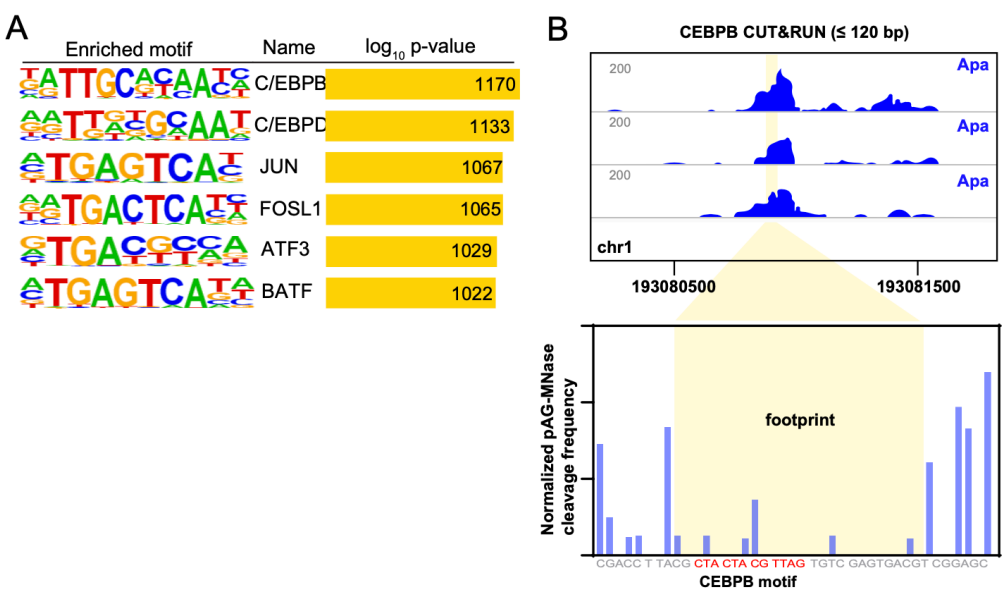
(D) Representative images of C/EBPB-immunofluorescent staining on MDApar and MDAapa cells (left panel). Quantification (right panel) was calculated by the percentage of C/EBPB localized in the nucleus compared to the cytoplasm for MDA-MB-231 and HCC1806 parental and PA-adapted cell lines.

(E-F) CD133+/CD44high cells population in adapted HCC1806 (E) and MDA-MB-231 (F) cells overexpressing LAP2. Cells were stained by CD133-APC and CD44-FITC

antibodies and measured by flow cytometry. Quantification data is shown as mean ± SEM of three replicates.

For E-F, statistical significance determined with unpaired, two-tailed Student's t-test. (NS, P value > 0.05; *, P value < 0.05).

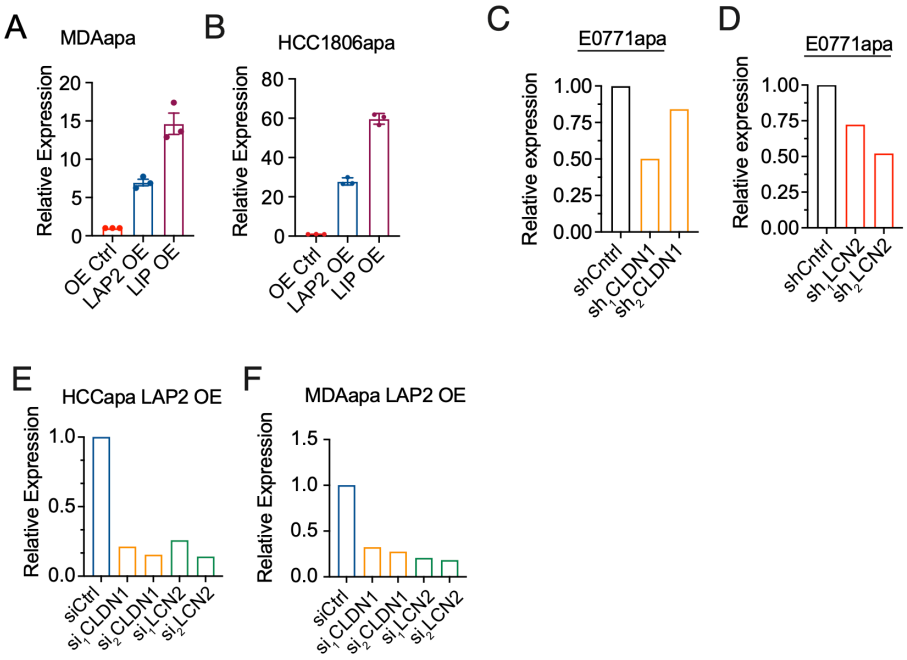
Supplementary Figure 5.



(A) Motifs enriched in C/EBPB Cut&Run footprints in MDAapa cells. The p-values shown in the figure were reported by HOMER using HOCOMOCO motifs.

(B) Single locus footprint analysis of C/EBPB Cut&Run experiments in adapted MDA-MB-231 cells. Upper panel shows representative genome browser tracks of C/EBPB Cut&Run signal in the specified region in chromosome 1 (chr1). Lower panel shows the total normalized pA/G-MNase cut frequency of the three biological replicates at each nucleotide around the *C/EBPB* motif within the identified footprint in the specified region.

Supplementary Figure 6.



(A-B) RT-qPCR was used to measure changes in the expression of *C/EBPB* upon the overexpression of *C/EBPB LAP2* and *LIP* isoforms on adapted MDA-MB-231 (A) and HCC1806 (B) cells. The relative expression is shown as relative fold change over control cells. Data shown as mean ± SEM of three independently repeated experiments.

(C-D) RT-qPCR was used to measure efficiency of *Cldn1* (C) and *Lcn2* (D) knockdown in adapted E0771 cells. Knockdown was performed by using two independent shRNAs for each gene.

(E-F) RT-qPCR was used to measure efficiency of *CLDN1* (yellow) *and LCN2* (green) knockdown relative to knockdown control (siCtrl, blue) in adapted HCC1806 (E) and MDA-MB-231 (F) cells. Knockdown was performed by using two independent siRNAs for each gene.

Supplementary Table S1. List of genes included in the targeted sequencing of PM/ER$^-$/PR$^-$

ABL1	ERCC2	MAPK10	ROS1
ABL2	ERCC3	MAPK7	RPS6KB1
ACVR2A	ERCC4	MAPK8	RPTOR
AKT1	ERCC5	MAPK9	RRM2B
AKT2	ESR1	MCL1	RSPO2
AKT3	ETV1	MDM2	RSPO3
ALK	EZH2	MDM4	RUNX1
APC	FADD	MED12	SETD2
AR	FAM123B	MED12L	SF3B1
ARAF	FANCA	MED13	SFTPA1
ARFRP1	FANCC	MED29	SHC1
ARID1A	FANCD2	MEN1	SKP2
ARID1B	FANCE	MET	SLIT2
ARID2	FANCF	MITF	SMAD2
ASXL1	FANCG	MLH1	SMAD3
ATM	FAS	MLL	SMAD4
ATR	FBXO11	MLL2	SMARCA4
ATRX	FBXW7	MLL3	SMARCB1
AURKA	FGFR1	MPL	SMO
AURKB	FGFR2	MRAS	SMURF1
AXIN1	FGFR3	MRE11A	SOCS1
BAG4	FGFR4	MSH2	SOX10
BAP1	FH	MSH6	SOX2
BCL11A	FLT1	MST1	SOX9
BCL2	FLT3	MTDH	SPOP
BCL2A1	FLT4	MTOR	SRC
BCL2L1	FOXA1	MUTYH	SRSF2
BCL2L2	FOXL2	MYB	STAT3
BCL6	FOXO1	MYC	STK11
BCOR	FOXP4	MYCL1	SUFU
BIRC2	GAB2	MYCN	TBX22
BIRC7	GABRG1	MYD88	TBX3
BLM	GATA1	MYO3A	TERT
BPTF	GATA2	MYO5B	TET2
BRAF	GATA3	MYOC	TGFBR2
BRCA1	GATA6	NBN	TNFAIP3
BRCA2	GNA11	NCOA2	TOP1
BRIP1	GNAQ	NCOA3	TP53
BUB1B	GNAS	NF1	TP63
C11orf30	GPC5	NF2	TP73
CARD11	GPR124	NFE2L2	TRAF2
CASP8	GRB2	NGFR	TSC1
CBL	GRB7	NKX2-1	TSC2
CCND1	GRID1	NOTCH1	TSHR
CCND2	GUCY1A2	NOTCH2	U2AF1
CCND3	H3F3A	NOTCH3	USP9X
CCNE1	HIST1H3B	NOTCH4	VEGFA

Supplementary Table S2. Antibody panel used for mass cytometry analysis

Isotope	Antigen	Cell Location	Epitope	Phenotype	Clone
168Er	Axl	Extracellular	Total	Stemness/EMT	1H12
160Gd	CD133	Extracellular	Total, Epitope 1	Stemness	AC133
173Yb	CD44	Extracellular	Total, Surface	Stemness	IM7
158Gd	E-cadherin	Extracellular	CD324/E-Cadherin	Epithelial	"24E10"
170Er	EGFR	Extracellular	Total EGFR	Epithelial organs/ initiates MAPk, Akt and JNk signalling	AY13
143Nd	N-cadherin	Extracellular	CD325/N-Cadherin	Stemness/mesenchymal	8C11
156Gd	p38	Intracellular	p38 [T180/Y182]	MAPK for stress response	D3F9
152Sm	pAkt	Intracellular	pAkt [S473]	PI3K pathway	D9E
176Yb	pCreb	Intracellular	pCREB [S133]	Transcription factor- stress and growth	87G3
151Eu	pEGFR	Intracellular	pEGFR [Y1068]	Activated EGFR	Y38
154Sm	pErk1/2	Intracellular	pT202/pY204	Branch of MAPK-Mek pathway	20A
175Lu	pHistone H3	Intracellular	pHistone H2A.X [Ser139]	Metaphase. Activated dwnstream of p38 or Erk1/2	HTA28
159Tb	pMAPKAPK2	Intracellular	pMAPKAPK2 [T334]	ERk1/2 activated protein downstream of p38. Response to stress	27B7
166Er	pNFKB	Intracellular	pNF-kB p65 [S529]	Transcription Factor mediator of inflammatory and immune responses	K10-895.12.50
162Dy	pPLCgamma 2	Intracellular	pPLCg2 [pY759]	Mediator of inflammatory and immune responses	K86-689.37
150Nd	pRb	Intracellular	pRb [S807/811]	G1 to S cell cycle phase	J112-906
172Yb	pS6	Intracellular	pS6 [S235/S236]	Protein translation	N7-548
141Pr	pSHP2	Intracellular	Y580	RTK phosphatase promotes signaling of JAK/STAT, PI3K/Akt Ras/MAPK pathway	D66F10
153Eu	pStat1	Intracellular	Y704		4a
145Nd	pStat3	Intracellular	pY705		4/p
146Nd	pStat5	Intracellular	pY694		00047
149Sm	pStat6	Intracellular	Y641		18/P-stat6
163Dy	TGFbeta	Intracellular	Total		TW4-6H10
154Sm	Vimentin	Intracellular	Total	Mesenchymal	D21H3
167Er	YAP	Intracellular	CTD 379-407	Stemness Hippo	H9
172YT	CC3	Intracellular	Cleavage at D175	Apoptosis	5A1E
164DY	CK7	Intracellular	Total	Luminal marker	RCK105

Supplementary Table S3. PCR Primer sequences used for barcode amplification

	Sequence	Length
WS PCR Forward Primer	AATGATACGGCGACCACCGAGATCTACACACTGACTGCAGTCTGAGTCTGACAG	54
WS_Rev_Index_011	CAAGCAGAAGACGGCATACGAGATGTATCACGACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94
WS_Rev_Index_013	CAAGCAGAAGACGGCATACGAGATAGCGTCTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94
WS_Rev_Index_014	CAAGCAGAAGACGGCATACGAGATCAGCATGTCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94
WS_Rev_Index_015	CAAGCAGAAGACGGCATACGAGATGTACTCATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94
WS_Rev_Index_016	CAAGCAGAAGACGGCATACGAGATTCTGCAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94
WS_Rev_Index_017	CAAGCAGAAGACGGCATACGAGATACTGTACTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94
WS_Rev_Index_018	CAAGCAGAAGACGGCATACGAGATCGACAGCTATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTAGCACTAGCATAGAGTGCGTAGCT	94

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230854686 (print)
9788230857724 (PDF)