TFAP2B as a Regulator of Growth and Survival in Invasive Lobular Breast Cancer

by

Dorothy Carter

BS, Pennsylvania State University, 2017

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

Department of Integrated Systems Biology

This thesis was presented

by

Dorothy Carter

It was defended on

May 31, 2021

and approved by

Adrian Lee, Professor, PhD, Department of Pharmacology and Chemical Biology

Lin Zhang, Professor, PhD, Department of Pharmacology and Chemical Biology

Xiaosong Wang, Professor, MD, PhD, Department of Pathology

Thesis Advisor, Steffi Oesterreich, Professor, PhD, Department of Pharmacology and Chemical Biology

TFAP2B as a Regulator of Growth and Survival in Invasive Lobular Breast Cancer

Dorothy Carter, MS

University of Pittsburgh, 2021

Breast cancer impacts 1 in 8 women over their lifetime in the United States, and the death toll from breast cancer is expected to reach over 42,000 this year alone. Approximately 10-15% of invasive breast cancer are invasive lobular carcinoma (ILC) which are characterized by loss of E-Cadherin (CDH1), and expression of estrogen receptor (ER). ILC differs from invasive ductal carcinoma (IDC, ~80% of invasive breast cancer) with distinct clinical, pathological and molecular characteristics. Both ILC and IDC are treated the same way clinically, although ILC tumors have been shown to be less responsive to endocrine therapy with increased long-term recurrences 5-10 years after the original diagnosis. There is need to understand the distinct molecular, clinical and pathological differences of ILC compared to IDC and how these may be exploited for therapeutic benefit. An area that is specifically understudied are potential epigenetic differences between ER+ ILC and IDC, and how this affects the biology of disease.

Advances in sequencing technologies and deeper exploration in underrepresented invasive lobular carcinoma samples has given researchers an opportunity to understand mechanism of genes and transcription factors in ILC. Publicly available sequencing data such as TCGA, METABRIC and internal cohorts allows researchers to explore queries and strengthen rationale through individual sequencing studies or pooled metadata analyses. I have investigated both public and house generated RNA sequencing data and methylation data, two in cell lines with alterations, either overexpressing CDH1 in ILC cell lines, or knocking out CDH1 expression in IDC cell lines, in-house data through sequencing a large panel of cell lines

consisting of ILC, IDC, and ILC-like cell lines [4], as well as publicly available methylation data from TCGA in tumors and CCLE in cell lines. Exploring gene expression and pathway analysis suggests possibly therapeutic or prognostic pathways or genes effected uniquely in ILC.

We found a transcription factor, part of the AP-2 family, TFAP2B that is hypomethylated in ILC, and is important for regulation of many genes. Members of this transcription factor family have shown to control expression of cancer related genes, such as ERα and CDH1 [1].

Table of Contents

List of Tables

List of Figures

1.0 Background and Significance

1.1 Invasive breast cancer - Overview

Breast cancer is the most frequently diagnosed women's cancer, with an estimated 284,200 new cases of invasive breast cancer and 44,130 breast cancer deaths in 2021; it remains to be the second leading cause of cancer death among women [1]. Outcome of diagnosis has improved over time, where the current 5-year survival is 90%, and 10-year survival is 84% [2]. Improvement in outcome is largely due to enhanced screening methods for early detection and to molecularly targeted therapies.

1.2 Molecular Subtypes of Breast Cancer

Breast cancer is a heterogeneous disease, categorized through cellular composition, molecular alterations and how they behave clinically. The first comprehensive expression analysis was performed by Perou and colleagues in the early 2000s, on sixty-five surgical specimens of breast tumors and were able to separate sample subgroups according to gene expression profile clustering [3]. The data has since been added upon by multiple other groups [4,5]. There are five major molecular intrinsic subtypes: luminal A, luminal B, HER2 positive, basal like and normal like. Luminal tumors are driven by the expression of ERα and progesterone receptor (PR). Luminal A has lower Ki67 scores and is associated with less aggressive breast cancers, where luminal B is associated with more aggressive breast cancer and

poorer clinical outcomes [6]. Molecularly, luminal A tumors have less frequent cyclin D1 amplifications and TP53 mutations, but have higher ER signaling [7]. The HER2 subtype has a HER2 gene expression signature and signaling through Map kinase, PI2k/Akt, phospholipase C and STAT pathways, which can be treated by HER2-targeted therapies [6]. Basal-like breast cancer is distinguished by a lack of ERα, PR, HER2 or luminal cytokines. These have higher genome instability and express genes associated with proliferation. Overall, basal-like tumors have poorer prognosis than luminal subtypes [7]. Clinical implications of molecular subtypes were examined, and are commonly called the 'PAM50' based subtypes classified for the original 50 genes used to separate and define the subtypes [8]. The PAM50 gene panel is the first comprehensive test that guides clinical decision for individual breast cancer diagnosis and therapy.

Through sequencing genomic and big data studies, there are many datasets available in breast cancer, such as exome/whole genome sequencing, methylation, transcriptomics and proteomics that allow further characterization of breast cancer. Institutional coordinated efforts and collaboration allow us to better understand the molecular basis of cancer, including the cancer genome atlas (TCGA) and the molecular taxonomy of breast cancer international consortium (METABRIC), as well as ScanB, the Cancer Cell Line Encyclopedia (CCLE), and MET500 [27-30]. These studies have shown PIK3CA, TP53, and GATA3 in more than 10% of breast cancer patients and enrichment of mutations in those genes in the luminal A subtype.

1.3 Histological Subtypes of Invasive Breast Cancer

Breast cancer is a complex disease, made up of multiple histological and PAM50 subtypes with distinct biological features that effect response to therapy and clinical outcomes [9]. Typical classification includes tumor size, lymph node involvement, histological grade, age, estrogen receptor (ER) expression, progesterone receptor (PR) expression, and human epidermal growth factor receptor (HER2) expression. There are two major histological subtypes of invasive breast cancer, invasive ductal carcinoma (IDC) that makes up ~85% of cases and invasive lobular carcinoma (ILC), which makes up 10-15% of cases. In 2020, ILC will affect ~40,000 patients in the United States. It is characterized by a loss of E-cadherin (*CDH1*) in ~90% of cases, which distinguishes it through unique clinicopathological features, such as the single-file growth pattern of small cells, in contrast to the cluster-like, lump pathology characteristic of IDC. Loss of function *CDH1* mutations are found in 50-60% of ILC tumors, thought to be an early event, and typically occur in combination with chromosome 16q loss [10]. Patients with ILC are shown to suffer from late recurrences and metastasis to unique sites such as the ovary and the gastrointestinal tract as well as common sites such as liver, lung and bone [11]. ILC are also almost exclusively ER-positive, and are largely characterized as luminal with characteristics such as ER-responsive genes, and luminal cytokeratins. Due to the unique growth patterns of ILC, there is a high false-negative mammography rate (19-43%) [11]. Classic ILC typically exhibits more favorable prognostic features such as ER+, PR+, HER2-, and low Ki67 index. Despite the different aspects of ILC regarding clinical, histopathological and molecular features, clinical trials and guidelines treat ILC and IDC as a single disease.

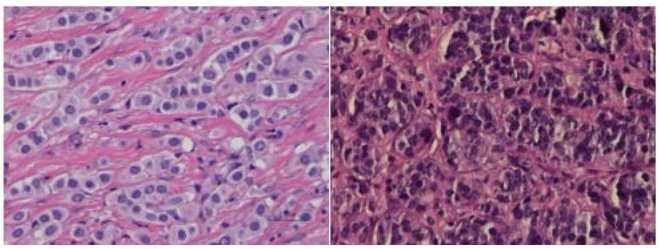


Figure 1 ILC and IDC have Histological Representation in Breast Tissue

H&E stains for ILC and IDC were performed on primary tumors by Ahmed Basudan. The ILC lesion displayed a classic discohesive growth pattern of cells, while IDC cells filled the section of interest.

Table 1 Overview of Characteristics of ILC and IDC

	ILC	IDC
Percent of Breast Cancers (%)	10 – 15	80
Patients Annually Diagnosed in U.S.[1] ***Numbers Calculated from 2021 Diagnoses Projections***	28420 – 42630	227,360
Average Age of Patient Diagnosed[19]	63.4 ± 12.7	59.5 ± 13.6
Growth Pattern	Discohesive linear cords	Distinct masses
Detection[20,21]	Hard to detect clinically and radiographically, Mammography has high false negative rate, MRI (magnetic resonance imaging)	More often palpable during self-examination, detection through mammography, MRI, ultrasound
ERα Expression[20,21]	+ 80-95%	+ 70-80%
Molecular Subtype[22]	Predominantly Luminal A	Luminal A Luminal B HER2 Basal
E-cadherin Expression[23]	Not expressed	Primarily expressed
Primary Therapeutic Modalities (ERα^{+})	Endocrine Therapies, adjuvant therapy	Endocrine Therapies

1.4 Estrogen Receptor Biology

Estrogen receptors are nuclear hormone receptors that play diverse roles in cellular processes, as well as enable extracellular signals in order to regulate tissue-specific intracellular events [12]. The estrogen receptor is made up of two classes, ERα, encoded by *ESR1*and ERβ, encoded by *ESR2*, which mediate the effects of estrogen. ER-α has been reported as the key estrogen signaling receptor in ER+ breast cancer [13]. The ER structure is made up of an A/B domain at the N-terminus, followed by the ligand-independent activation function (AF-1), DNA-binding domain (DBD), a hinge (D domain), then the E/F domain at the C-terminus, which is made up of the ligand-dependent activation function (AF-2), and ligand binding domain (LBD) [31].

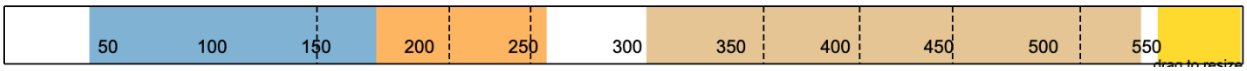


Figure 2 Schematic represenation of *ESR1* genomic and functional structure

The numbers represent amino acid sequence from N-terminus (left) to C-terminus (right). The region A/B contains the activation function 1 (AF-1) domain followed by DNA-binding domain (DBD) and a hinge in regions C and D, respectively. At the C-terminus, the region E/F contains the activation function 2 (AF-2) and ligand-binding (LBD) domains. Plot generated using ProteinPaint software

When ER is activated by a ligand, estrogen, it can dimerize and interact with estrogen response elements to initiate transcription. This is the classic genomic nuclear pathway of ER [32]. Another method, is ER bound to estrogen can interact with other DNA-binding transcription factors, such as AP-1, SP1, CREB, and STAT5, which allows activation of target genes through alternative, non-classical signaling [32]. Another method is that ER can activate

transcriptional activity without a ligand. Activation of growth factors and receptor tyrosine kinases, such as EGFR, HER2, IGF1R and FGFR can activate the ERK and PI3K/AKT pathways, which then can phosphorylate and activate ER [33].

1.5 Breast Cancer Treatments for ER+ breast cancer

Treatment of invasive breast cancer consists of surgical removal of the tumor, either through lumpectomy or mastectomy, and then subsequent treatment of adjuvant therapy, including chemotherapy, endocrine therapy, and radiotherapy, depending on the tumor size and representative biomarkers. Tumors that express the estrogen receptor (ER) account for 75% of breast cancer tumors. The estrogen receptor is a nuclear receptor that was discovered in 1958 by Elwood Jensen [21]. Estrogen acts through two receptors, $ER\propto$ and $ER\beta$, and regulates growth and development through proliferation and differentiation of cells [22]. ER positive tumors are typically treated through endocrine therapy, through selective estrogen receptor degraders (SERDs) and selective estrogen receptor modulators (SERMs), such as fulvestrant and tamoxifen respectively, or by an aromatase inhibitor (AIs), such as letrozole [14,15]. SERDs work by binding to ER, destabilizing the protein and prevents dimerization, leading to protein degradation. SERMs target ER and act as competitive antagonists and were developed on tissue selectivity to target ER signaling[16]. AIs inhibit the enzyme aromatase that converts androgens into estrogen. These are used in postmenopausal women, and are the first treatment for metastasis disease [25]. Tamoxifen blocks the estrogen receptor from binding its ligand estrogen through competitive binding and binding to DNA after metabolic activation, and leads to remission similar to a adrenalectomy or ovarectomy [23]. Fulvestrant was approved by the FDA

in postmenopausal patients with ER-positive metastasis disease [24]. Aromatase inhibitors (AIs) work by inhibiting aromatase in peripheral organs. AIs are typically used in hormone receptor positive post-menopausal women, because the enzyme is responsible for of the majority of estrogen production in this setting [8]. ILC tumors respond poorly to chemotherapy, leading to a mastectomy being a more common treatment. ILC usually respond well to endocrine therapy, with treatments including letrozole, which provide a better overall survival, especially compared to tamoxifen, which has shown an increased incidence of endocrine resistance [18]. HER2 positive tumors are typically treated by targeting HER2, with treatments such as trastuzumab, pertuzumab, lapatinib, and ado-trastuzumab [17]. HER2 is a transmembrane receptor tyrosine kinase, and is associated with a poor outcome. HER2 overexpression occurs in 15-20% of breast cancers [19]. It is activated with dimerization with other members of the EGFR/ErbB family members, which then triggers the PI3K/AKT, Raf/MAPK downstream pathways. HER2 pathway mediates mitogen-activated protein kinase (MAPK) and phosphatidylinositol (PI3K) pathways, and is important for cell growth, survival and differentiation [20].

Invasive lobular carcinoma is 95% estrogen receptor positive, is strongly estrogen driven and has unique context for estrogen receptor signaling. It is sensitive to and dependent on estrogen. From this, it supports the theory that ILC patients are ideal candidates for anti-estrogen therapies, but there is a large difference in the benefits compared to patients with IDC treated the same way. ILC patients treated with tamoxifen have poorer outcome and long-term outcomes >5-10-year post-diagnosis. In clinical studies, ILC models have shown tamoxifen resistance and *de novo* ER partial agonism when treated with anti-estrogens [26]. In an adjuvant BIG 1-98 trial, Metzger-Filho and his group investigated the outcome of treatment in IDC (n=2,599) and ILC (n=324) patients. Patients with ILC that are treated with tamoxifen are much more likely to

experience recurrence, suggesting that they receive less benefit from adjuvant therapy. There is a need to further investigate signaling pathways, therapeutic targets in ILC [27]. Understanding the endocrine response in ILC is essential in order to improve patient outcomes.

1.6 Genetics of Estrogen Receptor Positive Breast Cancer

To understand the genetics, biology and mechanism of invasive breast cancer and other diseases, there is a plethora of data available from TCGA, METABRIC, RATHER and CCLE, as the most comprehensive large-scale studies in breast cancer, with a focus on primary breast cancer tumors. From these studies, the most common mutated genes in early breast cancer pathogenesis were PIK3CA, PTEN, AKT1, TP53, GATA3, CDH1, RB1, MLL3, MAP3K1, and CDKN1B [38].

ILC is well characterized as having alterations in *CDH1*, copy number alterations with the gain of chromosome 1q, loss of chromosome 16q and amplifications in the *FGRF1* and *CCND1* locus. There are also mutations in driver genes *PIK3CA, PTEN, AKT1, TBX3, FOXA1*, and *ERBB2/3* which are seen in multiple studies. Common mutations found in ILC are *FOXA1* and *GATA3*, and in over half of all ILC samples in luminal A tumors, *PIK3CA, PTEN* and *AKT1* are mutated. ILC tumors can be more severe due to mutations in *TP53, ESR1*, and *ERBB2*, all of which drive more aggressive tumors and are also linked to endocrine therapy resistance [34].

In a comprehensive study by Cirello, they profiled 817 breast cancer tumors and 633 cases by RPPA, IDC (n=490), ILC (n=127) and mixed (n=88), other (n=112). Lobular tumors were primarily luminal A, ER+ tumors, characterized by low levels of proliferation markers. The

most significant commonly mutated genes in luminal A ILC were PIK3CA, RUNX1, CDH1, TP53, TBX3, PTEN, FOXA1, and MAP3K1. In luminal A IDC, PIK3CA, TUNX1, CDH1, TP53, PTEN, MAP3K1, GATA3, AKT1, NBL1, KMT2C, DCTD, SF3B1, CBFB, MAP2K4, KRAS, HLA-DRB1, CTCF, CDKN1B, GRIA2 and NCOR1 are the most significant. PTEN inactivation is a defining distinct feature between luminal A ILC and IDC, where there were PTEN genetic alterations across all ILC cases, including homozygous deletions, and somatic mutations. They also noted increased Akt signaling in ILC versus IDC, which is the most altered pathway in cancer that provides tumor cells with enhanced growth and survival capabilities. This makes the PI3K/Akt pathway a potential strategy for selective inhibition [40].

The main drivers in IDC are TP53 and PIK3CA mutations, which promote tumor progression, and MYC amplification. TP53, PIK3CA, GATA3, MAP3K1, and MAP2K4 are the most significant mutations and MYC, TRPS1, ANKRD46, ZBTB10 and ERBB2 gene amplification are the mostly highly associated with IDC histological grade. Although ILC and IDC tumors do have mutation and amplification of genes in common, they should not be treated the same way [39].

Table 2 Overview of Recurrently Mutated Genes in Breast Cancer

		ILC	IDC
Selected Top Mutated Genes from TCGA Study[40] ***Data are Adapted from the Above Publication*** ****Comparison of Luminal A Subtypes***	***PIK3CA***	9.18E-13	6.79E-13
	RUNX1	9.18E-13	1.32E-05
	TP53	2.22E-04	**6.79E-13***
	TBX3	**4.01E-06***	*NS*
	PTEN	**8.86E-09***	5.63E-03
	FOXA1	**6.53E-04***	*NS*
	GATA3	*NS*	**6.79E-13***

1.7 Methylation in Breast Cancer

Breast cancer is thought to be driven by both genetic and epigenetic alterations, where epigenetic alterations are reversible and susceptible to environmental factors. DNA methylation of "CpG islands" is a commonly studied epigenetic modification. Hypermethylation of a CpG island in the promoter region of a gene represses its expression. Methylation of APC, CDH1, and CTNNB1 were found to be related to breast cancer development, which suggests that progression and prognosis of breast cancer could be influenced by DNA methylation status. DNA methylation is important for cellular processes, including embryonic development, genomic imprinting, cell differentiation and senescence, and deregulation. DNA methylation markers are also chemically and biologically stable, so they are practical as diagnostic and prognostic markers [35].

Epigenetic markers, such as DNA methylation, have shown to be deregulated in cancer. This is typically an early event. Hypermethylation of CpGs in promoters, hypomethylation of non-CpG islands and increase in the variation in methylation are all cancer specific changes. Methylation patterns are also shown to be associated with hormone receptor status, TP53 mutation status, histological grade, stage and survival [36]. DNA methylation alterations are potential markers for ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) for outcome prediction, and could be used to identify cellular changes that effect tumor progression. Identification of methylation alterations in early lesions, can contribute to breast tumorigenesis and insight in the development of invasive disease. Progression in invasive breast cancer represents and increase of the change of mortality, and identification of disease progression markers is needed. Further examination of alterations is needed to see their utility in defining risk in invasive breast cancer, which would impact treatment decisions [37].

1.8 Identifying Unique Epigenetic Changes in ILC

The major epigenetic changes dysregulated in breast cancer are DNA methylation, histone modification, and nucleosome remodeling. Genes involved in proliferation, anti-apoptosis, invasion and metastasis have shown to undergo epigenetic changes in breast cancer. Due to these changes being reversible, there are efforts to understand the mechanism and target changes for a therapy. DNA methylation has led to transcriptional silencing of tumor suppressor and growth regulatory genes in breast cancer [40]. *CDH1*, a cell-to-cell adhesion protein that is mutated in ILC is hypermethylated in the promoter region in breast cancer, causing diverse biological implications. There is evidence of CpG methylator phenotype I liminal subtypes,

where there are fewer gains in CpG methylation in basal-like subtype [41]. We found a unique gene TFAP2B differentially methylated in ILC and IDC, correlated with change in expression. There are few genes that are distinct between the histological subtypes, so this holds value as a potential biomarker or therapeutic target. It allows us to further understand differences between ILC and IDC and how to target ILC differently through precision medicine.

1.9 TFAP2B

TFAP is a family made up of five members: TFAP2A, TFAP2B, TFAP2C, TFAP2D, and TFAP2E. They require homodimerization or heterdimerization with other members of the family. They interact through a helix-span-helix motif at the carboxyl terminus [42]. Each helix is an amphipathic alpha helix made up of ~80 AAs. This motif mediates protein dimerization and DNA binding of GC-rich regions [43]. Mutations in TFAP2B are associated with Char syndrome, which has 3 features: facial abnormalities, newborns suffer from abnormal connection between the aorta and pulmonary trunk, and most patients are born with abnormalities of the hand. Fusion of PAX3 and FKHR is a driver in this mutation. TFAP2B plays an anti-apoptotic tole with this protein fusion Missense mutation in conserved residues still dimerize with normal and mutant forms of the protein in vitro, but have poor binding of TFAP2B target genes. It has also shown to play a role in rhabdomyosarcoma, a childhood cancer of skeletal muscles. The AP-2 family members can activate the promoter of growth and differentiation related to *p21, ERα,* and *HER2*. These transcription factors also regulate multiple pathways and are involved in development, cell growth, differentiation and apoptosis. Knockdown of TFAP2B is limited due to death almost immediately after birth, probably due to polycystic kidney disease and epithelial

conversion of mesenchymal tissue [44]. In mice, loss of TFAP2B causes death due to apoptosis of epithelial cells in collecting ducts or lethal renal failure.

1.10 TFAP2

TFAP2B is one member of the AP2 transcription factor, made up of five transcription factors, TFAP2A, TFAP2B, TFAP2C, TFAP2D and TFAP2E, that orchestrate various cell processes. Studies have revealed that the AP-2 family members are encoded by separate genes where AP-2A, AP-2B, AP-2C, AP-2D, and AP-2E are located on chromosomes 6p24, 6p12, 20q13.2, 6p12 and 1p34 respectively [62]. These transcription factors have a conserved helix-span-helix dimerization domain, after a DNA-binding and transcription domain. The AP-2 proteins can form homo or heterodimers, which bind GC-rich DNA sequences of target genes, which act as both activating and repressing stimuli [62]. The TFAP2 factors orchestrate cell processes through regulation of expression of downstream genes and influence cell induction, differentiation, survival, differentiation and apoptosis. TFAP2 factors are also critical in regulating tissue development during embryogenesis. During this development, TFAP2 control the balance between proliferation and differentiation, confirming their roles in cell growth, differentiation, and apoptosis [63].

1.11 TFAP-2 Known Mechanism Of Action

In lung adenocarcinoma, TFAP2B was shown to regulate the ERK/p38 MAPK signaling pathway. It was shown that TFAP2B could be acting partially through the modulation of the ERK/p38 signaling pathway, which effects the regulation of tumor cell growth. To examine other pathways that play a role in tumor growth, they then examined the effect of TFAP2B on VEGF and PEDF proteins and found that TFAP2B regulated the VEGF/PEDF ratio, suggesting that TFAP2B impacts tumor growth and angiogenesis [61]. There are previous reports on TFAP2A and TFAP2C, where they have been shown to participate in tumorigenesis through controlling the expression of genes such as VEGF, P21, Rb, TP53, Era, BCL2, cKIT, MMP-2, E-cadherin, and c-myc [61]. TFAP2A and TFAP2C have been shown to induce changes in chromatin structure associated with *ER*∝ and there is a regulation loop between the Ap-2 proteins and ER/estrogen, where ER has extinguished some AP-1A transactivating properties and AP-2A mRNA was downregulated in response to estrogen treatment. In this study, TFAP2C showed increased expression, with no significant increase in TFAP2B expression, suggesting that AP-2C is the primary estrogen responsive gene [62]. AP-2 can activate the promoter of multiple growth and differentiation related genes, including *p21*, *ER*∝, and *HER2* [62]. Additionally, TFAP2A, TFAP2B, and TFAP2C are expressed in breast tissue and possibly could coordinate with *HER2* and *ER* [62]. In epithelial cells, reexpression of Rb or c-myc allows the reexpression of CDH1, mediated through AP-2 through the interaction of the N-terminal domain of AP-2 and the C terminal and oncoprotein binding domain of Rb. This suggest that AP-2 and E-cadherin are involved in a regulatory pathway [63].

1.12 TFAP2B in Breast Cancer

TFAP2B is rarely reported in breast cancer, although the AP-2 family members are essential for maintaining cellular homeostasis. In a paper by Zhenglin Li and his group, they saw that TFAP2B plays a tumor promoting function in breast cancer and there was higher expression in LCIS compared to DCIS. When they silenced expression, they saw inhibition of cell proliferation, colony formation, migration and invasiveness as well as down regulated expression of p57, MMP-2, MMP-9, C-Jun, p-ERK and STAT3. *In vitro*, TFAP2B knockdown blocked tumor growth, but they also noted that high TFAP2B expression was correlated with poor prognosis and advanced malignancy in patients. From this, we can confer that TFAP2B plays a role in tumor growth, and since it is associated with poor prognosis, it could be a potential therapeutic target [45].

1.13 Study Objectives

To determine if we could determine lineage based on methylation status, we took TCGA data and found differentiated methylated probes, aligned it with genomic data and found genes that were differentially methylated between ILC and IDC. From this, one gene was uniquely methylated between the two histology's, where TFAP2B expression is silenced in IDC and expressed in ILC, as the primary tumors and hypermethylated and hypomethylated respectively. To determine what this means from a mechanistic and phenotypic standpoint, we plan on exploring the implications in the following experiments.

    - Determine the effect of loss of TFAP2B on growth and survival of breast cancer cell lines in 2D and 3D conditions through growth, colony formation, and apoptosis and cell cycle assays
    - Determine the effect of loss of TFAP2B on estrogen and antiestrogen response
- Assess gene regulation by TFAP2B in breast cancer cell lines
    - Identify downstream targets of TFAP2B using an unbiased MOTIF analysis and a FIMO analysis of publicly available ATACseq data.
    - Identify genes co-expressed with TFAP2B in breast cancer cell lines and tumors, using a series of publicly available data sets, and in house generated data.
    - From the previous results, predict which genes and pathways play a role in mediating the effects of TFAP2B in ER+ LumA ILC

2.0 Materials and Methods

2.1 Tissue Culture

MCF-7, T47D, MDA-MB-134-VI, HCC2185, and MDA-MB-453 cells used in this study were obtained from the American Tissue Culture Collection (ATCC). Sum44PE cells were obtained from Asterand. Cultures were maintained as follows: MDA-MB-453 in Leibovitz-15 Gibco #11415-064) media supplemented with 10% FBS, HCC2185 in Roswell Park Memorial Institute (RPMI) 1640 Medium (Gibco #16600-082) supplemented with 5% FBS, SUM44PE, SUM44PE EV, and SUM44PR CDH1 OE in :1 mixture of DMEM and Ham's F-12 (F-12) (Gibco#11330-032) media supplemented with 2% dextran charcoal stripped fetal bovine serum (CSS; Gibco#12676-027, Lot#1747185 or Gemini#100-119, Lot#A67F02H), 5 mM Ethanolamine (Sigma#E0135-100ML), 1 µg/mL Hydrocortisone, 5 µg/mL Transferrin (Sigma#T2252-100MG), 10 nM Triiodothyronine (Sigma#T5516-1MG), 50 nM Sodium Selenite (Sigma#S9133-1MG), and 5 µg/mL Insulin, MDA-MB-134 (MM-134), MM134 EV, and MM134 CDH1 OE in a 1:1 mixture of DMEM with Leibovitz-15 media supplemented with 10% FBS, T47D, and T47D CDH1 KO cells in RPMI supplemented with 10% FBS, and MCF-7, and MCF7 CDH1 KO cells in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (Gibxo #26140-079). All cell lines were authenticated at the University of Arizona Genetics Core.

Routine culture was performed by removal of media, gentle rinsing of cells in 1X DPBS (Corning CellGro#21-031-CV), and treatment with 0.25% Trypsin-EDTA (Gibco#25200-056) until cells were in single-cell suspension. Trypsin was deactivated with base media. Cells were

always centrifuged at 1400 rpm for 4 minutes at room temperature to completely remove trypsin prior to plating. Plating densities for given assays are included in each respective section, while routine maintenance was performed with no harsher than 1:2 or 1:3 splits for ILCs and 1:5 to 1:20 cells for IDC or control cell lines, as deemed necessary.

2.2 Compounds and Reagents

17β-estradiol (E2) and fulvestrant (Ful) were obtained from Sigma and Tocris, respectively. TFAP2B siRNA smartpool was purchased from Dharmacon (Thermo #L-017730-00-0005). Lipfectamine&trade RNAiMAX was purchased from Invitrogen (#13778150).

2.3 Western Blot

Protein lysates were isolated using RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40 (Sigma #21-3277), 0.5% NaDeoxycholate, 0.1% SDS), supplemented with Protease and Phosphatase Inhibitor (Thermo #78442), sonicated in a cup horn sonicator (for 5 minutes in 30 second pulses), and centrifuged for 15 minutes at 14,000 rpm at 4°C. All samples were quantified for protein concentration using BCA Assay (Pierce #23225) and 30 - 50μg (actual amount noted in figure legend) were run on an 10% SDS-PAGE gel. Protein was then transferred to a PVDF membrane (Millipore #IPFL00010) and incubated in Odyssey PBS Blocking buffer (LiCor #927-40000) for one hour and probed with antibodies. Primary antibodies used in this project are as follows: TFAP2B (Cell Signaling Technology

#2509S) and β-actin (Sigma #A5441). All primary antibodies were probed overnight at 4°C. After removing primary, membranes were washed with PBST (1X PBS, 0.1% Tween 20) three times, for 5 minutes each, and then incubated with secondary antibody at 1:20,000 at room temperature for 55 minutes. Secondary antibodies used were anti-mouse 800CW (LiCor#925-32210) and anti-rabbit 800CW (LiCor#925-32211). Membranes were again washed 4 times in PBST prior to imaging on Odyssey Infrared Imaging System (LiCor). For blots that were stripped and re-probed, 1X NewBlot Stripping buffer (LiCor #928-40032) was used per manufacturer's protocol.

Alternative method was protein lysates were isolated using RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40 (Sigma #21-3277), 0.5% NaDeoxycholate, 0.1% SDS), supplemented with Protease and Phosphatase Inhibitor (Thermo #78442), sonicated in a cup horn sonicator (for 5 minutes in 30 second pulses), and centrifuged for 15 minutes at 14,000 rpm at 4°C. All samples were quantified for protein concentration using BCA Assay (Pierce #23225) and 30 - 50μg (actual amount noted in figure legend) were run on an 10% SDS-PAGE gel. Protein was then transferred to a PVDF membrane (Millipore #IPFL00010) and incubated in 5% BCA for one hour and probed with antibodies. Primary antibodies used in this project are as follows: TFAP2B (Cell Signaling Technology #2509S) and β-actin (Sigma #A5441). All primary antibodies were probed overnight at 4°C. After removing primary, membranes were washed with TBST (50mM Tris, 150mM NaCl, 0.1% Tween 20, pH 7.4)

Four times, for 5 minutes each, and then incubated with secondary antibody at 1:20,000 at room temperature for 55 minutes. Secondary antibodies used were Immun-Star Goat Anti-Rabbit(GAR)-HRP Conjugate (Bio-Rad #1705046) and Cytiva's Amersham ECL Mouse IgG,

times in TBST for 5-minute increments. Membranes were prepared using Clarity Western ECL Substrate, 500 mL (Bio-rad #1705061) and imaged.

Table 3 Antibodies used in IB experiments

Protein	**Company**	**Catalog Number**	**Host Species**	**Dilution**
Actin	Sigma	A5441	Mouse	1:5000
TFAP2B	Cell Signaling Technology	2509S	Rabbit	1:1000

2.4 Proliferation Assays

MM134, SUM44 PE cells were plated at 15,000 cells/well, MDA-MB-453 cells were plated at 7,500 cells/well, in 96-well 2D (Fisher #353072) or flat bottom ultra-low attachment (ULA) (Corning #3473) plates. Cells were seeded in full serum media and read over a 7-day period at days 1,3,5 and 7 with Cell Titer Glo. Vehicle controls were always included, as well as media blank wells, and unused wells were given equal volume 1X DPBS to keep consistent surface tension across the plate, and to avoid evaporation of inner wells. (CTG). experimental plates were obtained from the incubator and brought to room temperature without exposure to light for approximately 30 minutes. Then, room temperature reagent from the CellTiter-Glo Luminescent Cell Viability Assay kit (Promega$^{\#}$PR-G7573) were prepared, applied, and stored per manufacturer's protocol and added in a 1:1 ratio to well contents with rocking at 2 minutes without exposure to light to lyse cells, and further incubation at room temperature for 10 minutes to stabilize luminescence. Contents were then transferred to white walled luminescent plates

(Corning#3912) and were measured on a Promega Glo-Max Microplate Reader. Data were corrected for each experiment by subtracting average background fluorescence (2D) or luminescence (ULA) from values, and normalizing to respective vehicle controls as "1." Outliers were removed using Grubb's Test at $\alpha = 0.05$. Error was displayed as standard deviation of the mean (STDEV).

2.5 Hormone Deprivation

Hormone deprivation was performed for all estrogen response experiments. Cells were regularly maintained in FBS as described in cell culture. Shortly before the experiment, cells were split into appropriate vessels with 80-90% cell confluency. Start from the day next, cell culture medium was switched to phenol-red-free IMEM (Gibco, A10488) with 10% CSS-stripped serum (CSS, Gemini, #100-119). Medium were changed three times a day for three consecutive days followed by an additional PBS wash each time.

2.6 Generation of transient knockdown TFAP2B cell models

Cells were reverse transfected by Lipofectamine™ RNAiMAX manufacturer's protocol (Thermo Fisher Scientific#13778-150). For preliminary assessments, 25, 50, or 100 nM were utilized to assess best dose. For subsequent assays, 50 pmol were utilized for *TFAP2B* knockdown assays. For si*TFAP2B* assays, ON-TARGETplus Non-targeting Pool (Dharmacon#D-001810-10-05) was used. For si*TFAP2B* assays Dharmacon (Thermo #L-

017730-00-0005) was used. In basal expression assays, si*RNA* complexes were generated in Opti-MEM® I (Gibco#31985-070) for 48-72 hours prior to assessment of downstream applications. In deprivation assays, si*RNA* complexes were generated in Opti-MEM® I without phenol red (Life Technologies#11058-021).

Table 4 Reagents used for transient knockdown cell models

Reagent	Company	Catalog Number
TFAP2B Smartpool	Dharmacon (Thermo)	L-017730-00-0005
Lipofectamine RNAiMAX	Invitrogen	13778150
Non-targeting siRNAs	Dharmacon (Thermo)	D-001810-0X

2.7 Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

MM134, SUM44PE, HCC2185, and MDA-MB-453 cells were seeded into 6-well plate with 1.5M. 1.5M, 1.5M, 750k cells per well respectively with biological duplicates. After desired treatments, RNAs were extracted from each sample, and cDNA was synthesized using iScript kit (#1708890, BioRad, Hercules, CA). qRT-PCR reactions were performed with SybrGreen Supermix (#1726275, BioRad), and the ΔΔCt method was used to analyze relative mRNA fold changes and *RPLP0* levels were measured as the internal control. Relative mRNA levels between conditions were compared using unpaired t-tests.

Table 5 qRT-PCR primers

Gene	Forward Sequence (5' → 3')	Reverse Sequence (5' → 3')
TFAP2B 1	TGAAGATGCCAATAACAGCGGCA	GAGCAACAATATCGGT
TFAP2B 2	CTATGAGGACCGGCACGATG	TGCCGCTGTTATTGGCATCT
TFAP2B 3	TTGAACCGGCAGCACACA	CTTGGTGGCCAACAGCATATT
RPLP0	TGGCAATCCCTGACGCACCGC	GCTGCATCTGCTTGGAGCCCACA
GADPH	GGAGCGAGATCCCTCCAAAAT	GGCTGTTGTCATACTTCTCATGG

2.8 RNA-sequencing analysis – cell lines*

RNA-seq library was prepared by UGC using KAPA RNA hyperPrep kit and sequencing was performed obtaining >100 M reads per sample with UPMC genome center (UGC) (Pittsburgh, PA). Salmon was used for quantification of the transcripts using default options and hg38 genome build as the reference [20]. Log2 (TPM+1) value of each gene was used for output and downstream analysis. Genes with maximum transcripts per million (TPM) <1 across all samples were excluded from further analysis due to low gene expression. R was used for statistical analysis, and for plotting of expression levels.

*RNAseq analysis conducted by Anokhi Kashiparekh, B.S. University of Texas

2.9 RNA-sequencing analysis – publicly available datasets

Salmon was used for quantification of the transcripts using default options and hg38 genome build as the reference [20]. Log2 (TPM+1) value of each gene was used for output and downstream analysis. Genes with maximum transcripts per million (TPM) <1 across all samples were excluded from further analysis due to low gene expression. R was used for statistical analysis, and for plotting of expression levels.

2.10 Dose Response Assays

Cells were seeded at 5,000 cells/well for IDCs or 15,000 cells/well for ILCs in six replicate wells of a 96 welled 2D (Fisher#353072) or flat bottomed ultra-low attachment (ULA) (Corning#3473) plate for each concentration of drug tested. Vehicle controls were always included, as well as media blank wells, and unused wells were given equal volume 1X DPBS to keep consistent surface tension across the plate, and to avoid evaporation of inner wells. Cells were seeded 1 day prior to treatment and typically collected at day 7. Presto blue dye was added at 1:10 ratio and measurements assessed for fluorescence per manufacturer's protocol of the FluoReporter Blue Fluorometric dsDNA Quantitation Kit (ThermoScientific#F2962) on a PerkinElmer 2030 Multilabel Reader with VictorX software. Data was corrected for each experiment by subtracting average background fluorescence, and normalizing to respective vehicle controls as "1." Outliers were removed using Grubb's Test at $\alpha = 0.05$. Error was displayed as standard deviation of the mean (STDEV).

2.11 Cell Cycle

Cells were plated in triplicate at 1.5M cells/well in 6 welled plates, treated per experimental design siRNA transient knockdown or scramble NTC at 50 pmol, and were collected by standard trypsinization and centrifugation, diluted to 1 x 10^6 cells/mL in standard media, with pellets were treated to 20 µg/mL Hoechst 33342 (Life Technologies$^{\#}$H3570) at 37°C, 5% CO_2 while protected from light for 30 minutes. Just prior to measuring, 2.5 µg/mL Propidium Iodide (PI; Sigma$^{\#}$P4170) were spiked into the mixture. Cells were quantified on a BD LSR II Flow Cytometry machine.

Table 6 Cell Cycle Cytometer Gating

Cell Line	FSC	SSC	PE	APC
MDA-MB-134-VI	295	315	500	358
MDA-MB-453	300	340	420	342
HCC2185	300	340	420	342
SUM44PE	300	340	420	342

2.12 Apoptosis

Cells were plated in triplicate at 1.5M cells/well in 6 welled plates, treated per experimental design siRNA transient knockdown or scramble NTC at 50 pmol, and were collected by standard trypsinization and centrifugation, diluted to 1 x 10^6 cells/mL in standard media, with pellets were treated to 100 uL master mix, made up of 1X binding buffer, 5 uL APC annexin V (BD Pharmingen) and PI (Sigma#P4170) at room temperature for 15 minutes. Control was a well in 6-well plate treated with 70% ethanol for 90 minutes. Cells were quantified on a BD LSR II Flow Cytometry machine.

Table 7 Apoptosis Cytometer Gating

Cell Line	**FSC**	**SSC**	**PE**	**APC**
MDA-MB-134-VI	320	360	490	470
MDA-MB-453	334	360	490	430
HCC2185	334	360	500	450

3.0 Results

3.1 TFAP2B is uniquely methylated in invasive lobular breast cancer

In a study investigating if lineages can be defined using DNA methylation patterns to separate invasive lobular and invasive ductal carcinoma, the TCGA breast cancer cohort with a specific focus on ER+ luminal A tumors was examined. The TCGA Infinium human methylation array 450k data was utilized to elucidate the most variable probes differentiating ILC and IDC, from which the differentially methylated probes were isolated and then clustered based on methylation and expression patterns (Figure 3A). From this, TFAP2B was found to be an outlier, hypomethylated in ILC compared to IDC and a clear separator in lineage DNA methylation markers (Figure 3B). From this, we looked at the differentially expressed probes and their effect on the expression levels of TFAP2B in primary tumors. As shown in Figure 3C, TFAP2B is primarily hypomethylated in ILC tumors and exhibits higher TFAP2B expression compared to IDC. Copy number variations (CNVs) in genes that are involved in the development and progression of cancer can alter gene expression and drive disease. In Figure 3D, probe cg25593948 was used to further explore methylation levels and expression levels of TFAP2B in primary tumors, which demonstrates that ILC tumors are less methylated and also exhibit higher TFAP2B expression. To validate that the expression of TFAP2B is largely influenced by DNA methylation, we analyzed CNVs in breast cancer cell lines with data from the Broad Institute [46]. We observed a negative correlation between TFAP2B CNVs and mRNA expression, except in ER negative IDC cell lines, but the results were not significant (Figure S1).

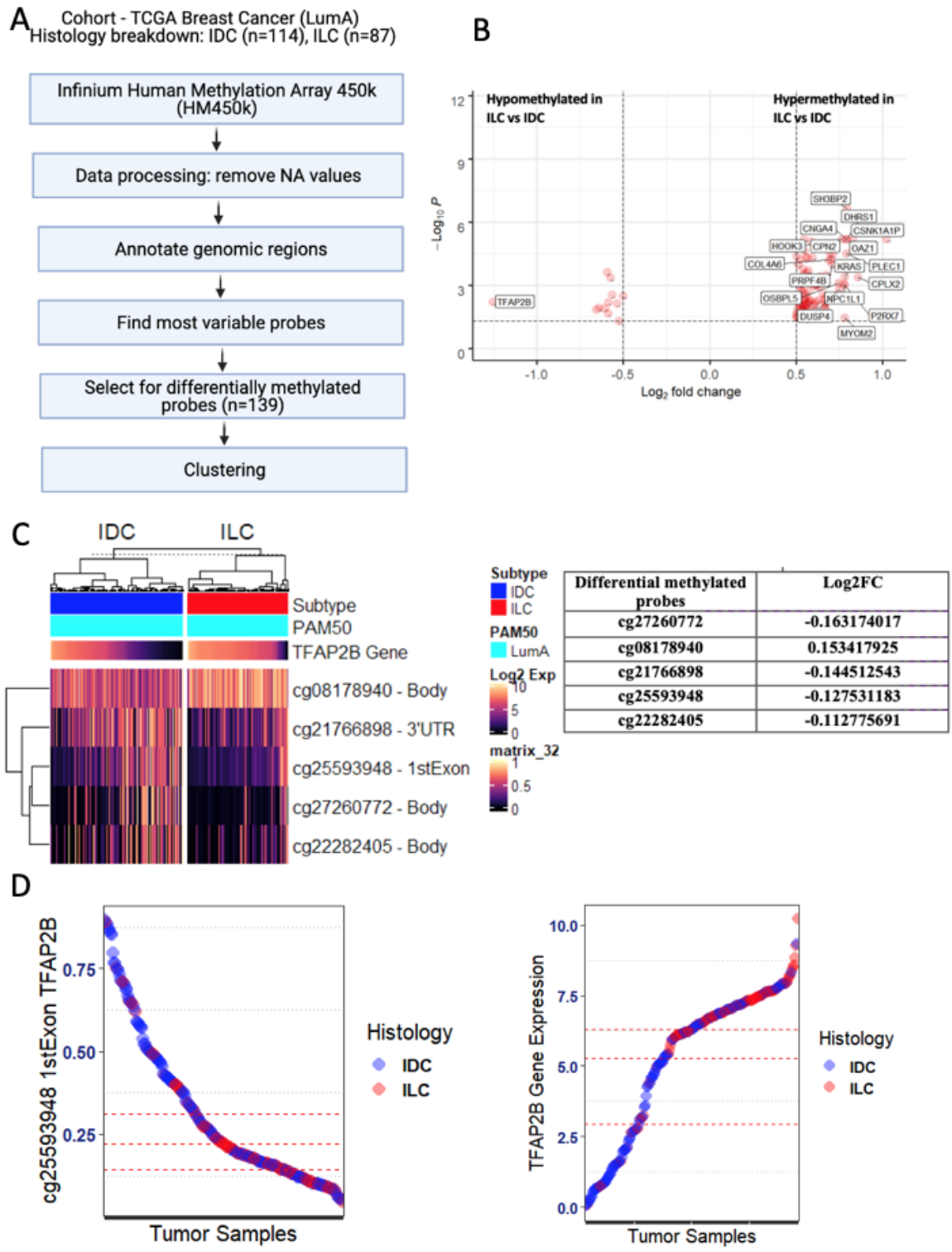


Figure 3 DNA methylation analysis in primary breast tumors from TCGA reveal that TFAP2B is uniquely methylated in ILC

(A) Method for analyzing Infinium human methylation array 450k for luminal A cohort (B) Clustering of differentiated methylated probes in ILC and IDC lineages (C) Supervised clustering of top differentially methylated probes of TFAP2B gene expression in ILC and IDC primary tumors (D) Methylation patterns and gene expression for differentially methylated probe cg25593948. Analyses were performed in collaboration with Osama Shiraz, BS.

3.2 TFAP2B expression in primary tumor cohorts confirms upregulation in ILC

In order to validate our findings in Figure 3, we explored the expression of TFAP2B in a plethora of database, in primary tumors and cell lines. TCGA and METABRIC as well as cell data from CCLE and publicly available RNAseq cell datasets were utilized [48-51]. TFAP2B is significantly more highly expressed in ILC tumors, data taken from TCGA and METABRIC (Figure 4A), and in cell lines (Figure 4B). In Figure 4B, where the expression data is further divided into subgroups, *TFAP2B* exhibits the largest difference in luminal A tumors in ILC, followed by luminal B and normal tissue. There is higher *TFAP2B* expression in ILC in the Her2 subtype in the METABRIC dataset, but the sample size is small. In paired primary and metastatic tumors from an in-house cohort, TFAP2B was expressed more highly in the primary tumor compared to its metastasis (Figure S3), whereas expression in patient derived organoids (Figure S2) was mixed.

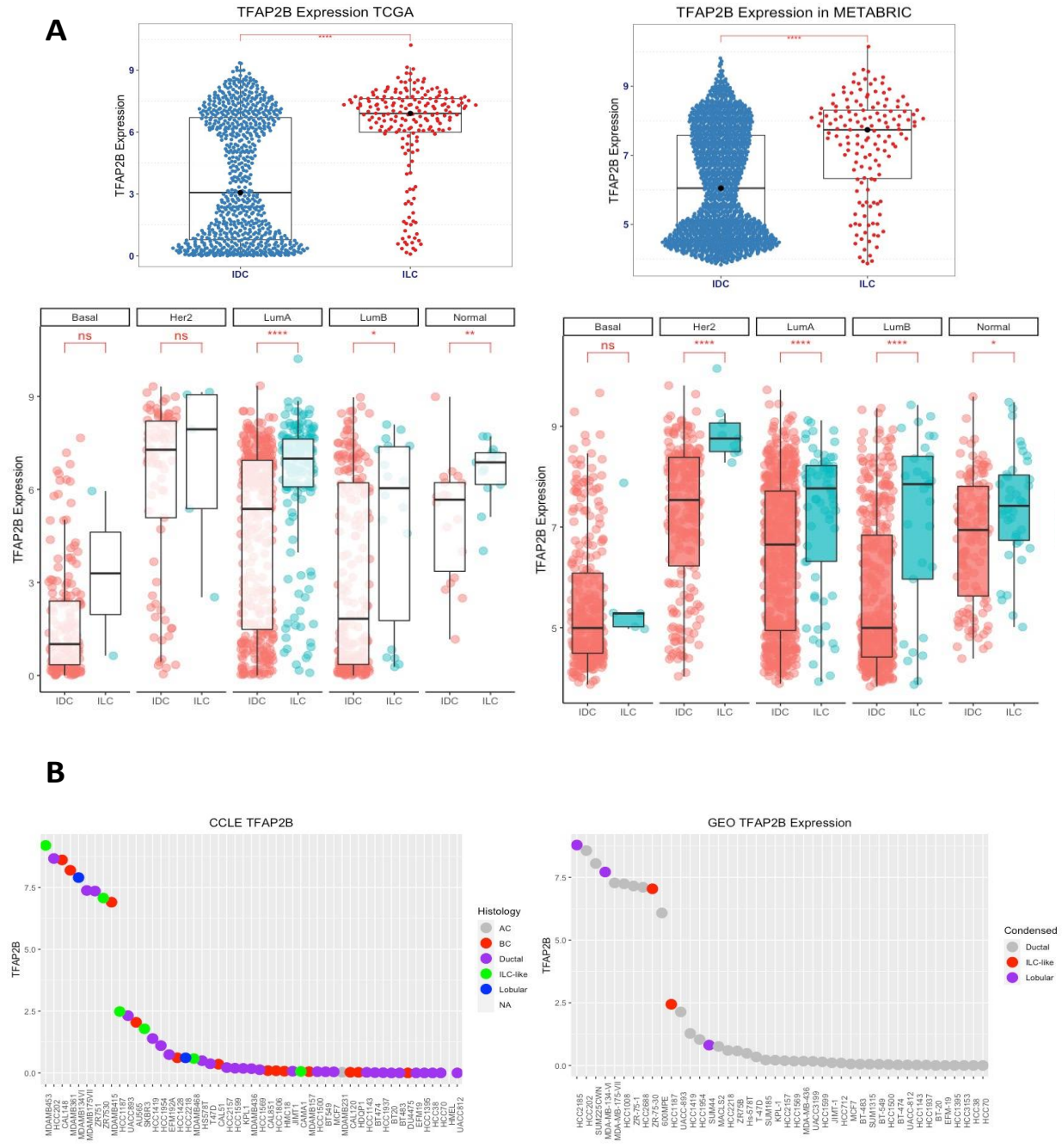


Figure 4 TFAP2B exhibits higher expression in ILC LumA primary tumors and in ILC cell lines

(A) Method for analyzing Infinium human methylation array 450k for luminal A cohort (B) Clustering of differentiated methylated probes in ILC and IDC lineages (C) Supervised clustering of top differentially methylated probes of TFAP2B gene expression in ILC and IDC primary tumors (D) Methylation patterns and gene expression for differentially methylated probe cg25593948. Analyses were performed in collaboration with Osama Shiraz, BS.

3.3 Analysis of the epigenetic alterations of the TFAP2B locus reveals expression is silenced through methylation

Since Figure 3 showed TFAP2B methylation and silenced expression in IDC, we next aimed to investigate if epigenetic alterations of TFAP2B are associated with aberrant *TFAP2B* expression. In Figure 5A, we first ranked cell lines according to their methylation status to rationalize which cell lines would regain the most expression. Cell line methylation data came from methylation profiling of 55-well characterized breast cancer lines using Illumina HumanMethylation27 BeadChip platform [46]. The arrows indicate the methylation status of the cell lines selected for further study, where ILC cell lines are denoted in red, IDC in blue and 'ILC-like' in purple. To confirm and test the methylation status of these cell lines, we treated them with DNA methyltransferase 5-aza-2'-deoxycytidine (DAC) (Figure 5B) for 3 days and then assessed gene expression of TFAP2B. We hypothesized that we may see a recovery of TFAP2B expression in cell lines that previously did not exhibit any detectable expression due to hypermethylation, such as MCF7 and T47D. As shown in Figure 5C, SUM44PE and MCF7 had the highest methylation status and when treated with DAC, they showed the largest increase in expression recovery. We expected to see a higher TFAP2B expression level in T47D when treated, but as seen in Figure 5A, T47D has a lower methylation level (~0.4), so there is less of a recovery effect. MM134 methylation level (~0.1), HCC2185 (~0.43), also showed limited

recovery effect. In HCC2185, there was higher *TFAP2B* levels in the DMSO control, suggesting that DAC has some alternative demethylating effects in this cell line.

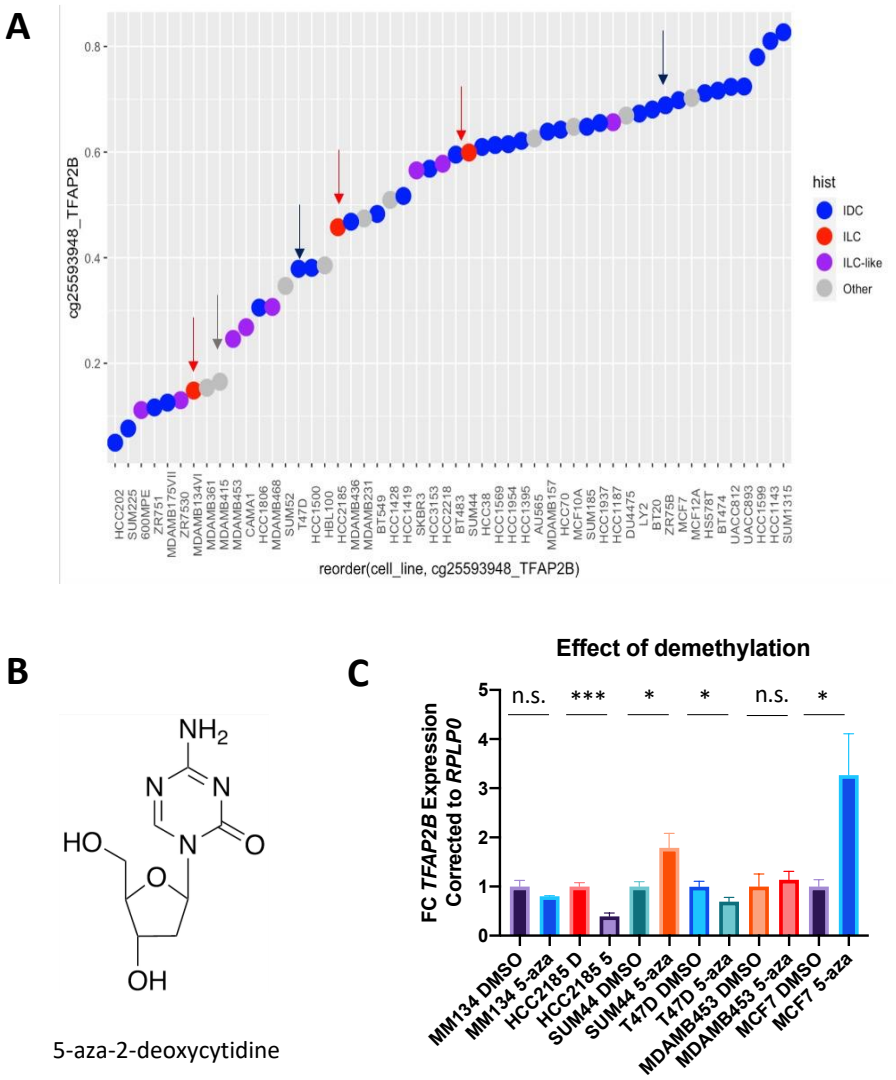


Figure 5 Examining the epigenetic alterations effect on TFAP2B expression

(A) The methylation status of cell lines was mapped to one of the top 5 differentially methylated probes, cg25593948, utilized in Figure 1. Arrows indicate the methylation status of the cell lines selected for further study, where ILC cell lines are denoted in red, IDC in blue and 'ILC-like' in purple (B) Demethylating agent utilized (C) qPCR data (n=1) assessing recovery of TFAP2B expression in cell lines following DAC treatment for 3 days. *TFAP2B* expression was normalized to *RPLP0* and each cell line was normalized to DMSO control. Students t-test (*p<0.05, **p<0.005, ***p<0.0005) was used to compare control and treated group. Error bars represent s.e.m.

3.4 TFAP2B expression in cell lines

From our preliminary methylation study and expression in primary tumors, metastatic tumors and in human cell lines, TFAP2B expression differs significantly, with predominately higher expression in ILC. We explored the mRNA and protein expression of TFAP2B in a panel of cell lines in order to select models for further phenotypic characterization of TFAP2B. We included ILC, 'ILC-like', which have ILC characteristics, predominantly E-cadherin loss, as defined by RATHER [57], and IDC cell lines. Previously, RNAseq was used to investigate transcriptome levels (Figure 4) so to validate this data we performed qRT-PCR and western blotting. The models we chose were MDA-MB-134-VI (MM134), HCC2185, MDA-MB-453 (MDA-MB-453), and SUM44PE (SUM44). These are appropriate models because three are ILC and one 'ILC-like' that have sufficient expression of TFAP2B, detectable at a transcriptome and protein level. SUM44 has less expression, but is still detectable. ILC cell models exhibit moderate to slow *in vitro* growth characteristics. We first validated primers for TFAP2B expression (Figure S5).

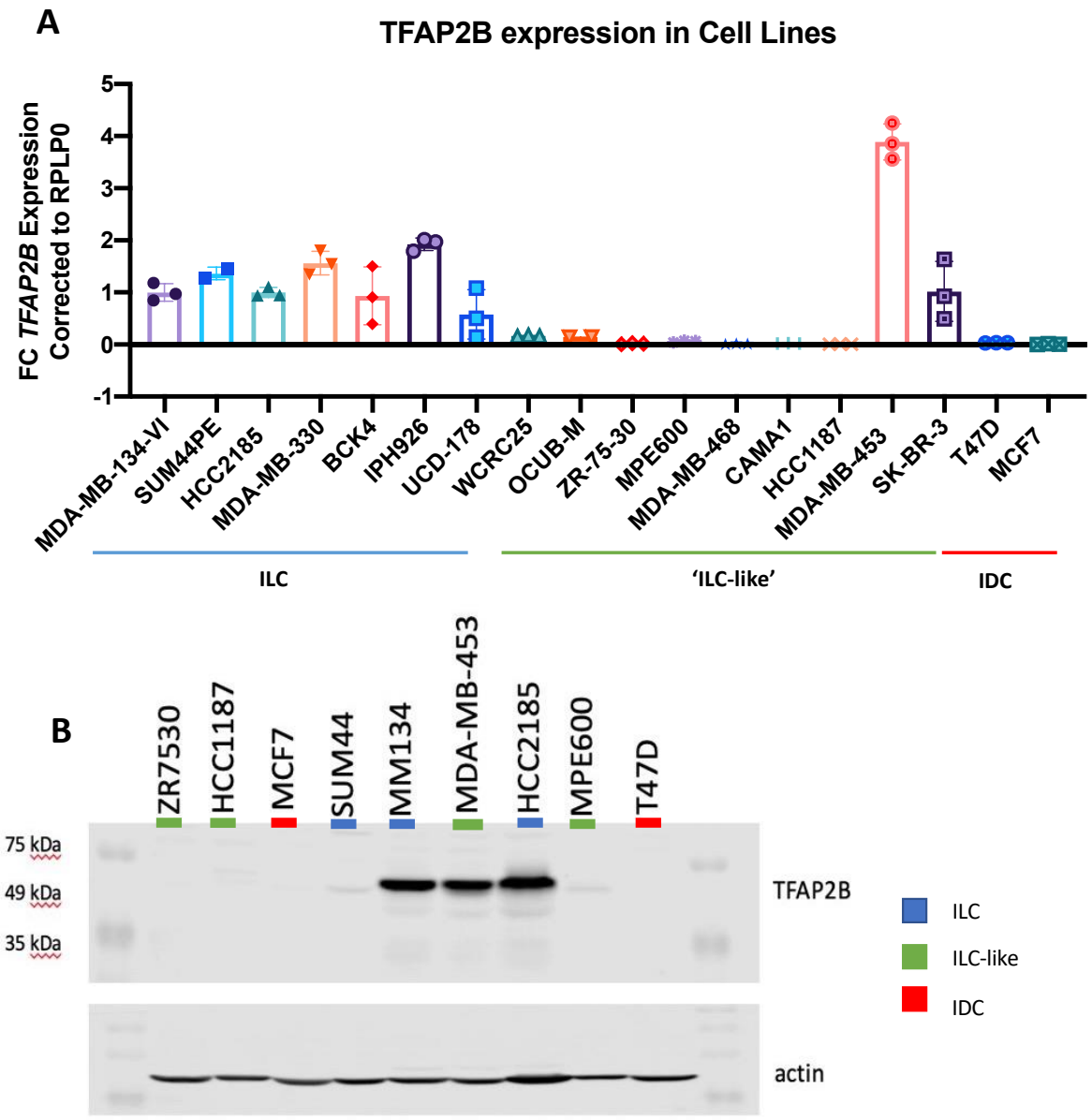


Figure 6 TFAP2B expression in a panel of cell lines

Cell line panel utilized to confirm *TFAP2B* expression at the mRNA and protein levels to determine optimal cell models to use for phenotypic experiments. (A) Quantitative real-time PCR in human breast cancer cell lines (B) Total cellular protein (30ug) was separated by 10% SDS-PAGE and probed with anti-TFAP2B antibody. Detection of actin verified equal loading.

3.5 TFAP2B loss causes diminished growth in both 2D and ULA

Due to the higher protein level determined from western blot and transcription level from RNAseq/qRT-PCR of TFAP2B in MM134, SUM44PE, HCC2185, and MM453 cells, we decided to transiently knock down *TFAP2B* to assess differences in growth and proliferation in 2D and ultra-low attachment (ULA) conditions. Optimization experiments with siRNA demonstrated consistent knockdown at all levels, but we ultimately settled on conducting our experiments with 50 pmol TFAP2B targeting siRNA (Figure S4). We pursued alteration of the 2D and ULA proliferation phenotypes in full serum for all cell lines. The expected outcome was that *TFAP2B* loss would lead to decreased proliferation in both conditions, as TFAP2B is known to effect proliferation and knockdown has shown similar results in other cancer types [45]. Knockdown of *TFAP2B* led to decreased 2D and ULA proliferation in all cell lines tested in both 2D and ULA conditions (Figure 7). There is a more robust response in MM134 and MM453 compared to SUM44, possibly due to the fact that those cell lines have higher TFAP2B expression. These results suggest that *TFAP2B* is important for cell growth and proliferation and aligns with published literature in multiple cancer types.

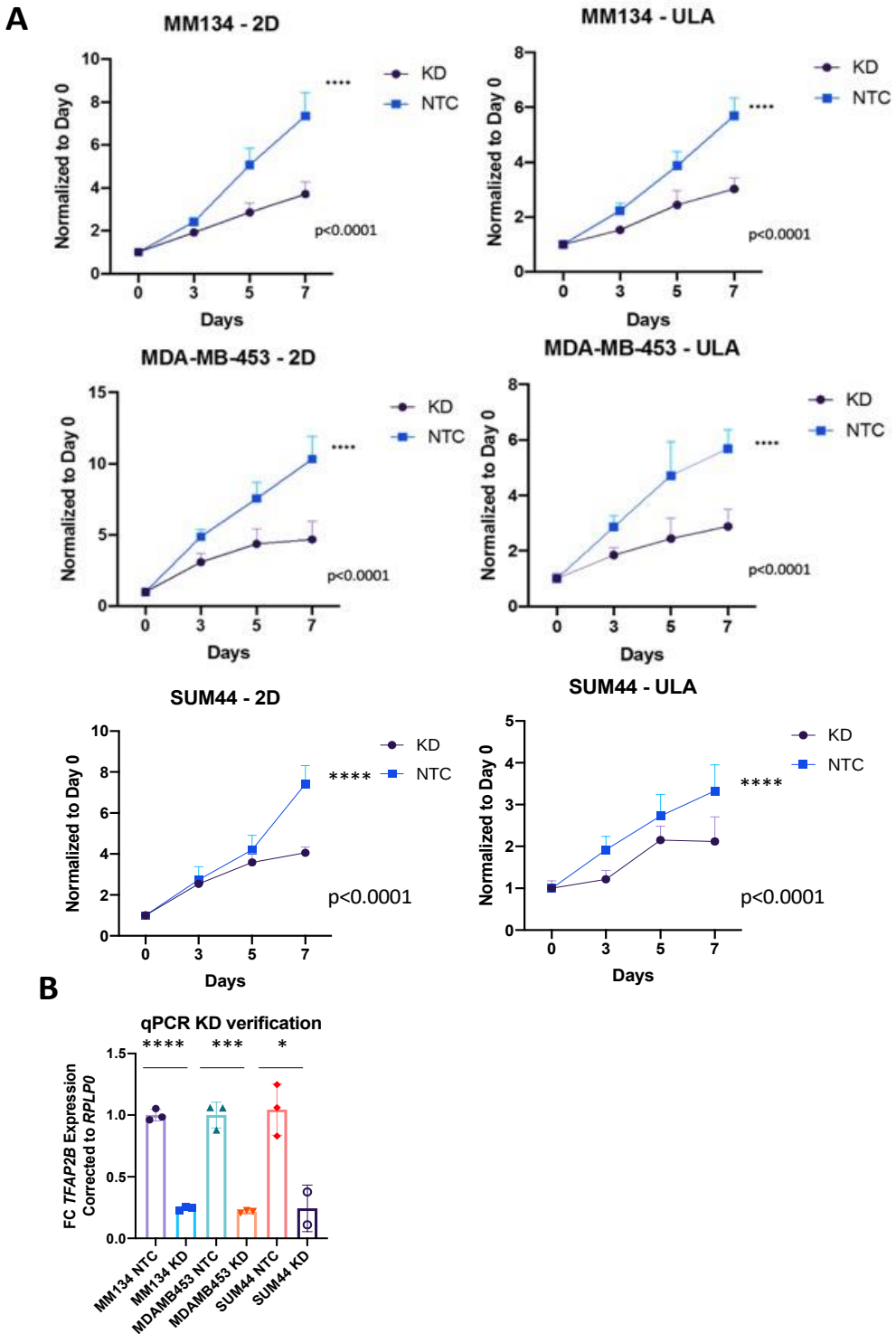


Figure 7 Loss of TFAP2B reduces cell proliferation in MM134, MM453, and SUM44PE cell lines

Cells were exposed to transient TFAP2B knockdown for 48 hours before being used to measure proliferation phenotypes. (A) 2D proliferation was measured with six technical replicates per treatment group and data was quantified following normalization to D0 values with non-linear regressions applied to test differences between slopes (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) (n=2) for each cell line (B) qRT-PCR for TFAP2B was performed after 48 hours with technical triplicates and students t-test for each siNTC versus siTFAP2B comparisons shown to verify the success of TFAP2B knockdown with siRNA (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). Data normalized to RPLP0, and then to NTC. NTC = non-targeting control, KD = TFAP2B knockdown, FC = fold change.

3.6 TFAP2B loss causes cell arrest in G_0/G_1

Decreased expression of *TFAP2B* has been previously described to contribute to decreased proliferative phenotypes due to cell cycle arrest. We first measured cell cycle profiles in MM134, MM453, HCC2185, and SUM44PE +/- transient TFAP2B knockdown (72 hours). We observed cell arrest in G_0/G_1 with decreases in S phase and no changes in the G_2/M phase (Figure 4). Additionally, gene set enrichment analysis was performed on TFAP2B MOTIF analysis (Figure 15 and 8C), which showed a downregulation of the REACTOME cyclin E associated events during G1 to S transition. The G1 to S transition is controlled by the Cyclin R: Cdk2 complexes. The transition from G1 to S phase during the cell cycle is important for cell proliferation and misregulation of this process promotes oncogenesis. During the G1 phase, cyclin-dependent kinase (CDK) activity initiates the G1 to S phase transition and promotes DNA replication. The G1-S transcripts then encode proteins that regulate downstream cell cycle events [56]. Since this pathway is downregulated, cells in the G1 phase are reduced moving into the S phase. This is what is illustrated by Figure 8C.

7

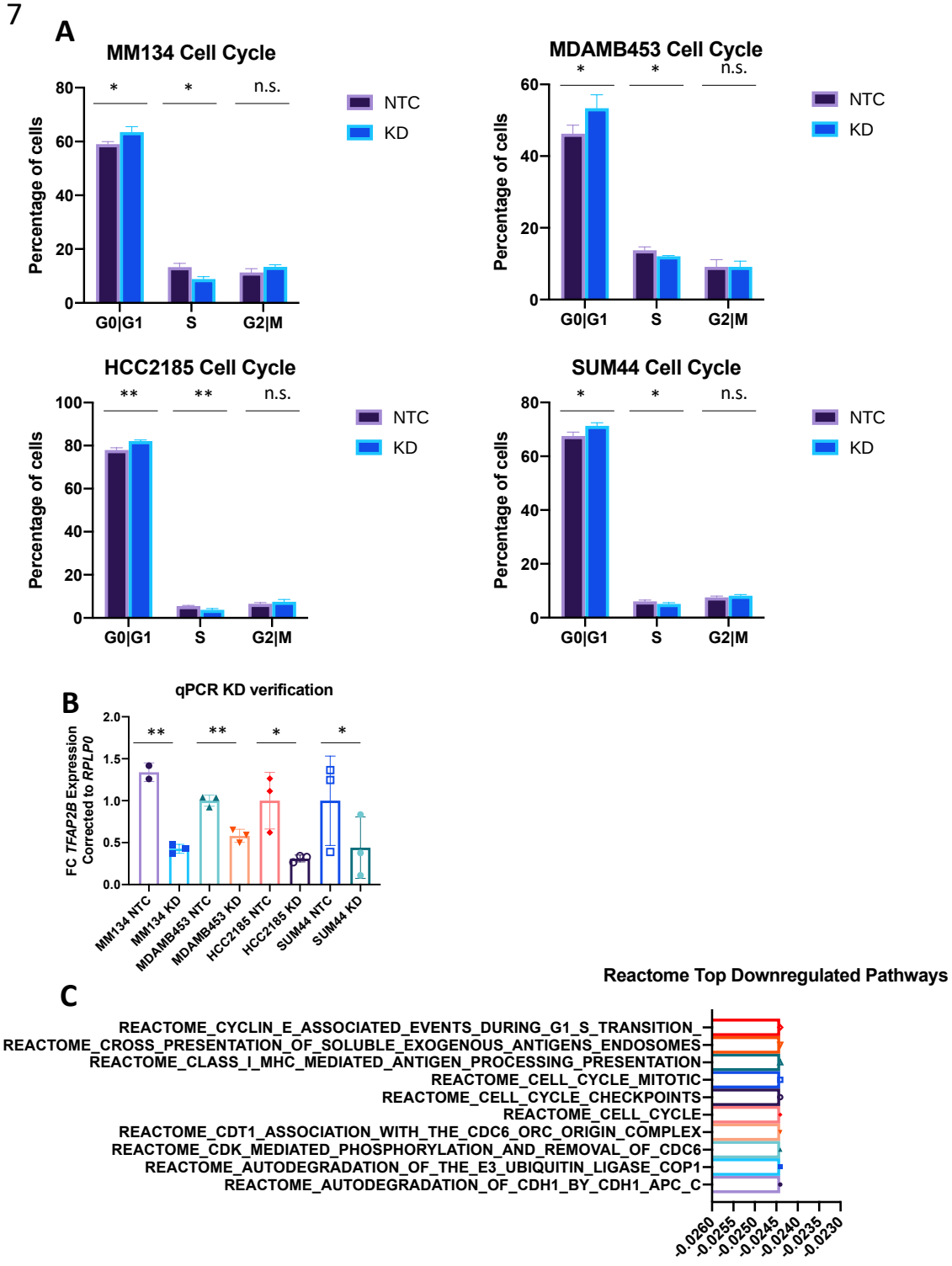


Figure 8 Transient Knockdown of TFAP2B causes arrest in G0/G1

MM134, MDA-MB-453, HCC2185, and SUM44PE were plated in triplicate for cell cycle analyses. Data are representative of two independently performed experiments with similar outcomes. Students t-test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) was used to compare groups for each given phase of the cell cycle. Error bars represent s.e.m. (B) qRT-PCR for TFAP2B was performed after 48 hours with technical triplicates and students t-test for each siNTC versus siTFAP2B comparisons shown Results normalized to NTC. (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). KD= TFAP2B knockdown and NTC = non-targeting control (C) GSEA was performed on MOTIF analysis of TFAP2B (Figure 13) and shows that the REACTOME cyclin E associated events during G1-S transition.

3.7 TFAP2B loss regulates cell proliferation exhibited through colony formation

Transient *TFAP2B* knockdown in breast cancer cell lines was used to further investigate the function of *TFAP2B* in cell growth and proliferation using colony formation assays. Cells were plated in 6-well plates with serial dilution (from 40,000 cells to 1250 cells). The cells were knocked down with siTFAP2B initially and then again after sufficient colonies were formed. After that, media was changed every four to five days before the cells were fixed and stained with crystal violet. We observed *TFAP2B* knockdown significantly reduced colony formation in MDA-MB-453 cells as shown by both colony number (Figure 9A) and percentage surface area (Figure 9B).

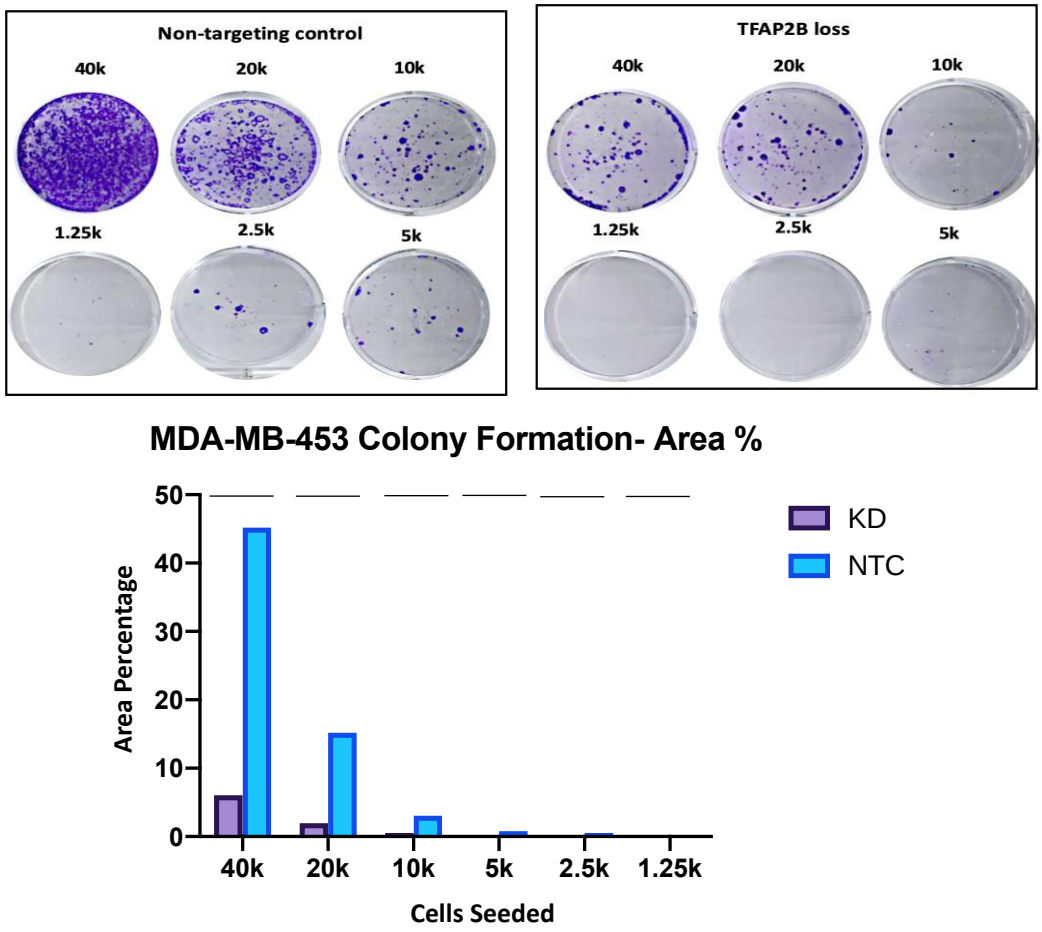


Figure 9 TFAP2B knockdown reduced colony formation ability in MM-453 cells.

MM-453 cells were transfected with siTFAP2B and assessed for colony formation ability after 5 weeks. (A) cells were stained with crystal violet and imaged (n=1) (B) The colony formation rate and colony area were calculated using ImageJ. Data is presented as the percentage area of colonies per well. Student's t-test (**p<0.0001, *p<0.005) significant differences between treatment group and control group. NTC = siScramble non-targeting control, KD = siTFAP2B knockdown.

3.8 TFAP2B loss impacts apoptosis in BC cells

Apoptosis is a highly conserved cell death mechanism. It occurs during normal development and turnover. Improper apoptosis contributes to diseases such as cancer, viral infection and autoimmune diseases, and can be triggered by a wide array of signals. The cell cycle is also a conserved mechanism of cell replication. Maintenance of a homeostatic balance of cell loss and gain, is achieved through the coupling of proliferation and programmed cell death. In other cancer models, loss of *TFAP2B* expression leads to cell proliferation and induces apoptosis during tissue development in mice [44]. To determine what could be causing the cell arrest as seen in Figure 8, we decided to investigate apoptosis in our models. We first measured apoptosis in MM134, MM453, HCC2185 cells with transient TFAP2B knockdown, at 72 hours. Although the results were insignificant in MM134 and MM453, HCC2185 non-targeting control had a higher percentage of apoptosis compared to knockdown, suggesting that *TFAP2B* may play a role in reducing apoptosis. In Figure 10A, one of three technical triplicates are represented for si*TFAP2B* and si*Scramble* with images taken from FlowJo. In Figure 10B, apoptosis percentage is the combination of Q1 and Q2, the late apoptotic regions and are quantified on the plot. In Figure 10C, qPCR verification was conducted on the cell lines, where *TFAP2B* expression is normalized to *RPLP0* and KD of each cell line is normalized to its own NTC. Figure 10D is a GSEA from the MOTIF analysis that shows that Hallmark apoptosis is a top upregulated pathway.

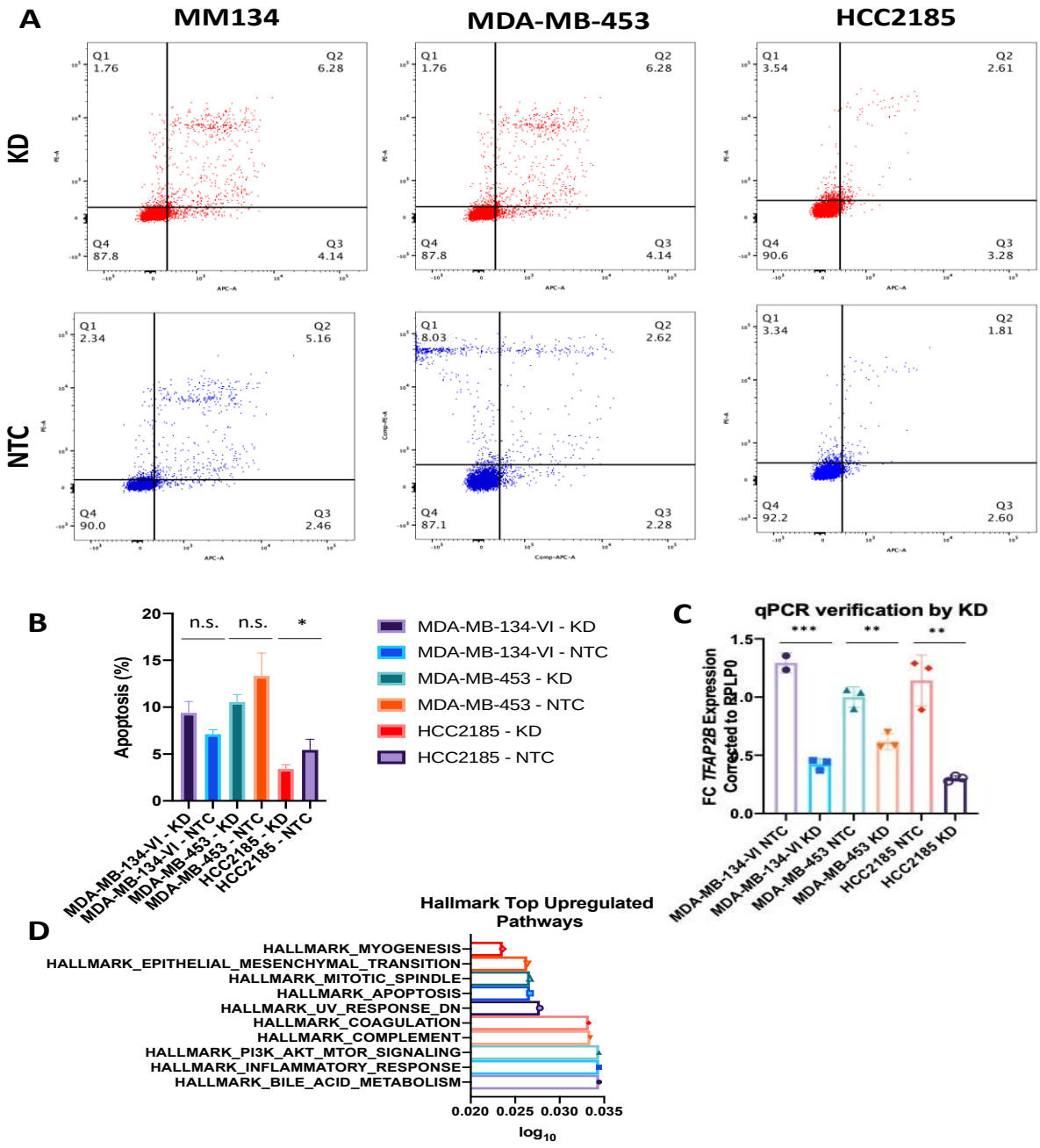


Figure 10 TFAP2B loss could impact apoptosis

(A) MDA-MB-134-VI, MDA-MB-453 and HCC2185 were plated in triplicate for Annexin/PI apoptosis analyses. X-axis APC-A, and Y-axis PE-A Data are representative of two independently performed experiments with similar outcomes. Students t-test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) was used to compare groups for apoptosis percentage between KD and NTC. Error bars represent s.e.m. (n=2) experiments for MDA-MB-453 and HCC2185, MDA-MB-134-VI (n=1) NTC = si*Scramble* non-targeting control, KD = si*TFAP2B* knockdown (B) Data

quantified as % apoptotic cells. (C) qPCR verification of knockdown (D) Pathway analysis on Hallmark dataset shows upregulation of Apoptosis.

3.9 TFAP2B may affect estrogen receptor signaling

As suggested by Figure 1C, *TFAP2B* expression is more prevalent in ER+ LumA tumors, suggesting that *TFAP2B* may be related to estrogen receptor signaling. To examine the effect of *TFAP2B* loss on estrogen receptor signaling, MM134 cells were hormone-deprived for 3 days and then transfected with siTFAP2B for 48 hours and exposed to vehicle, 1 nM E2, , 1μM ICI 182,780 (Fulvestrant, ICI) or a combination of E2 and ICI. Cell growth was monitored over 5 days and the relative growth of cells was compared day 5 to day 0, as shown in Figure 11A. As expected, cells treated with E2 had significantly higher growth compared to untreated samples and cells treated with ICI or ICI plus E2 showed no growth. As shown in Figure 11B, the siTFAP2B cells did show less of an E2 growth response under the hormone deprived conditions. This suggests that *TFAP2B* could be associated with estrogen signaling in ILC. In Figure 11C, qPCR verification was conducted on the cell lines, where *TFAP2B* expression is normalized to *RPLP0* and KD is normalized to NTC. This is corroborated by Figure 11D, where there is a slight positive correlation between *TFAP2B* and *ESR1* in ILC from TCGA primary tumor data, with an opposite trend in IDC.

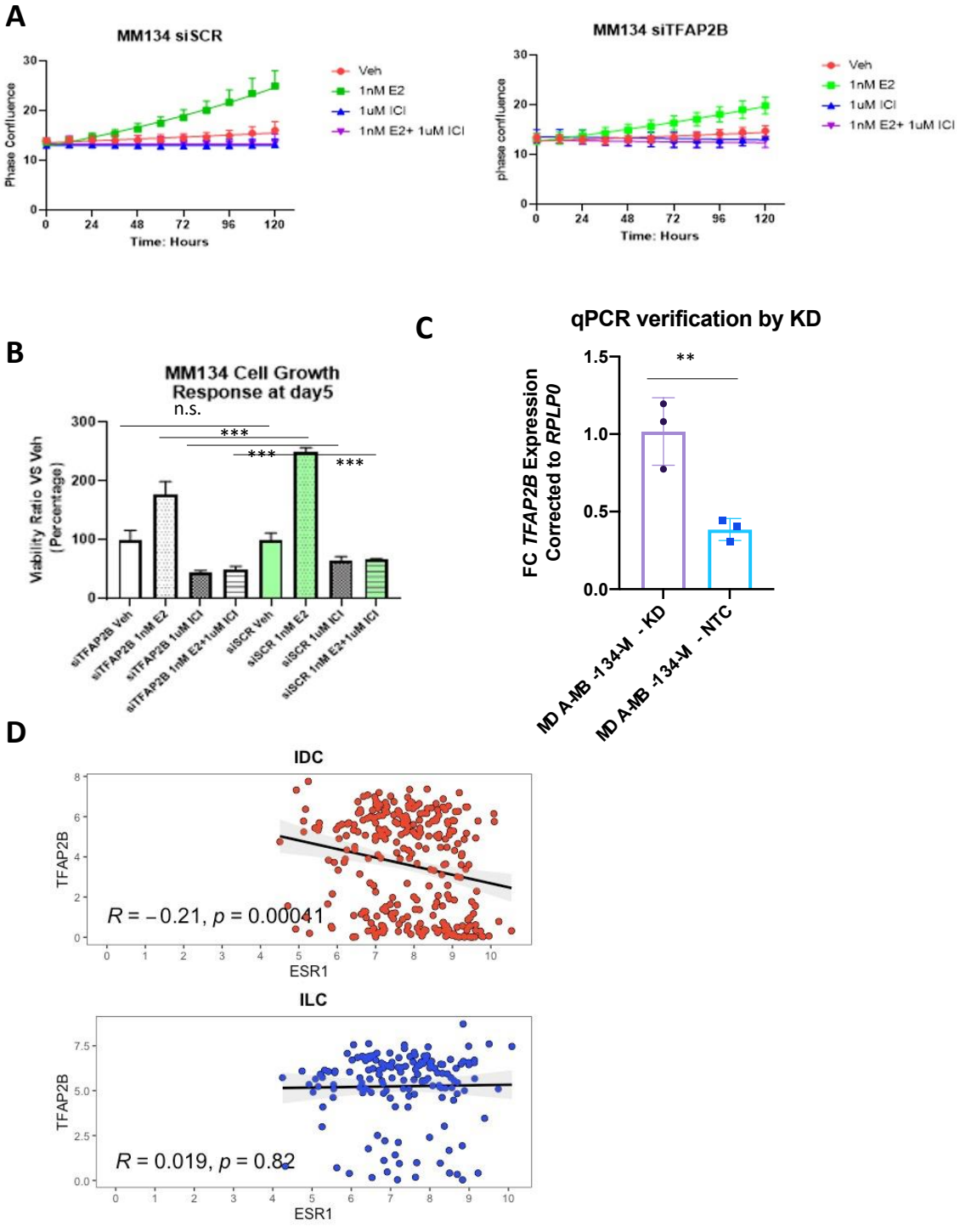


Figure 11 TFAP2B may play a role in estrogen receptor signaling

(A) Cells were maintained in hormone deprived media over 5 days and confluency measured using the Incucyte. (B) Bar graph shows the average of relative growth on day 5 with the error bar representing the SEM of six technical replicates, as measured by PrestoBlue viability assay. Students t-test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) was used to compare groups between KD and NTC. (C) qPCR verification of knockdown (D) TFAP2B, ESR1 correlation in TCGA ER+ Luminal A primary tumors. Estrogen response assay performed by Jian Chen, MS.

3.10 TFAP2B has a negative association with CDH1

CDH1 loss is the hallmark ILC. The E-cadherin protein encoded by the CDH1 gene plays a critical role in cell-cell adhesion and cell-cell junctions and communication. *CDH1* has significantly lower expression in ILC compared to IDC, so we assessed the relationship between *TFAP2B* and *CDH1*. To do this, in Figure 12A the correlation between the two genes in primary breast tumors from TCGA were plotted. As anticipated IDC tumors showed a negative correlation (higher CDH1, lower TFAP2B), and ILC showed a slightly positive, but non-significant association. In Figure 12B, RNAseq of *CDH1* knockout and overexpression cell models was analyzed. The observation was that with lower *CDH1* expression, the models exhibited higher *TFAP2B* expression, and the reverse is true with higher *CDH1* expression. We expected that *CDH1* to might play a role in controlling TFAP2B expression in ILC and while this data supports that notion, further work is needed eludicate the mechanisms by which this occurs. As a control, *CDH1* expression levels were confirmed in the *CDH1* knockout and overexpression cell models (Figure S6).

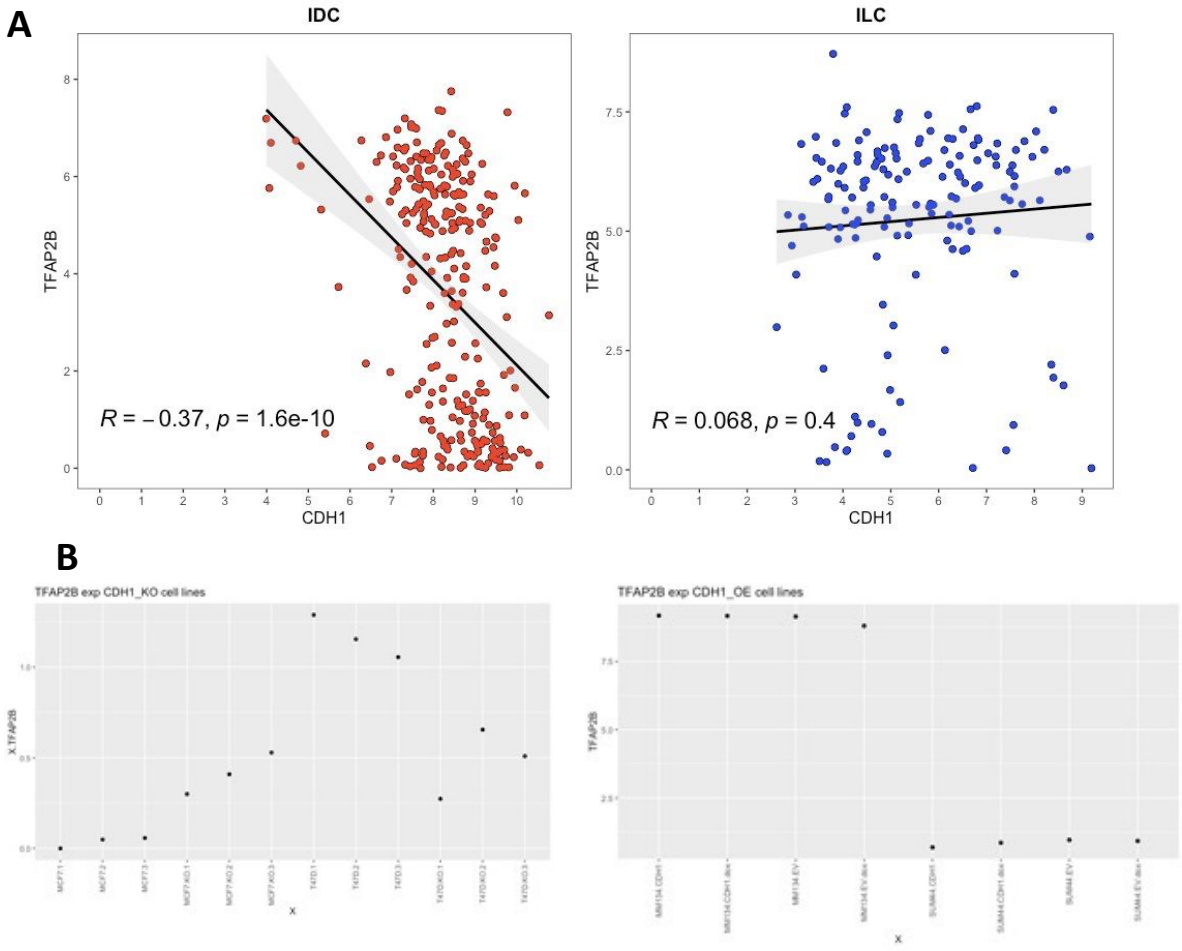


Figure 12 TFAP2B has a negative correlation with CDH1 expression in CDH1 KO and OE cell lines

(A) TFAP2B and CDH1 correlation in ER+ LumA primary tumors in IDC and ILC. (B) TFAP2B expression in CDH1 KO and OE cell lines. X axis is cell lines, and y axis is TFAP2B expression. Cell lines are in triplicate for KO cell lines and single for OE cell line

3.11 Differentially expressed pathways in TFAP2B high expressing ILC and IDC tumors

In order to determine pathways which may be altered in TFAP2B expressing tumors, we again turned to TCGA data. In Figure 13 the distribution of ILC and IDC primary tumors from

TCGA is plotted. It shows that IDC has some high *TFAP2B* expressing tumors, although the majority show little to no expression. To explore what these high expressing IDC tumors and ILC tumors have in common, differential analysis and GSEA pathway analysis was conducted. The cutoff point for calling high *TFAP2B* expressing tumors was the median. From this, our analysis revealed 56 differentially expressed genes between ILC and IDC in high *TFAP2B* expressing tumors with FDR-adjusted p value <0.05 and fold change ≥ 2 (Data Supplement: Table S1). In Figure 13A, the bimodal distribution in ILC and IDC primary tumors TFAP2B expression is evident. Clustering this gene list into differentially expressed genes between ILC and IDC is illustrated in Figure 13B. The most notable upregulated pathways included estrogen response early and late and androgen response (Figure 13C). A majority of ILC tumors (80-95%) and 70-80% of IDC tumors express estrogen receptor [20,21]. Since the estrogen response is upregulated in Figure 13, it makes sense that *TFAP2B* could be related to estrogen receptor signaling. Additional pathway analysis is illustrated by Figure S10.

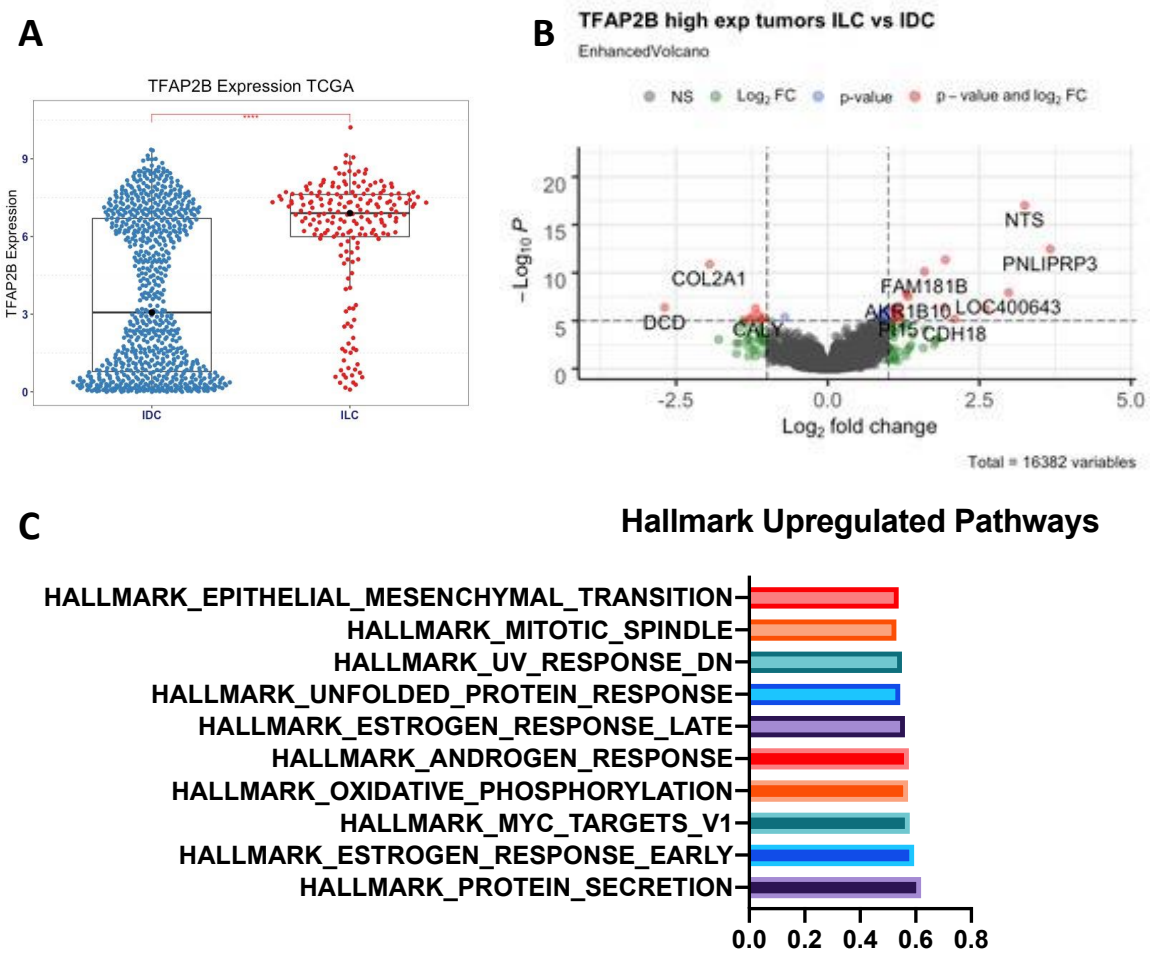


Figure 13 Differentially expressed pathways in high TFAP2B expressing ER+ LumA ILC and IDC tumors reveals potential mechanism

(A) TFAP2B expression in TCGA ER+ LumA primary tumors. (B) Volcano plot showing significantly differentiated genes in IDC and ILC. (C) Upregulated HALLMARK pathway. P-value is indicated for each pathway.

3.12 TFAP2B expression explored through MOTIF analysis potential mechanistic pathways

To determine induced or downregulated pathways in ILC and IDC tumors, I took the top 100 downstream genes predicted by Integrated Motif Activity Response Analysis (ISMARA) as a TFAP2B signature. ISMARA uses a methodology that models gene expression or chromatin modifications as genome-wide predictions of regulatory sites. This tool identifies key transcription factors and microRNAs that drive expression and chromatin changes and makes predictions regarding their regulatory roles. The regulatory roles include genome-wide targets, and enriched gene categories, as well as direct interactions between regulators. ISMARA uses motif activity response analysis to identify key regulators driving gene expression and chromatin state changes across samples, the activity of these regulators, target genes and sites on the genome where the regulators act. ISMARA works by taking raw gene expression or chromatin data, and computationally predicting regulatory sites [51]. For the TFAP2B Motif prediction, the data came from GNF SymAtlas and NCI-60 cancer cell lines [52,53]. In Figure 14A, the methodology for this exploratory technique is explained and processed using the TCGA LumA ER+ cohort. In Figure 14B, the ISMARA Motif, collaborated by HOCOMOCO Motif were generated for TFAP2B. HOCOMOCO is a database of transcription factor binding models that is used for sequence analysis of transcription regulation in models. It is useful for predicting transcription factor binding sites. It uses ChIP-seq datasets from over 5000 experiments and is regularly updated [55]. The most notable upregulated pathways identified through these analysis were apoptosis and PI3K AKT MTOR signaling pathways (Figure 14C, S11C) and downregulated pathways were estrogen response late, MAPK signaling pathway, and cyclin E associated events during G1 S phase transition (Figure 14D, S11A, S11D).

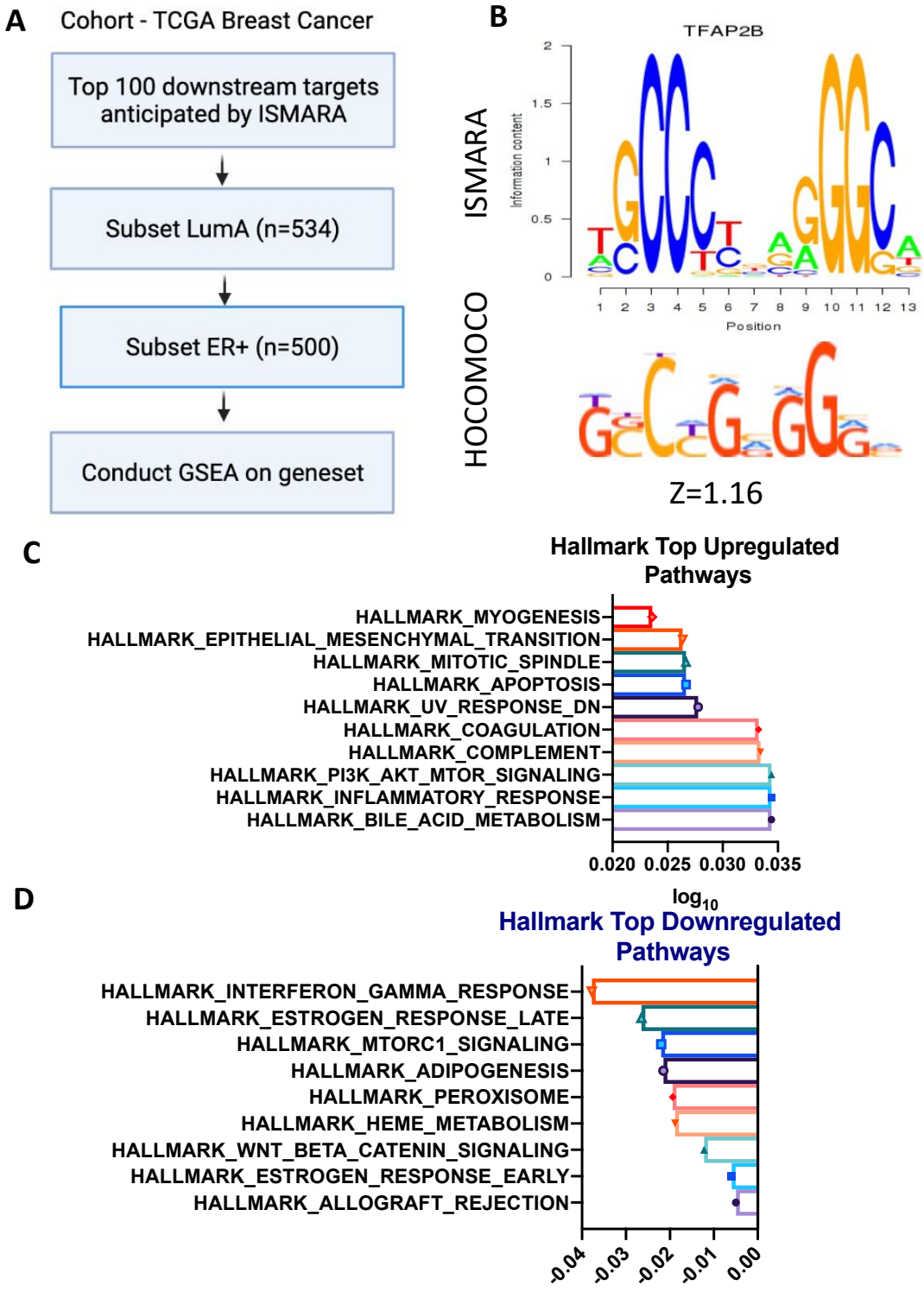


Figure 14 TFAP2B expression explored through MOTIF analysis reveals insights for potential downstream signaling

(A) Methodology for investigating MOTIF analysis generated by ISMARA (B) MOTIF generated by ISMARA and corroborated by HOCOMOCO. (C-D) Upregulated and downregulated HALLMARK pathways. P-value is indicated for each pathway.

4.0 Discussion and Conclusions

ILC is a unique disease with distinct pathological, clinical and molecular features compared to IDC. Although it represents 10-15% of invasive breast cancers, little research has focused on this breast cancer subtype until recent years and it is treated the same as IDC clinically despite unique properties. Although nearly 40,000 women will be diagnosed ILC in 2021, it is highly understudied [1]. ILC is slow growing, difficult to detect, and more likely to recur about a decade after the primary diagnosis, compared to IDC. A more thorough understanding of ILC is needed in order to understand drivers of disease progression and identify therapeutic vulnerabilities.

Endocrine therapy is a typical treatment for ER positive breast cancer, and since 80-95% of ILC patients express ER, primary treatment is endocrine therapy [20,21]. In ER positive breast cancer, ER is a successful therapeutic target and one of the first examples of precision medicine. Unfortunately, endocrine therapy resistance develops and leads to endocrine resistant metastatic disease. In addition to limitations in the treatment of ILC, there are limited prognostic markers to differentiate between different breast cancer histologies.

TFAP2B is important for embryonic organ development, and it is part of various gene lists used for breast cancer subtyping and prognostics [20-25]. Despite this, there are limited studies that explore the role of *TFAP2B* in breast cancer and due to the unique upregulation of TFAP2B in ILC compared to IDC, we aimed to characterize *TFAP2B* expression and function in invasive breast cancer.

Primary breast tumors show significantly enhanced *TFAP2B* expression in ILC compared to IDC, consistent with the higher levels of TFAP2B methylation in IDC tumors (Figure 2,3).

This data was accompanied by observations of an upregulation of estrogen early and late response genes in high *TFAP2B* expressing primary tumors, which suggest *TFAP2B* may be related to ER signaling, a finding which should be further explored and with potential therapeutic implications (Figure 11). *TFAP2B* positivity was found to be enhanced in lobular carcinoma *in situ*, as well as in lobular carcinoma, and showed associated with loss of E-cadherin, positive ER status, low Ki67, low recurrence scores and prolonged event-free survival [45]. Additionally, due to the difference in expression illustrated in Figure 3, *TFAP2B* could be a differentiation marker in ILC.

To study *TFAP2B* on a functional level, we utilized human breast cancer cell lines, and noted that TFAP2B is more highly expressed in ER+ derived carcinomas (Figure 4). As verification that DNA methylation is the driver of the epigenetic alterations effecting the expression levels in ILC, we used DNA demethylation with 5-aza-2-deoxycytidine to look for expression recovery. Indeed, SUM44PE and MCF7, the most highly methylated cell lines, re-expressed TFAP2B when treated with DAC. SiRNAs focused on diminishing TFAP2B levels in cell lines demonstrated that TFAP2B contributes to a proliferation phenotype *in vitro* in both two-dimensional and ultra low adhesion conditions (Figure 6). This is consistent with the published literature where neuroblastoma resulted in induction of TFAP2B expression, suggesting that TFAP2B is silenced through genomic methylation [59]. They also saw that low TFAP2B expression was associated with CpG methylation of the TFAP2B locus in neuroblastomas. TFAP2B knockdown in lung adenocarcinoma showed inhibition of proliferation, where TFAP2B overexpression significantly increased cell viability [42]. This effect was also seen in breast cancer and lung carcinoma [45, 60, 61]. Pathway enrichment also showed a deregulation of the cell cycle and MAPK signaling pathway, verifying that TFAP2B

plays a role in the cell cycle. To understand the changes observed in cell growth, flow cytometry to assess the cell cycle revealed an increase in the G0/G1 phase and decrease in the S phase, through cell arrest (Figure 7), again consistent with literature in breast cancer [45]. GSEA pathway enrichment from our MOTIF analysis showed a downregulation of REACTOME cyclin E associated events during the G1 S phase, suggesting that with TFAP2B knockdown there was a limiting factor that stopped the transition from the G1 to the S phase (Figure 7C) and suggesting that TFAP2B may have a role in the progression from G1 to S phase.

Based on other studies [42, 61] we sought to understand if apoptosis may also play a role in the cell growth phenotype, and this should be further explored (Figure 9) especially since pathway enrichment showed hallmark apoptosis upregulated in TFAP2B knockdown cells (Figure 9D). In lung adenocarcinoma, an initially flow experiment showed that TFAP2B knockdown played a role in apoptosis. This was then verified through caspase activation, examining the expression of pro-apoptotic and anti-apoptotic proteins, caspase-3/9, BAX and BCL-2 [61]. Additionally, a colony formation was conducted over a five week period in MM453, which showed reduced colony formation ability in TFAP2B knockdown cells (Figure 10). These functional observations indicate that TFAP2B controls cell proliferation and growth and could also modify the apoptotic response in ILC.

Our analysis of tumors highly expressing TFAP2B revealed estrogen early and late response genes were upregulated and prompted us to explore the relationship between TFAP2B and ER signaling. Using siRNA to deplete TFAP2B and assessing proliferative response to estradiol in stripped serum conditions we observed a weaker proliferative response to estradiol in TFAP2B depleted cells, indicating a role for TFAP2B in estrogen response which is an area warranting further investigation. As previously discussed, studies conducted on TFAP2A,

TFAP2B and TFAP2C role in estrogen regulation showed that AP-2A and AP-2C induced changes in the chromatin structure associated with ER∝ and there is a regulation loop between AP-2 protein and ER/estrogen where ER can extinguish the AP-2A transactivating properties. The AP-2 proteins take part in a mechanism related to ER regulation, but the molecular events have not yet been elucidated. Additionally, in clinical studies, nuclear AP-2A and AP-2B showed a positive association with ER in ER-positive breast cancer, where the *in vitro* studies show a lack of association between AP-2 and ER [62].

In order to understand the role it plays with estrogen signaling, siRNAs targeted at TFAP2B expression in MM134 showed a reduced estrogen response in the knockdown compared to the control; coupled with higher TFAP2B in ER+ LumA tumors and TCGA positive (non-significant) correlation in ILC, this suggests that TFAP2B plays a role in estrogen receptor signaling (Figure 10). However, a study in breast cancer, treated with ICI did not affect *TFAP2B* expression, so they determined that *TFAP2B* is independent of ER signaling [45]. They only blotted for TFAP2B and actin, but did not show ER. Therefore, *TFAP2B*'s interaction with estrogen signaling should be further explored.

To understand TFAP2B's role with E-cadherin, *TFAP2B* expression was examined in CDH1 knockout and overexpression cell lines. From this, the overexpression cell lines showed a minimal difference and the knockdown models showed opposite response. Interestingly in MCF7 cells, *CDH1* knockdown showed an increase in *TFAP2B* expression, suggesting that CDH1 could drive *TFAP2B.* In previous reports, AP-2 was shown to mediate the activation of Rb and Myc, which in epithelial cells have shown to transcriptionally activate the expression of E-cadherin [63]. Considering that we see an increase of *TFAP2B* expression with CDH1 loss, it could be driven by transcriptional changes upstream of CDH1.

Overall, the results presented here support an important role of *TFAP2B* in ILC. The growth, apoptosis and signaling findings contained herein warrant further detailed exploration in order to mechanistically elucidate the role of TFAP2B in breast cancer. It will be crucial to conduct further comprehensive functional studies of TFAP2B in ILC including gene regulation and signaling (ChIP-seq) in further models such as stable expressing cell lines or genome edited models followed by *in vivo* assessment of the contribution of TFAP2B in ILC. Intraductal xenografts of MM134 do show significant *TFAP2B* expression [58] (Figure S9), and particular organoids, including B046, B030 (ILC), and B075 (IDC), that would be optimal models for a more accurate clinical representation (Figure S3).

Finally, data so far indicates that TFAP2B could be a valuable biomarker in ILC or indeed represent a new therapeutic vulnerability for targeting ILC tumors, potentially representing a new avenue for precision therapeutics in ILC.

Appendix A Supplementary Figures

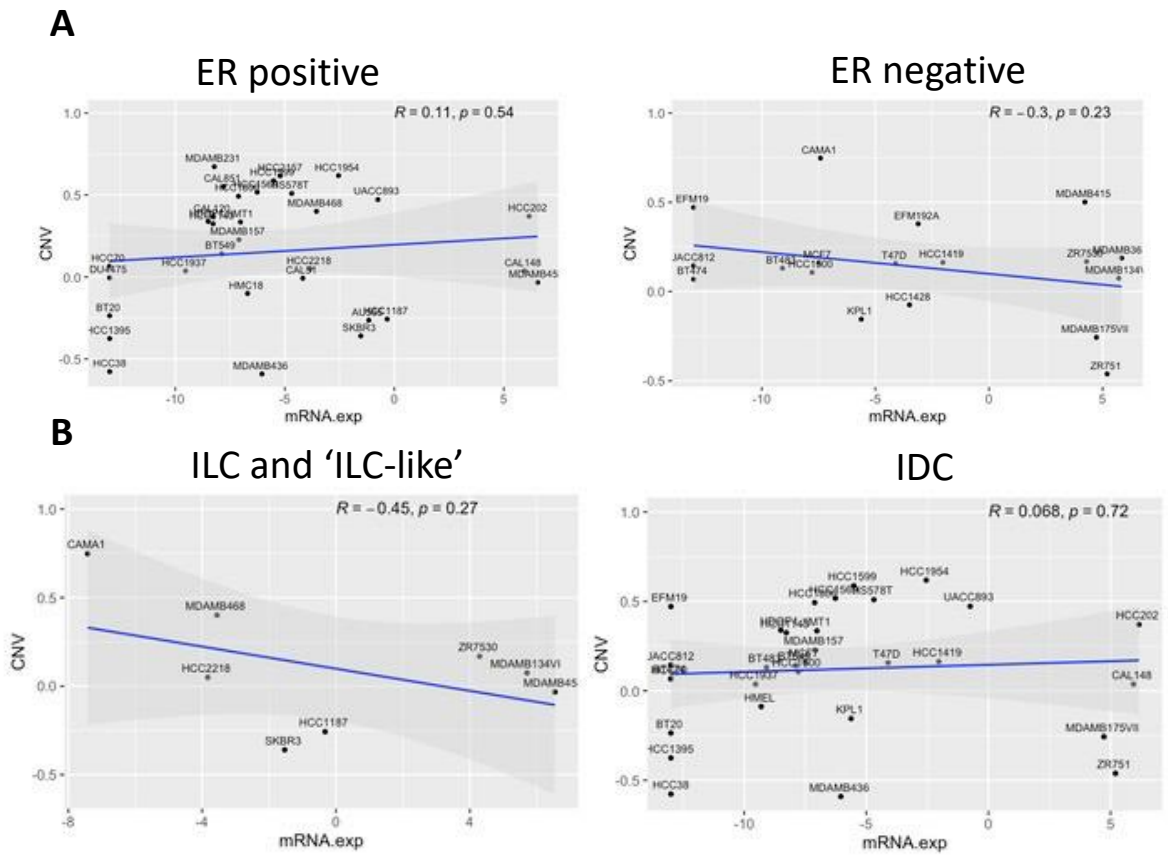


Figure S1 Copy Number variations have no significant impact on TFAP2B expression

Cell line data obtained from the Broad institute for ILC, 'ILC-like' and IDC cell lines. (A) Comparison of TFAP2B expression and copy number variations extrapolated between ER positive and ER negative cell lines. (B) Comparison of TFAP2B expression and copy number variations separated by histological subtypes

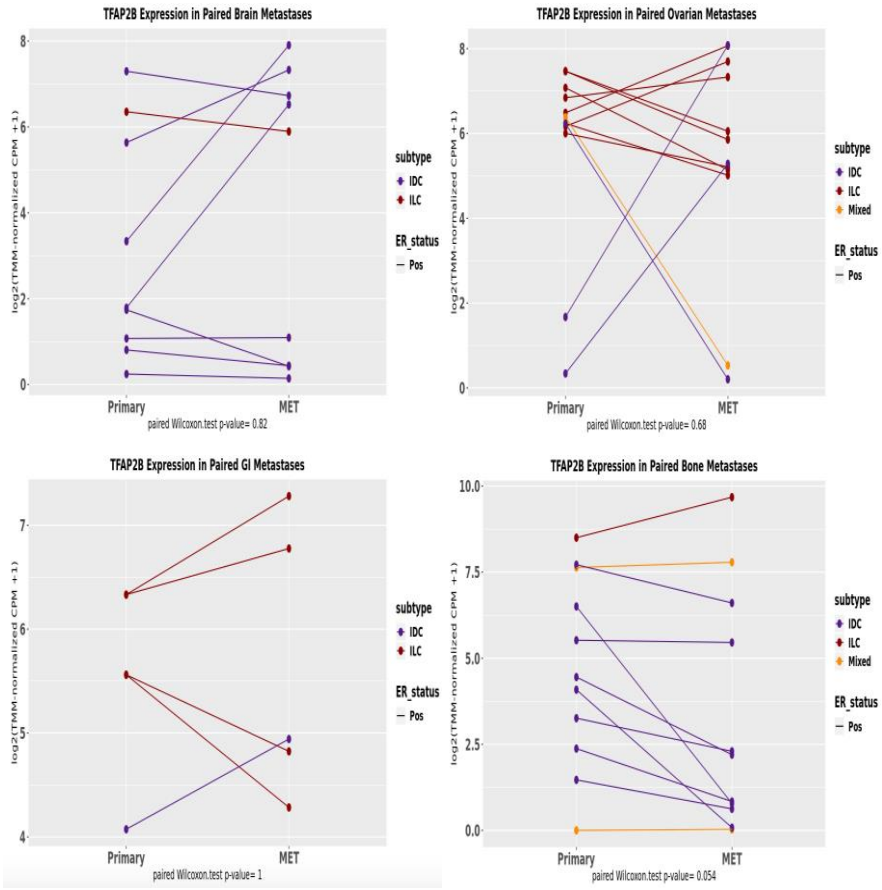


Figure S2 TFAP2B in paired metastasis in brain, ovarian, GI and bone shows consistently higher ILC expression in primary tumors

Mined data was generated from samples collected by Nolan Priedigkeit (NP), PhD, Ahmed Basudan (AB), PhD, and Rebecca Watters (RW), PhD for n=11 matched primary/metastatic ILC pairs from brain (NP), ovary (AB), GI tract (AB) and bone (RW) metastasis-afflicted patients.

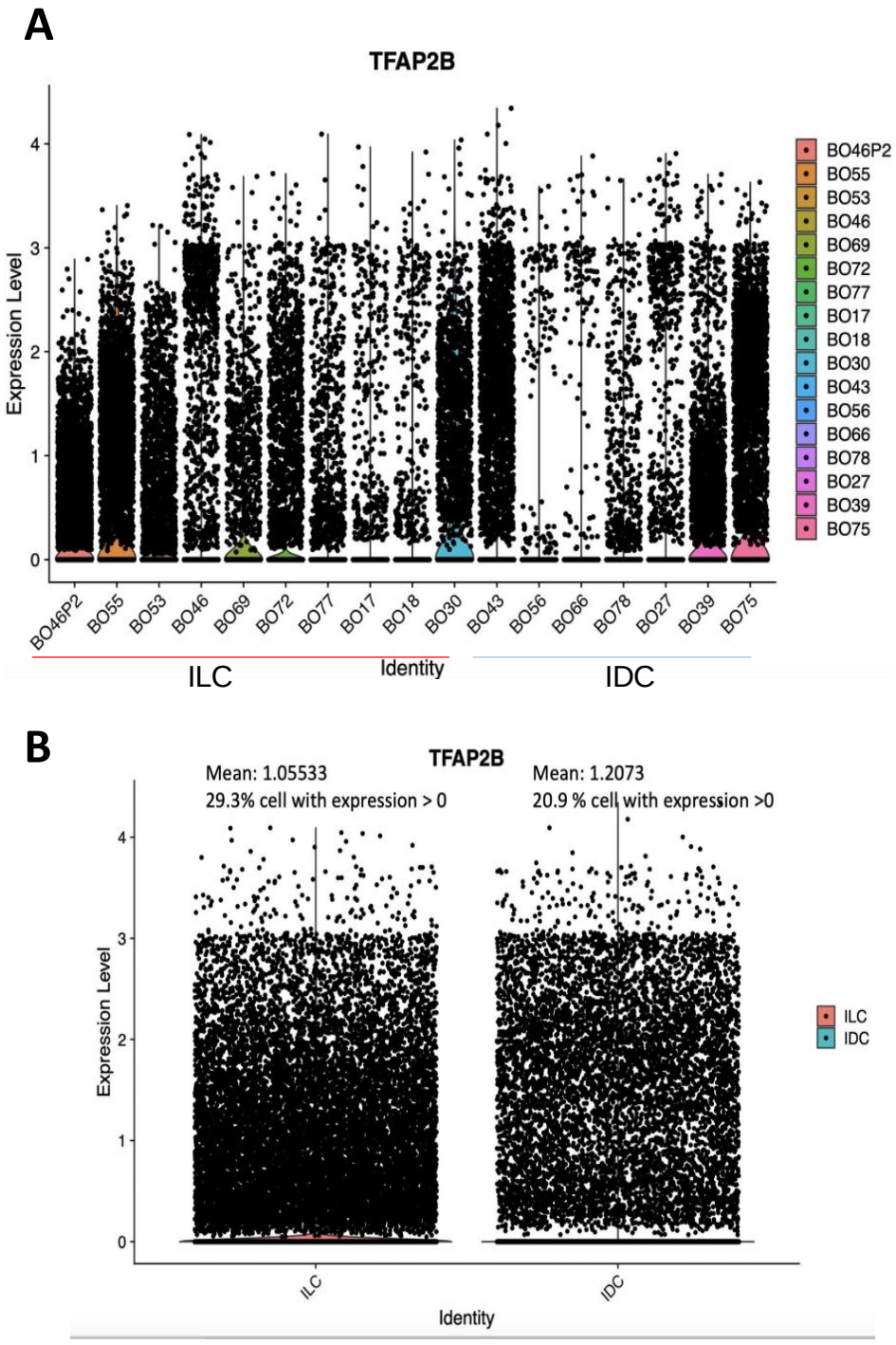


Figure S3 TFAP2B expression in bone metastasis, ovarian metastasis and primary tumors

TFAP2B expression in individual organoids, data obtained through single cell sequencing by Kai Ding, BS. BO46P2-BO30 are ILC organoids, BO43-BO75 are IDC organoids. (B) Overall TFAP2B across all organoids, extrapolated by histology. Bioinformatics completed in collaboration with Kai Ding, BS.

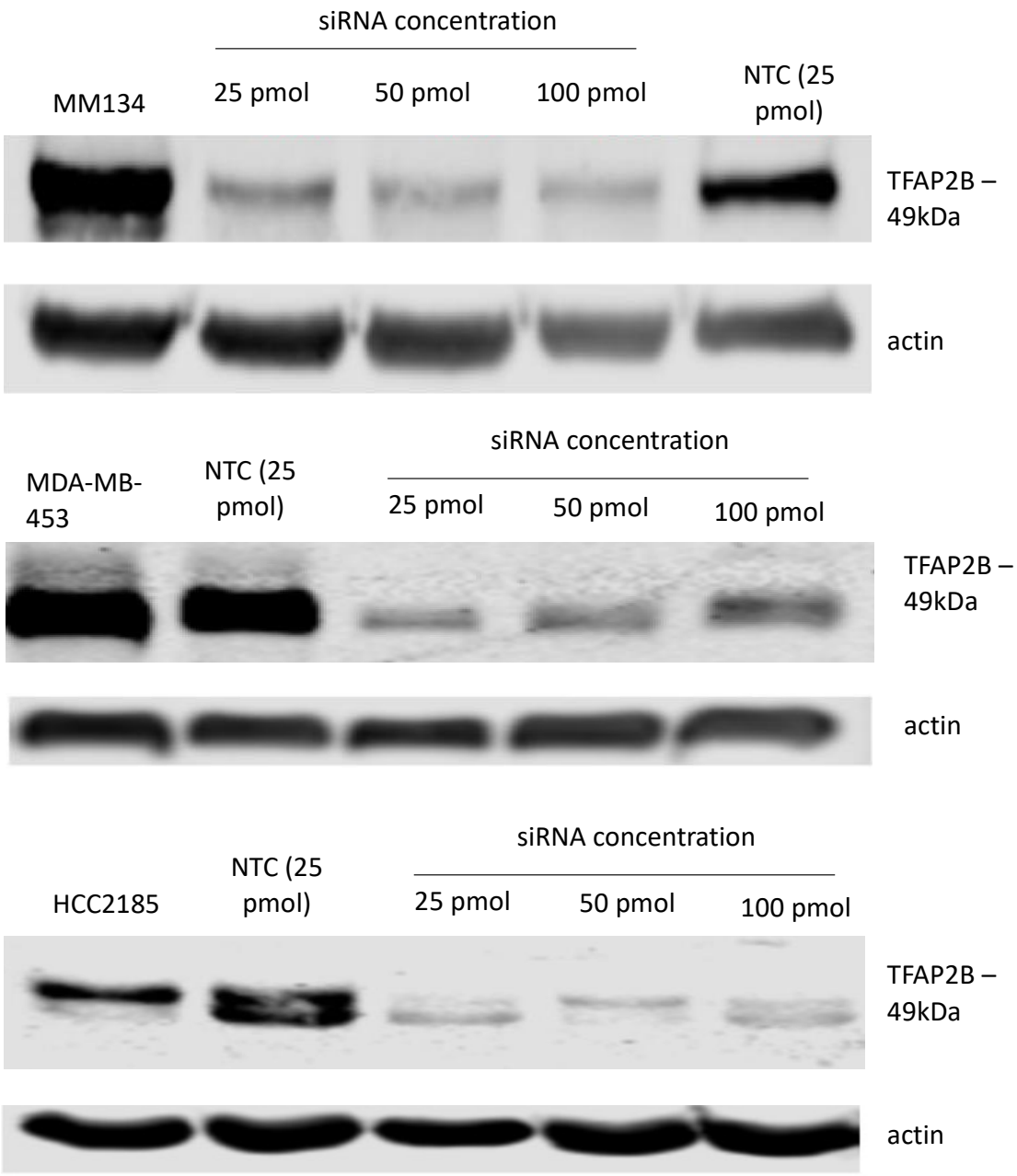


Figure S4 siRNA mediated knockdown optimization in cell lines

Knockdown of MDA-MB-134-VI, MDA-MB-453 and HCC2185 cell lines were treated with concentrations of siRNA (25 pmol, 50 pmol, 100 pmol) to find optimal KD siRNA concentration.

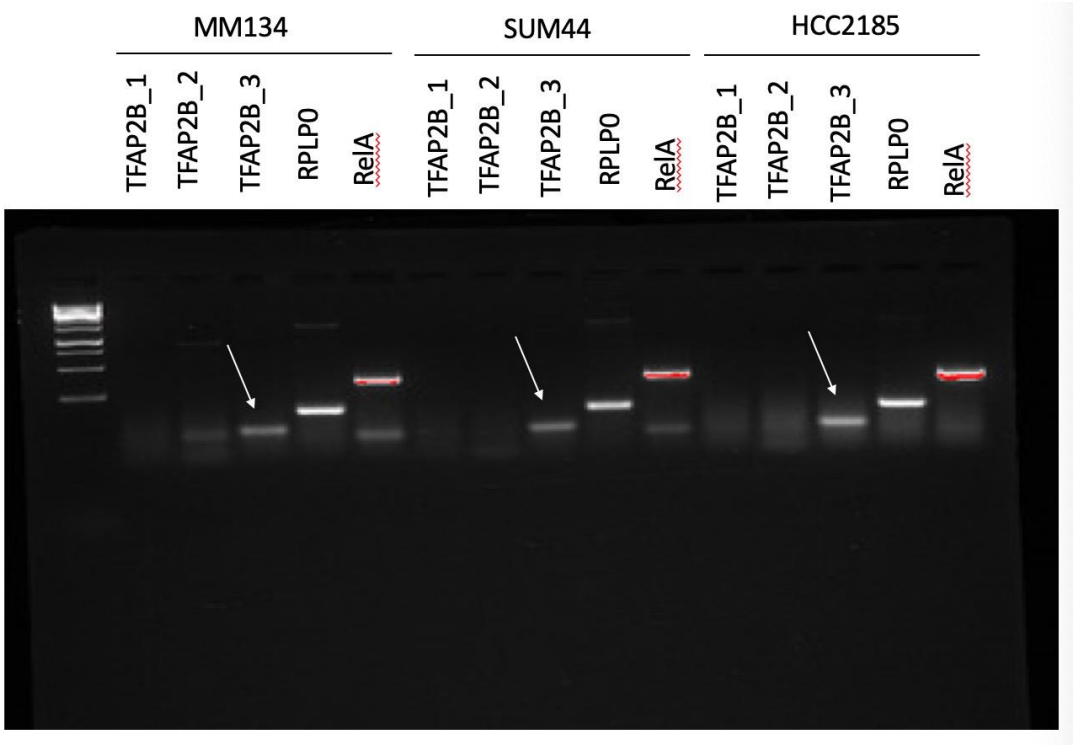


Figure S5 TFAP2B primer optimization for qRTPCR

TFAP2B taken from previous publications optimized settings for PCR. TFAP2B primer 3 was selected. RPLP0 and RelA primers are controls.

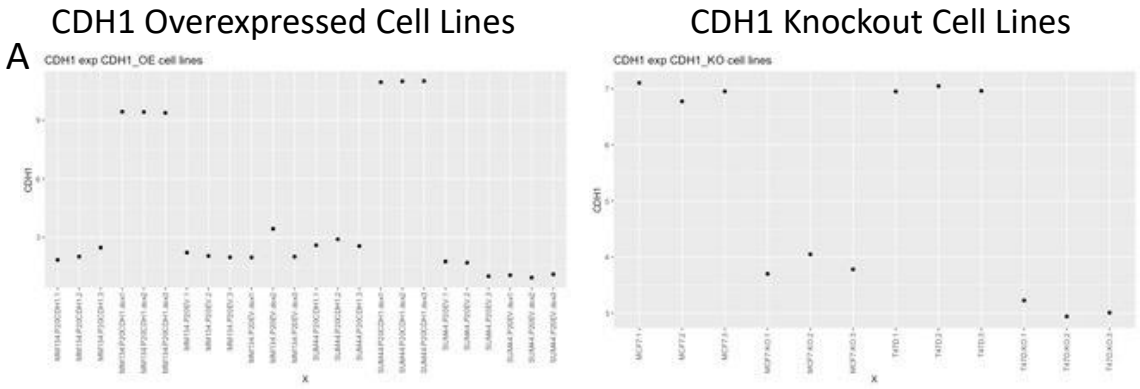
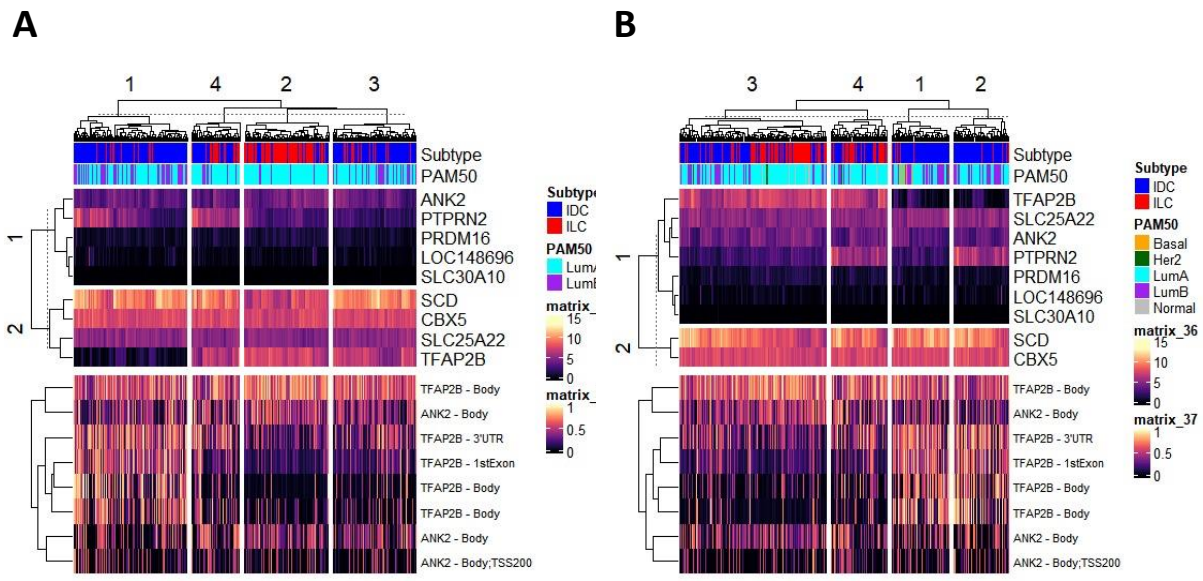


Figure S6 *CDH1* controls in *CDH1* overexpression and knockdown cells

CDH1 expression in CDH1 KO and OE cell lines



	LumA	LumB
IDC	131	83
ILC	94	5

	Basal	Her2	LumA	LumB	Normal
IDC	5	4	116	76	1
ILC	0	0	87	4	9

Figure S7 Unsupervised clustering of top differentially methylated probes of TFAP2B gene expression in ILC and IDC primary tumors

. (A) Including luminal A and B tumors (B) Comparing ILC and IDC and the top differentially methylated probes in ER+ tumors

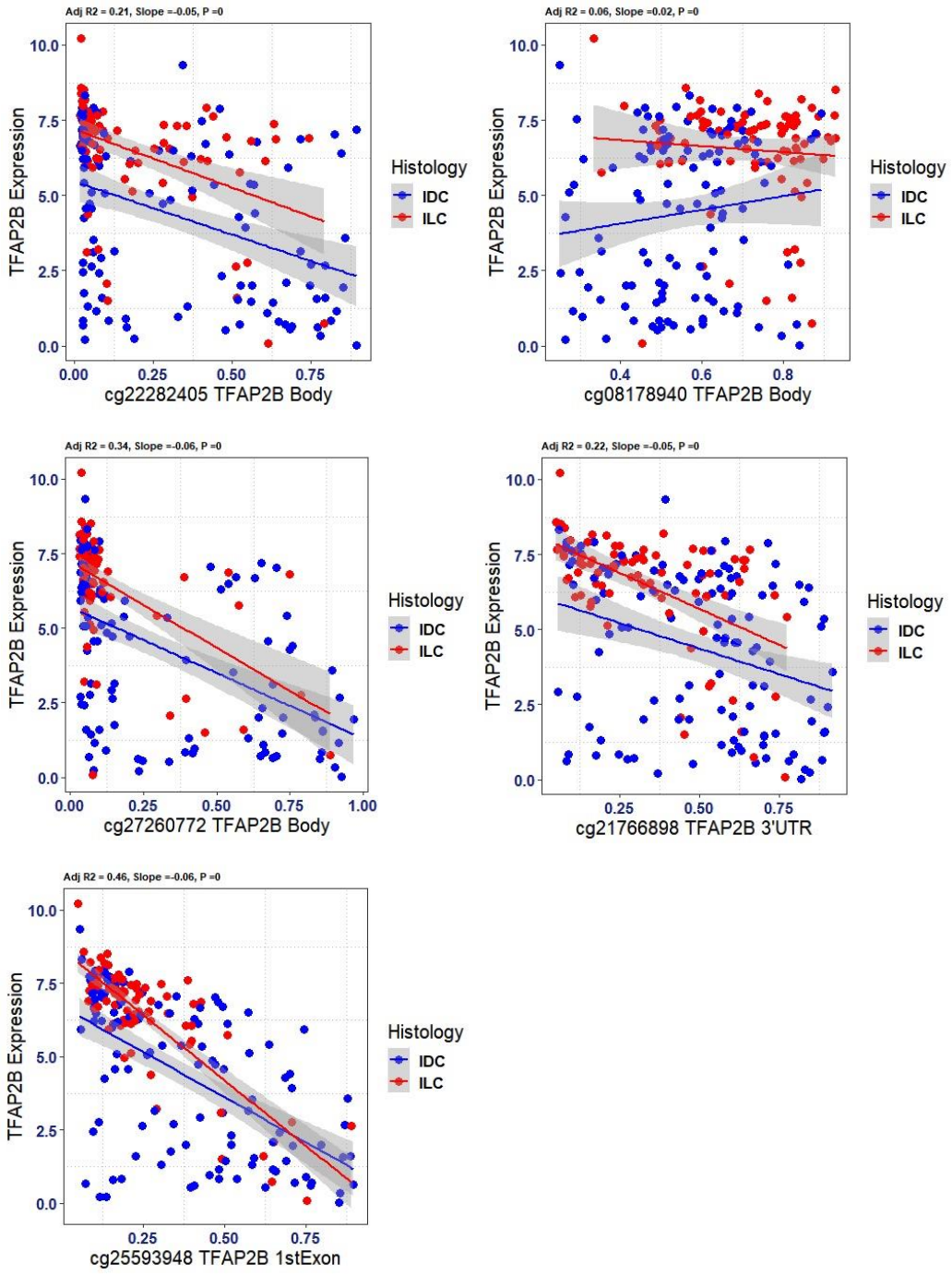


Figure S8 TFAP2B clqustering in ILC and IDC tumors based on methylation and expression

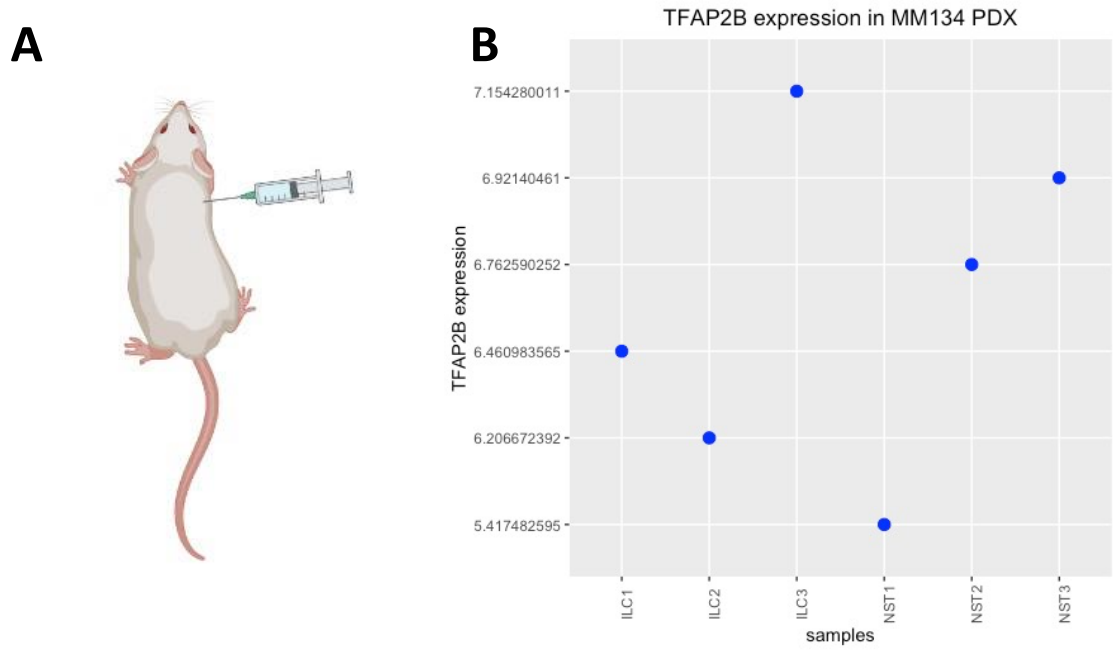
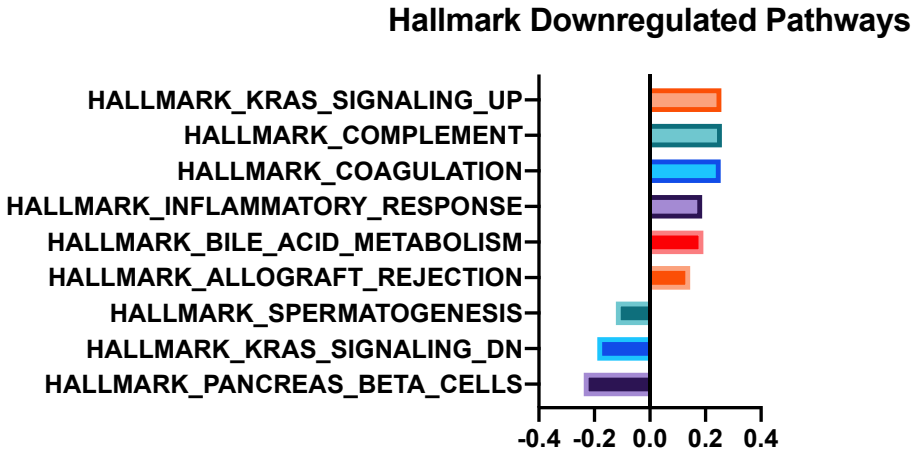


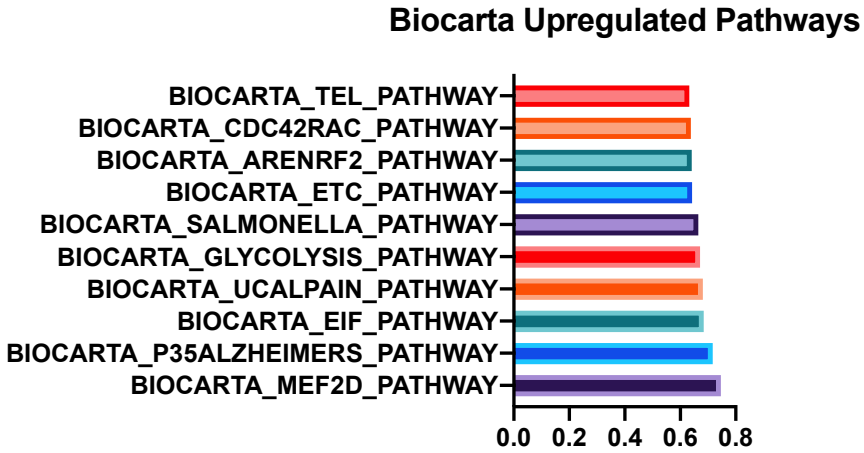
Figure S9 ILC xenografts exhibit high *TFAP2B* expression

Intraductal xenografts developed from ILC-derived cell line models. (A) Cell lines were infected with RFP-luciferase 2 expressing lentivirus, selected by flow cytometry and injected through milk ducts of immunocompromised mice. (B) TFAP2B expression in ILC mice and ER+ HER2- non-special type (NST).

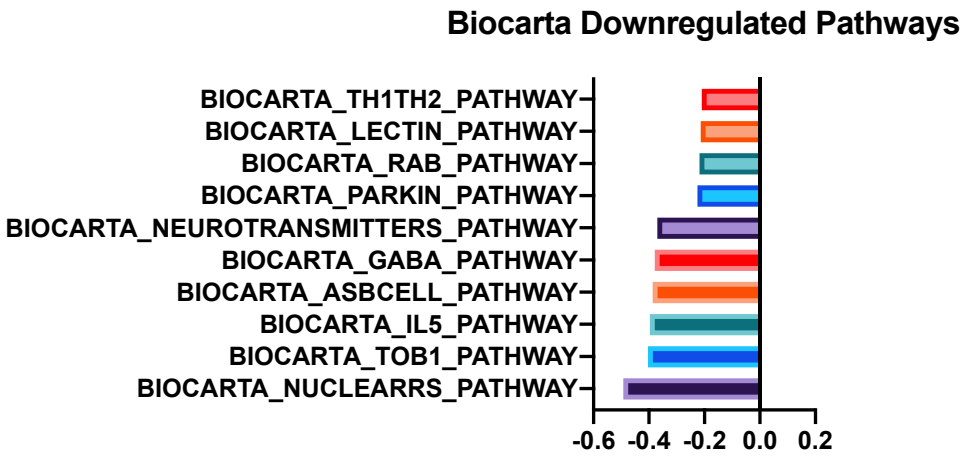
A

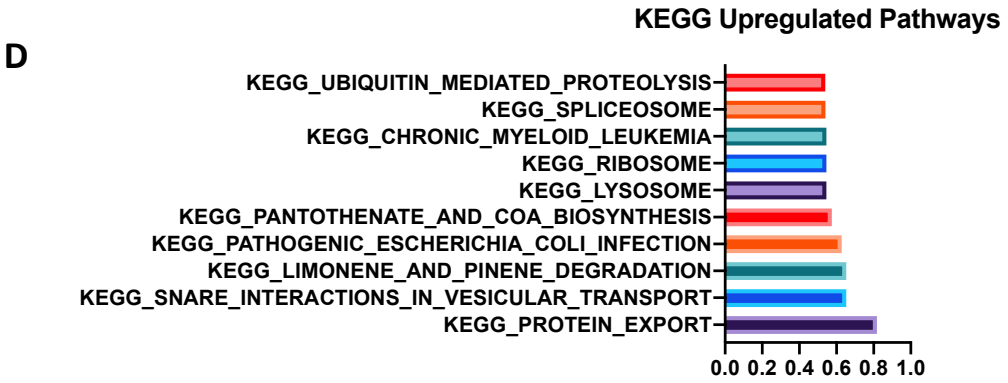

Hallmark Downregulated Pathways
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_COMPLEMENT
HALLMARK_COAGULATION
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_SPERMATOGENESIS
HALLMARK_KRAS_SIGNALING_DN
HALLMARK_PANCREAS_BETA_CELLS
-0.4 -0.2 0.0 0.2 0.4

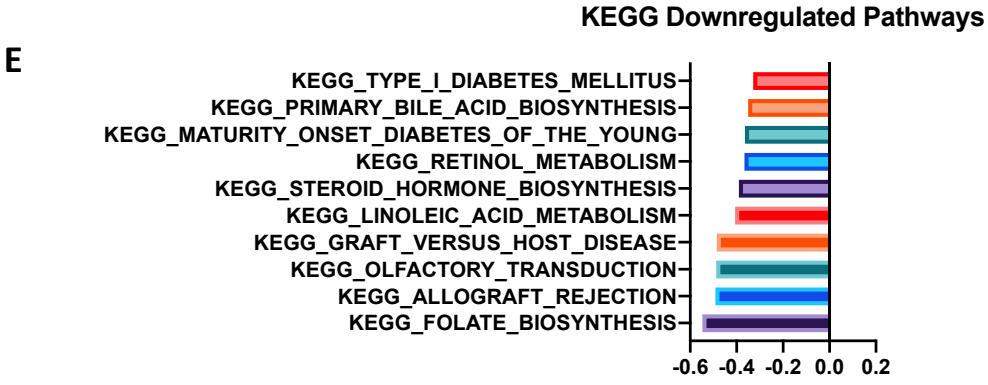
B

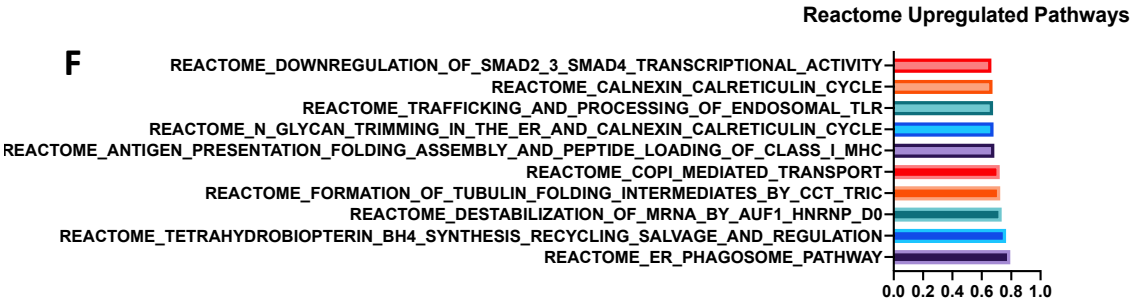

Biocarta Upregulated Pathways
BIOCARTA_TEL_PATHWAY
BIOCARTA_CDC42RAC_PATHWAY
BIOCARTA_ARENRF2_PATHWAY
BIOCARTA_ETC_PATHWAY
BIOCARTA_SALMONELLA_PATHWAY
BIOCARTA_GLYCOLYSIS_PATHWAY
BIOCARTA_UCALPAIN_PATHWAY
BIOCARTA_EIF_PATHWAY
BIOCARTA_P35ALZHEIMERS_PATHWAY
BIOCARTA_MEF2D_PATHWAY
0.0 0.2 0.4 0.6 0.8

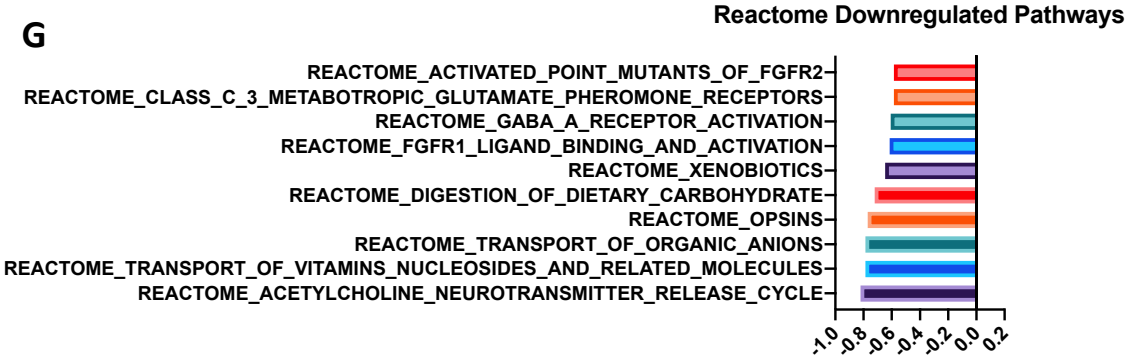
C


Biocarta Downregulated Pathways
BIOCARTA_TH1TH2_PATHWAY
BIOCARTA_LECTIN_PATHWAY
BIOCARTA_RAB_PATHWAY
BIOCARTA_PARKIN_PATHWAY
BIOCARTA_NEUROTRANSMITTERS_PATHWAY
BIOCARTA_GABA_PATHWAY
BIOCARTA_ASBCELL_PATHWAY
BIOCARTA_IL5_PATHWAY
BIOCARTA_TOB1_PATHWAY
BIOCARTA_NUCLEARRS_PATHWAY
-0.6 -0.4 -0.2 0.0 0.2

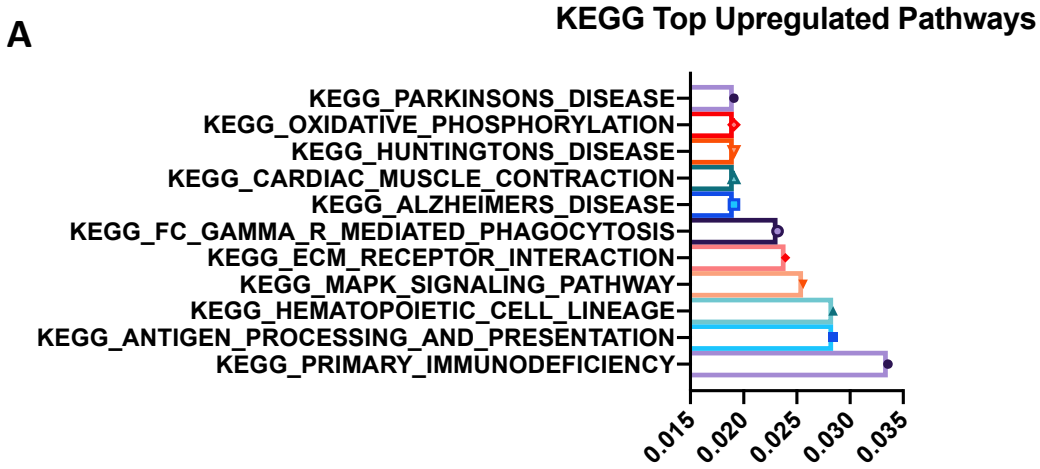

D
KEGG Upregulated Pathways
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS
KEGG_SPLICEOSOME
KEGG_CHRONIC_MYELOID_LEUKEMIA
KEGG_RIBOSOME
KEGG_LYSOSOME
KEGG_PANTOTHENATE_AND_COA_BIOSYNTHESIS
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION
KEGG_LIMONENE_AND_PINENE_DEGRADATION
KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT
KEGG_PROTEIN_EXPORT
0.0 0.2 0.4 0.6 0.8 1.0


E
KEGG Downregulated Pathways
KEGG_TYPE_I_DIABETES_MELLITUS
KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS
KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG
KEGG_RETINOL_METABOLISM
KEGG_STEROID_HORMONE_BIOSYNTHESIS
KEGG_LINOLEIC_ACID_METABOLISM
KEGG_GRAFT_VERSUS_HOST_DISEASE
KEGG_OLFACTORY_TRANSDUCTION
KEGG_ALLOGRAFT_REJECTION
KEGG_FOLATE_BIOSYNTHESIS
-0.6 -0.4 -0.2 0.0 0.2

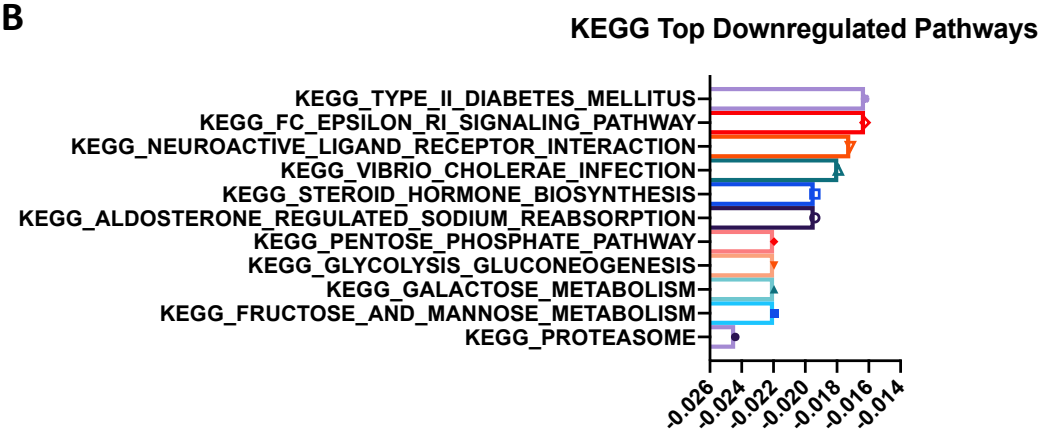

F
Reactome Upregulated Pathways
REACTOME_DOWNREGULATION_OF_SMAD2_3_SMAD4_TRANSCRIPTIONAL_ACTIVITY
REACTOME_CALNEXIN_CALRETICULIN_CYCLE
REACTOME_TRAFFICKING_AND_PROCESSING_OF_ENDOSOMAL_TLR
REACTOME_N_GLYCAN_TRIMMING_IN_THE_ER_AND_CALNEXIN_CALRETICULIN_CYCLE
REACTOME_ANTIGEN_PRESENTATION_FOLDING_ASSEMBLY_AND_PEPTIDE_LOADING_OF_CLASS_I_MHC
REACTOME_COPI_MEDIATED_TRANSPORT
REACTOME_FORMATION_OF_TUBULIN_FOLDING_INTERMEDIATES_BY_CCT_TRIC
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0
REACTOME_TETRAHYDROBIOPTERIN_BH4_SYNTHESIS_RECYCLING_SALVAGE_AND_REGULATION
REACTOME_ER_PHAGOSOME_PATHWAY
0.0 0.2 0.4 0.6 0.8 1.0


G
Reactome Downregulated Pathways
REACTOME_ACTIVATED_POINT_MUTANTS_OF_FGFR2
REACTOME_CLASS_C_3_METABOTROPIC_GLUTAMATE_PHEROMONE_RECEPTORS
REACTOME_GABA_A_RECEPTOR_ACTIVATION
REACTOME_FGFR1_LIGAND_BINDING_AND_ACTIVATION
REACTOME_XENOBIOTICS
REACTOME_DIGESTION_OF_DIETARY_CARBOHYDRATE
REACTOME_OPSINS
REACTOME_TRANSPORT_OF_ORGANIC_ANIONS
REACTOME_TRANSPORT_OF_VITAMINS_NUCLEOSIDES_AND_RELATED_MOLECULES
REACTOME_ACETYLCHOLINE_NEUROTRANSMITTER_RELEASE_CYCLE
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2

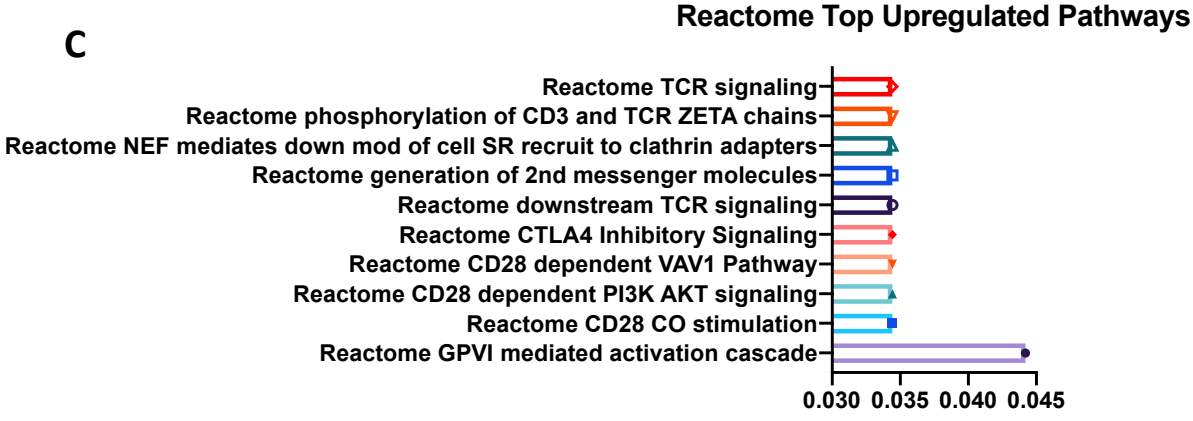
A


KEGG Top Upregulated Pathways
KEGG_PARKINSONS_DISEASE
KEGG_OXIDATIVE_PHOSPHORYLATION
KEGG_HUNTINGTONS_DISEASE
KEGG_CARDIAC_MUSCLE_CONTRACTION
KEGG_ALZHEIMERS_DISEASE
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS
KEGG_ECM_RECEPTOR_INTERACTION
KEGG_MAPK_SIGNALING_PATHWAY
KEGG_HEMATOPOIETIC_CELL_LINEAGE
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
KEGG_PRIMARY_IMMUNODEFICIENCY
0.015
0.020
0.025
0.030
0.035

B


KEGG Top Downregulated Pathways
KEGG_TYPE_II_DIABETES_MELLITUS
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION
KEGG_VIBRIO_CHOLERAE_INFECTION
KEGG_STEROID_HORMONE_BIOSYNTHESIS
KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION
KEGG_PENTOSE_PHOSPHATE_PATHWAY
KEGG_GLYCOLYSIS_GLUCONEOGENESIS
KEGG_GALACTOSE_METABOLISM
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM
KEGG_PROTEASOME
-0.026
-0.024
-0.022
-0.020
-0.018
-0.016
-0.014

C


Reactome Top Upregulated Pathways
Reactome TCR signaling
Reactome phosphorylation of CD3 and TCR ZETA chains
Reactome NEF mediates down mod of cell SR recruit to clathrin adapters
Reactome generation of 2nd messenger molecules
Reactome downstream TCR signaling
Reactome CTLA4 Inhibitory Signaling
Reactome CD28 dependent VAV1 Pathway
Reactome CD28 dependent PI3K AKT signaling
Reactome CD28 CO stimulation
Reactome GPVI mediated activation cascade
0.030
0.035
0.040
0.045

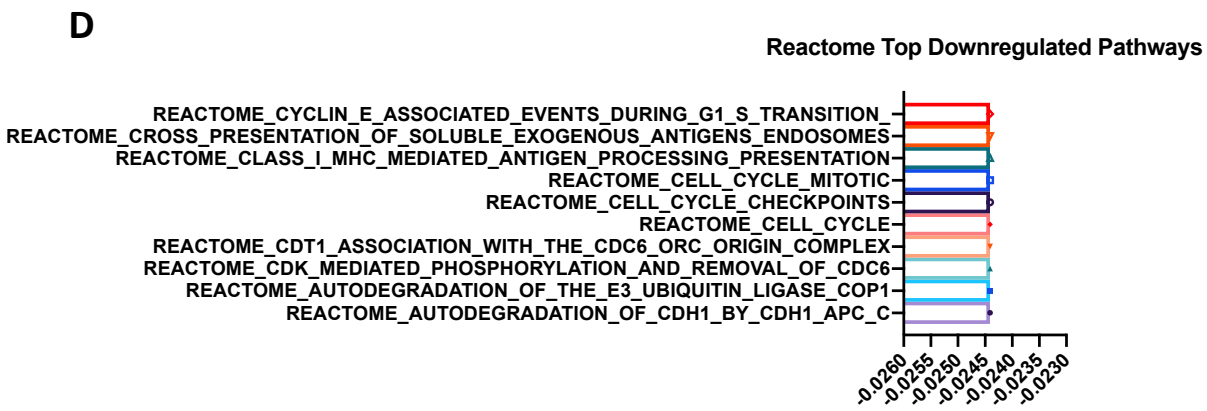


Figure S10 Differentially expressed pathways in high and low TFAP2B expressing ER+ LumA ILC and IDC tumors reveals potential mechanism

(A-G)Upregulated and downregulated HALLMARK, BIOCARTA, KEGG and REACTOME pathways. P-value is indicated for each pathway.

Table S1 Fold Changes of Individual genes high *TFAP2B* expression in ER+ LumA tumors between ILC and IDC

Rank	Genes	Log2FC in TCGA, TFAP2B high ILC/IDC
1	DCD	-2.6893249
2	COL2A1	-1.9501804
3	MYL1	-1.7999967
4	HS3ST5	-1.4680836
5	PHGR1	-1.3951797
6	IGFBP1	-1.3856613
7	FBXL21	-1.3235915
8	INSL6	-1.3119345
9	CRCT1	-1.3113283
10	P2RX2	-1.2820137
11	MIR7.3HG	-1.2614281
12	DPY19L2P4	-1.2591158
13	GLDC	-1.1892743
14	FAM71E2	-1.1461757
15	CALY	-1.1419399
16	ADGB	-1.1396018
17	IRS4	-1.1138267
18	BPIFB2	-1.1136522
19	ONECUT1	-1.107128
20	KLK15	-1.1024549
21	CHRM1	-1.0979901
22	FLJ34503	-1.0974161
23	C20orf141	-1.0557548
24	CDHR3	-1.0411197
25	KLK13	-1.0357184
26	DDC	1.03363575
27	CHRNB4	1.04041088
28	ADRB1	1.04257375
29	CXCR5	1.0650713
30	PI3	1.06737015
31	CALML5	1.08737818

Table S1 continued		
32	CD19	1.09617095
33	CHGB	1.10646759
34	IL29	1.12307889
35	PI15	1.16303314
36	KCNJ4	1.16400371
37	LOC100216479	1.1703468
38	FCER2	1.18133534
39	PLA2G2D	1.20011704
40	CXCR2P1	1.23401494
41	IFNB1	1.28073829
42	AKR1B15	1.29688524
43	CLEC4C	1.30548151
44	AKR1B10	1.32587527
45	ARHGAP36	1.40313417
46	FAM181B	1.59498184
47	LOC340017	1.60111851
48	LACRT	1.77734453
49	MAGEB1	1.83294393
50	GRIA2	1.91541669
51	PCDH10	1.93973113
52	CDH18	2.0917741
53	MAGEA1	2.60565668
54	LOC400643	2.98166713
55	NTS	3.24792183
56	PNLIPRP3	3.66732838

Bibliography

1. "U.S. Breast Cancer Statistics." *Breastcancer.org,* American Cancer Society, 4 Feb. 2021.

2. "Breast Cancer - Statistics." *Cancer.Net*, 24 Feb. 2021.

3. Inic, Zorka, et al. "Difference between Luminal A and Luminal B Subtypes According to Ki-67, Tumor Size, and Progesterone Receptor Negativity Providing Prognostic Information." *Clinical Medicine Insights: Oncology*, vol. 8, 2014, doi:10.4137/cmo.s18006.
4. ZHANG, MEI HONG, et al. "Estrogen Receptor-Positive Breast Cancer Molecular Signatures and Therapeutic Potentials (Review)." *Biomedical Reports*, vol. 2, no. 1, 2013, pp. 41–52., doi:10.3892/br.2013.187.
5. Jia, Zhenyu. "Expression of HER2 in Breast Cancer Promotes a Massive Reorganization of Gene Activity and Suggests a Role for Epigenetic Regulation." *Journal of Data Mining in Genomics & Proteomics*, vol. 03, no. 03, 2012, doi:10.4172/2153-0602.1000e102.

6. Badowska-Kozakiewicz, Anna M., and Michał P. Budzik. "Immunohistochemical Characteristics of Basal-like Breast Cancer." *Współczesna Onkologia*, vol. 6, 2016, pp. 436–443., doi:10.5114/wo.2016.56938.

7. Yaşar, Pelin, et al. "Molecular Mechanism of Estrogen-Estrogen Receptor Signaling." *Reproductive Medicine and Biology*, vol. 16, no. 1, 2016, pp. 4–20., doi:10.1002/rmb2.12006.

8. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-752.

9. Yersal, Ozlem. "Biological Subtypes of Breast Cancer: Prognostic and Therapeutic Implications." *World Journal of Clinical Oncology*, vol. 5, no. 3, 2014, p. 412., doi:10.5306/wjco.v5.i3.412.

10. Ciriello, Giovanni, et al. "Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer." *Cell*, vol. 163, no. 2, 2015, pp. 506–519., doi:10.1016/j.cell.2015.09.033.

11. McCart Reed, Amy E., et al. "Invasive Lobular Carcinoma of the Breast: the Increasing Importance of This Special Subtype." *Breast Cancer Research*, vol. 23, no. 1, 2021, doi:10.1186/s13058-020-01384-6.

12. Yaşar, Pelin, et al. "Molecular Mechanism of Estrogen-Estrogen Receptor Signaling." *Reproductive Medicine and Biology*, vol. 16, no. 1, 2016, pp. 4–20., doi:10.1002/rmb2.12006.

13. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-752.

14. Burszein, Harold J., et al. “Adjuvant Endocrine Therapy for Women With Hormone Receptor–Positive Breast Cancer: ASCO Clinical Practice Guideline Focused Update.” *Journal of Clinical Oncology*, vol. 37, no. 5, 2019, pp. 423–438., doi:10.1200/jco.18.01160.

15. Burstein, Harold J., et al. “Adjuvant Endocrine Therapy for Women With Hormone Receptor–Positive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update on Ovarian Suppression Summary.” *Journal of Oncology Practice*, vol. 12, no. 4, 2016, pp. 390–393., doi:10.1200/jop.2016.011239.

16. Patel, Hitisha K., and Teeru Bihani. “Selective Estrogen Receptor Modulators (SERMs) and Selective Estrogen Receptor Degraders (SERDs) in Cancer Treatment.” *Pharmacology & Therapeutics*, vol. 186, 2018, pp. 1–24.,

17. Wang, Jiani, and Binghe Xu. “Targeted Therapeutic Options and Future Perspectives for HER2-Positive Breast Cancer.” *Signal Transduction and Targeted Therapy*, vol. 4, no. 1, 2019, doi:10.1038/s41392-019-0069-2.

18. Wilson, Natalie, et al. “Lobular Breast Cancer: A Review.” *Frontiers in Oncology*, vol. 10, 2021, doi:10.3389/fonc.2020.591399.

19. Furrer, Daniela, et al. “The Human Epidermal Growth Factor Receptor 2 (HER2) as a Prognostic and Predictive Biomarker: Molecular Insights into HER2 Activation and Diagnostic Implications.” *Cancer Prognosis*, 2018, doi:10.5772/intechopen.78271.

20. Tai, Wanyi, et al. “The Role of HER2 in Cancer Therapy and Targeted Drug Delivery.” *Journal of Controlled Release*, vol. 146, no. 3, 2010, pp. 264–275., doi:10.1016/j.jconrel.2010.04.009.

21. Fuentes, Nathalie, and Patricia Silveyra. “Estrogen Receptor Signaling Mechanisms.” *Advances in Protein Chemistry and Structural Biology*, 2019, pp. 135–170., doi:10.1016/bs.apcsb.2019.01.001.

22. Shanle, Erin K., and Wei Xu. “Selectively Targeting Estrogen Receptors for Cancer Treatment.” *Advanced Drug Delivery Reviews*, vol. 62, no. 13, 2010, pp. 1265–1276., doi:10.1016/j.addr.2010.08.001.

23. Yu, F, and W Bender. “The Mechanism of Tamoxifen in Breast Cancer Prevention.” *Breast Cancer Research*, vol. 3, no. S1, 2001, doi:10.1186/bcr404.

24. Dowsett, Mitch, et al. “Biological Characteristics of the Pure Antiestrogen Fulvestrant: Overcoming Endocrine Resistance.” *Breast Cancer Research and Treatment*, vol. 93, no. S1, 2005, pp. 11–18., doi:10.1007/s10549-005-9037-3.

25. Fabian, C. J. “The What, Why and How of Aromatase Inhibitors: Hormonal Agents for Treatment and Prevention of Breast Cancer.” *International Journal of Clinical Practice*, vol. 61, no. 12, 2007, pp. 2051–2063., doi:10.1111/j.1742-1241.2007.01587.x.

26. Shackleford, Madeleine T., et al. “Estrogen Controls MTOR Signaling and Mitochondrial Function via WNT4 in Lobular Carcinoma Cells.” 2019, doi:10.1101/684530.

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

28. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature***486,** 346–52 (2012).

29. Nik-Zainal, S. *et al.* Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **149,** 979–993 (2012).

30. Thomas, C. & Gustafsson, J.-Å. Estrogen receptor mutations and functional consequences for breast cancer. *Trends Endocrinol. Metab.* **26,** 467–476 (2015).

31. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).

32. Björnström, L. & Sjöberg, M. Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes. *Mol. Endocrinol.* **19,** 833–842 (2005).

33. Lee, A. V, Cui, X. & Oesterreich, S. Cross-talk among estrogen receptor, epidermal growth factor, and insulin-like growth factor signaling in breast cancer. *Clin. Cancer Res.* **7,** 4429s–4435s; discussion 4411s–4412s (2001).

34. McCart Reed, A.E., Kalinowski, L., Simpson, P.T. *et al.* Invasive lobular carcinoma of the breast: the increasing importance of this special subtype. *Breast Cancer Res* 23, 6 (2021). https://doi.org/10.1186/s13058-020-01384-6

35. Chen, Xiao, et al. “DNA Methylation Profiles Capturing Breast Cancer Heterogeneity.” *BMC Genomics*, vol. 20, no. 1, 2019, doi:10.1186/s12864-019-6142-y.

36. Fleischer, Thomas, et al. “Genome-Wide DNA Methylation Profiles in Progression to in Situand Invasive Carcinoma of the Breast with Impact on Gene Transcription and Prognosis.” *Genome Biology*, vol. 15, no. 8, 2014, doi:10.1186/s13059-014-0435-x.

37. Johnson, Kevin C., et al. “DNA Methylation in Ductal Carcinoma in Situ Related with Future Development of Invasive Breast Cancer.” *Clinical Epigenetics*, vol. 7, no. 1, 2015, doi:10.1186/s13148-015-0094-0.

38. Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature, 490*(7418), 61-70. doi:10.1038/nature11412

39. Ping, Zheng, et al. “A Microscopic Landscape of the Invasive Breast Cancer Genome.” *Scientific Reports*, vol. 6, no. 1, 2016, doi:10.1038/srep27545.

40. Lustberg, Maryam B., and Bhuvaneswari Ramaswamy. “Epigenetic Therapy in Breast Cancer.” *Current Breast Cancer Reports*, vol. 3, no. 1, 2010, pp. 34–43., doi:10.1007/s12609-010-0034-0.

41. Stefansson, Olafur A., and Manel Esteller. “Epigenetic Modifications in Breast Cancer and Their Role in Personalized Medicine.” *The American Journal of Pathology*, vol. 183, no. 4, 2013, pp. 1052–1063., doi:10.1016/j.ajpath.2013.04.033.

42. Williams T, Tjian R. Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. 1991 Apr;5(4):670-82.

43. Eckert D, Buhl S, Weber S, Jager R, Schorle H. The AP-2 family of transcription factors. Genome Biol. 2005;6(13):246. Epub 2005 Dec 28.

44. Sun P, Yoshizuka N, New L, Moser BA, Li Y, Liao R, Xie C, Chen J, Deng Q, Yamout M, Dong MQ, Frangou CG, Yates JR 3rd, Wright PE,Han J. PRAK is essential for rasinduced senescence and tumor suppression. Cell. 2007 Jan 26;128(2):295-308.

45. Raap, Mieke, et al. “Lobular Carcinoma in Situ and Invasive Lobular Breast Cancer Are Characterized by Enhanced Expression of Transcription Factor AP-2β.” *Laboratory Investigation*, vol. 98, no. 1, 2017, pp. 117–129., doi:10.1038/labinvest.2017.106.

46. Cope LM, Fackler MJ, Lopez-Bujanda Z, Wolff AC, Visvanathan K, Gray JW, Sukumar S, Umbricht CB. Do breast cancer cell lines provide a relevant model of the patient tumor methylome? PLoS One. 2014 Aug 26;9(8):e105545. doi: 10.1371/journal.pone.0105545. PMID: 25157401; PMCID: PMC4144876.

47. *Broad Institute Cancer Cell Line Encyclopedia (CCLE)*, MIT and Harvard, portals.broadinstitute.org/ccle/data.

48. “Table 1: The Single Nucleotide Polymorphisms in Cathepsin B Protein Mined from Literature (PMID: 16492714).” doi:10.7717/peerj.7425/table-1.

49. Barretina, Jordi, et al. “The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity.” *Nature*, vol. 483, no. 7391, 2012, pp. 603–607., doi:10.1038/nature11003.

50. Gendoo, Deena M., et al. “Genefu: an R/Bioconductor Package for Computation of Gene Expression-Based Signatures in Breast Cancer.” *Bioinformatics*, vol. 32, no. 7, 2015, pp. 1097–1099., doi:10.1093/bioinformatics/btv693.

51. Balwierz, P. J., et al. “ISMARA: Automated Modeling of Genomic Signals as a Democracy of Regulatory Motifs.” *Genome Research*, vol. 24, no. 5, 2014, pp. 869–884., doi:10.1101/gr.169508.113.

52. Su, A. I., et al. “A Gene Atlas of the Mouse and Human Protein-Encoding Transcriptomes.” *Proceedings of the National Academy of Sciences*, vol. 101, no. 16, 2004, pp. 6062–6067., doi:10.1073/pnas.0400782101.

53. Ross, Douglas T., et al. “Systematic Variation in Gene Expression Patterns in Human Cancer Cell Lines.” *Nature Genetics*, vol. 24, no. 3, 2000, pp. 227–235., doi:10.1038/73432.

54. Kulakovskiy, Ivan V, et al. “HOCOMOCO: towards a Complete Collection of Transcription Factor Binding Models for Human and Mouse via Large-Scale ChIP-Seq Analysis.” *Nucleic Acids Research*, vol. 46, no. D1, 2017, doi:10.1093/nar/gkx1106.

55. Bertoli, Cosetta, et al. “Control of Cell Cycle Transcription during G1 and S Phases.” *Nature Reviews Molecular Cell Biology*, vol. 14, no. 8, 2013, pp. 518–528., doi:10.1038/nrm3629.

56. Michaut, M., Chin, SF., Majewski, I. *et al.* Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* 6, 18517 (2016). https://doi.org/10.1038/srep18517

57. Sflomos, George et al. “Intraductal xenografts show lobular carcinoma cells rely on their own extracellular matrix and LOXL1.” *EMBO molecular medicine* vol. 13,3 (2021): e13180. doi:10.15252/emmm.202013180

58. Ikram, Fakhera, et al. “Transcription Factor Activating Protein 2 Beta (TFAP2B) Mediates Noradrenergic Neuronal Differentiation in Neuroblastoma.” *Molecular Oncology*, vol. 10, no. 2, 2015, pp. 344–359., doi:10.1016/j.molonc.2015.10.020.

59. Li, Zhenglin, et al. “Activator Protein-2β Promotes Tumor Growth and Predicts Poor Prognosis in Breast Cancer.” *Cellular Physiology and Biochemistry*, vol. 47, no. 5, 2018, pp. 1925–1935., doi:10.1159/000491463.

60. Wajapeyee, Narendra, and Kumaravel Somasundaram. “Cell cycle arrest and apoptosis induction by activator protein 2alpha (AP-2alpha) and the role of p53 and p21WAF1/CIP1 in AP-2alpha-mediated growth inhibition.” *The Journal of biological chemistry* vol. 278,52 (2003): 52093-101. doi:10.1074/jbc.M305624200

61. Fu, Lingyi, et al. “TFAP2B Overexpression Contributes to Tumor Growth and a Poor Prognosis of Human Lung Adenocarcinoma through Modulation of ERK and VEGF/PEDF Signaling.” *Molecular Cancer*, vol. 13, no. 1, 2014, p. 89., doi:10.1186/1476-4598-13-89.

62. Pellikainen, Johanna M., and Veli-Matti Kosma. “Activator Protein-2 in Carcinogenesis with a Special Reference to Breast Cancer-A Mini Review.” *International Journal of Cancer*, vol. 120, no. 10, 2007, pp. 2061–2067., doi:10.1002/ijc.22648.