Exposure of mammary cells to lipid activates gene expression changes associated with ER-negative breast cancer via chromatin remodeling

Short title: Lipids, ER-negative breast cancer and chromatin

Shivangi Yadav[1], Ranya Virk[2], Carolina H Chung[3], David Van Derway[2], Duojiao Chen[7], Kirsten Burdett[8], Hongyu Gao[7], Zexian Zeng[9], Manish Ranjan[1], Gannon Cottone[1], Xiaoling Xuei[7], Sriram Chandrasekaran[3,4,5,6], Vadim Backman[2], Robert Chatterton[10], Seema Ahsan Khan[1*], Susan E Clare[1*]

[1]Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA.

[2] Department of Biomedical Engineering, Northwestern University, Evanston, IL 60208-2850, USA.

[3]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

[4] Program in Chemical Biology, University of Michigan, Ann Arbor, MI 48109, USA.

[5] Rogel Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA.

[6] Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

[7]Center of for Medical Genomics, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

[8] Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA.

[9]Department of Data Sciences, Dana Farber Cancer Institute, Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA.

[10] Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.

*To whom correspondence should be addressed:

Susan E Clare, MD, PhD

Research Associate Professor, Department of Surgery

Feinberg School of Medicine, Northwestern University

Robert H Lurie Medical Research Center Room 4-105

250 E Superior, Chicago, IL 60611

Email: susan.clare@northwestern.edu

Tel: 312-926-3021

Seema A Khan, MD

Professor, Department of Surgery

Feinberg School of Medicine, Northwestern University

NMH/Prentice Women's Hospital Room 4-420

250 E Superior, Chicago, IL 60611

Email: s-khan2@northwestern.edu

Tel: 312-503-4236

Abstract

Improved understanding of local breast biology that favors the development of estrogen receptor negative (ER-) breast cancer (BC) would foster better prevention strategies. We have previously shown that overexpression of specific lipid metabolism genes is associated with the development of ER- BC. We now report results of exposure of MCF-10A cells and mammary organoids to representative medium- and long-chain polyunsaturated fatty acids. This exposure caused a dynamic and profound change in gene expression, accompanied by changes in chromatin packing density, chromatin accessibility and histone posttranslational modifications (PTMs). We identified 38 metabolic reactions that showed significantly increased activity, including reactions related to one-carbon metabolism. Among these reactions are those that produce S-adenosyl-L-methionine for histone PTMs. Utilizing both an *in-vitro* model and samples from women at high risk for ER- BC, we show that lipid exposure engenders gene expression, signaling pathway activation, and histone marks associated with the development of ER- BC.

Introduction

Breast cancer is a heterogeneous disease with different molecular subtypes that are characterized, at a minimum, by the expression of the estrogen receptor (ER), progesterone receptor (PR) and Human epidermal growth factor receptor 2 (HER2)/neu (*1*). Although multiple statistical tools have been developed to quantify breast cancer risk (*2*), they do not predict breast cancer subtypes. Current breast cancer prevention with selective estrogen receptor modulators (SERM) and aromatase inhibitors decreases the risk of estrogen-receptor (ER) positive breast cancer sub-types, but not those without ER expression (*3-5*). Thus, determining the etiologic/biologic factors that favor the development of ER-negative breast cancer will potentially enable the development of both strategies to identify women at risk for ER-negative disease as well as targeted preventive and therapeutic agents.

Given the poor understanding of the genesis of sporadic ER-negative breast cancer, we set out to study this using the contralateral, unaffected breast of patients with unilateral breast cancer as a model. Studies of metachronous contralateral breast cancer show a similarity in the ER status of the contralateral cancer to the index primary *(6-8)*. Therefore, the contralateral unaffected breast (CUB) of women undergoing surgical therapy for newly diagnosed unilateral breast cancer can be employed as a model to discover potential markers of subtype-specific risk. In a previous study, we performed Illumina expression arrays on epithelial cells from the CUB of breast cancer patients, and identified a lipid metabolism (LiMe) gene signature which was enriched in the CUBs of women with ER- breast cancer (*9*). Among these are genes that control critical steps in lipid and energy metabolism. We validated this signature in an independent set of 36 human samples and re-confirmed the above results in fresh frozen tissues obtained from a new set of ER+ and ER- breast cancer patients, each time using laser capture microdissection (LCM) to obtain epithelial

cells from tumor and CUB samples (*10*). Again, we found significantly higher expression of LiMe genes in CUBs from women with ER- breast cancer, compared to both CUBS from women with ER+ breast cancer, and breast epithelium from a control group of women undergoing reduction mammoplasty. However, the specific genes comprising this overexpressed set had no specific function or group of functions in common and did not suggest specific mechanistic explanations as to why lipid metabolism pathways would aid ER- breast cancer development. In the present study, we address possible mechanistic explanations for our previous observations.

Cellular metabolism is a complex sequence of reactions in response to a cell's microenvironment that have profound effects on cellular function (*11*). Major reprogramming of cellular energetics is one of two emerging hallmarks of cancer (*11*). Metabolic re-wiring is required to provide the energy required to enable continuous growth and proliferation of the cancer cells. The past century has witnessed intensive investigation of metabolic pathways in cancer, in particular that of aerobic glycolysis commonly called as the Warburg effect (*12*). However, this is not the singular anomaly in the metabolically altered cancer cell. In addition to glucose and glutamine, fatty acids are an extremely important energy source (*13*). Altered lipid metabolism is posited to be a driver of carcinogenesis in various cancers, including ovarian (*14*), prostate (*15, 16*), liver (*17*) and triple negative breast cancer (*18, 19*). Increased lipid metabolism has also been shown to serve as a survival signal that enables tumor recurrence and has been suggested as an Achilles heel for combating breast cancer progression (*20*). Despite this recognition of the importance of fatty acid metabolism, its role of in the transformation of a normal cell to the malignant state is largely unknown. Metabolomic studies of the concentrations of several free fatty acids in primary breast tumors, including linoleate, palmitate, and oleate, as a function of breast cancer subtype have revealed significant differences across the subtypes, with the highest concentrations in basal-like

106 breast cancer (*21*). Conjugation of long-chain fatty acids to carnitine for transport into the
107 mitochondria and subsequent fatty acid oxidation (FAO) was observed to be highest in basal-like
108 breast cancers, followed by luminal B ~HER2-enriched, with luminal A tumors displaying the
109 lowest levels (*21*). Another study, which utilized Raman spectroscopy to interrogate tissue,
110 revealed that histologically normal breast tissue centimeters removed from the breast malignancy
111 have significantly higher polyunsaturated fatty acid levels compared with normal tissue from
112 cancer-free subjects (*22*).

113 Metabolites from intermediate metabolism are the substrates used to generate chromatin
114 modifications, underlining a complex relationship between metabolism and epigenetics. Key to
115 the crosstalk between metabolism and chromatin structure, is that the kinetic and thermodynamic
116 properties of the chromatin modification reactions are commensurate with the dynamic range of
117 the physiological concentrations of the corresponding intermediates in metabolism (*23*). For
118 example, the substrates for histone methylation and acetylation reactions often have cellular
119 concentrations that are commensurate with enzyme Km values, and thus are sensitive and
120 responsive to changes in metabolism. Historically, glucose-derived carbon has been considered
121 the primary source of acetyl-coA for histone acetylation. In the nucleus, acetate may be a minor
122 source. Recently, however, data from McDonnell and colleagues has revealed that lipids
123 reprogram metabolism to become a major carbon source for histone acetylation (*24*). This
124 reprogramming was shown to have significant effects on gene expression. Therefore, we sought
125 to determine if the LiMe signature we observed in the CUBs of ER- patients is associated with
126 chromatin modifications and histone PTMs secondary to changes in metabolism fostered by
127 exposure to medium and long chain fatty acids.

Results

Lipid facilitates transcriptional reprogramming in non-transformed mammary cells

We established an *in vitro* model by exposing estrogen and progesterone receptor (PR) negative MCF10A cells to octanoate, a medium chain eight-carbon fatty acid. Due to its small size and lipophilic nature octanoate does not depend on fatty acid transport proteins to traverse cell membranes and is readily oxidized in the mitochondria to form acetyl-CoA (*25, 26*). We performed RNA-seq to determine the effects of octanoate treatment on gene expression in the MCF10A cells. RNA-seq analysis revealed that 24 hours of octanoate treatment produces a transcriptional profile that is completely distinct from vehicle-treated controls (**Fig. 1A, Fig. S1A-B**). Genes with initially low expression (negative values of $\ln(E_{ctrl}/E_{ctrl,avg})$) are upregulated (corresponding to positive values of $\ln(E_{oct}/E_{ctrl})$) while genes with initially high expression (positive values of $\ln(E_{ctrl}/E_{ctrl,avg})$) are downregulated upon octanoate treatment (corresponding to negative values of $\ln(E_{oct}/E_{ctrl})$)(*27*). More specifically, there is a clear trend for initially highly expressed genes in the control condition to be downregulated upon octanoate treatment while genes with initial low expression in the control condition were upregulated. Differential expression analysis revealed a total of 2132 upregulated and 632 downregulated genes (FDR=0.01) in the octanoate treated cells (**Supplementary Fig. 1C**). Pathway enrichment analysis of the differentially expressed genes induced by the 5mM octanoate treatment was performed and the top 25 upregulated and downregulated pathways are shown in Fig. 1B. Specifically, this analysis revealed that among the top altered biological processes are second messenger mediated signaling, the Notch signaling pathway, adenylate cyclase-activating adrenergic receptor signaling, cell morphogenesis and differentiation. In contrast, downregulated genes are involved in cell cycle processes, transcriptional regulation of tumor suppressor genes such as p53, and cell cycle

checkpoints (**Fig. 1B**). Additional gene set enrichment analysis (GSEA) investigating top pathways with coordinated upregulation or downregulation of genes demonstrated that the top pathways associated with octanoate treatment included positive regulation of cell morphogenesis, a process involved in differentiation, as well as several oncogenic pathways associated with breast tumorigenesis, including *ERBB*, *WNT*, and *NOTCH* signaling pathways (**Fig. 1C**). Subsequent leading-edge analysis of these top upregulated signaling pathways- Lipid storage pathways (I), Wnt pathway (II), Notch signaling (III) and ERBB pathway (IV) shows clear association of core enrichment genes with octanoate treatment across replicates (**Fig. 1D**). Network analysis of octanoate-associated pathways identified by GSEA analysis revealed linked clusters involved with the nervous system and a second, separate group of linked clusters involved with growth factor stimulation, regulation of the *MAPK* cascade, and *ERBB* signaling (**Fig. 1E**). Finally, using real-time qPCR we validated the expression of a number of genes that GSEA analysis determined were significantly upregulated with octanoate treatment (**Fig. 1F**). Thus, treatment with medium chain fatty acids induces significant changes in transcription (*28*).

Evaluating the lipid composition of the serum of ER- and ER+ BC patients Next, we investigated whether dietary lipids, which are mainly long chain fatty acids (LCFAs), have a similar effect on the gene transcriptional profile to that of MCF10A cells. In order to determine the specific lipid(s) to evaluate experimentally, we sought to determine the differences in the percent composition of lipid species as a function of ER expression in serum from patients who had donated CUB samples for our original studies (*9, 10*). A comprehensive lipid profile of these serum samples was performed by the Northwest Metabolic Research Center at University of Washington, with measurement of more than 700 lipids. For each of the measurements, the association between the measured value and ER status was evaluated using regression models,

adjusting for BMI, age, and menopausal status. ER was a categorical variable used to describe subjects having ER + or ER – cancers, or controls undergoing reduction mammoplasty. As the purpose of this experiment was to identify a lipid for ensuing experiments, lipid species were ranked for effect size comparing serum from patients subjects with ER- disease to those with ER+ disease. Three of the top four lipid species with the largest effect size were noted to contain linoleic acid: cholesterol ester (CE) 18:2, phosphatidyl choline (PC)16:0/18:2 and triacylglycerol (TAG) 54:6-FA18:2. Linoleic acid as a free fatty acid ranked 11th in the analysis. Linoleic acid is the most highly consumed polyunsaturated fatty acid in the human diet (*29*), its presence in serum CE has been strongly correlated with intake (*30*), and its concentration in adult adipose tissue has more than doubled in the past half century (*31*). Therefore, all subsequent studies were carried out using linoleic acid (LA).

Linoleic acid influences chromatin packing behavior

The state of chromatin is intimately linked with the regulation of gene transcription, undergoing dynamic changes between transcriptionally active and inactive states. Thus, our next step was to explore the changes in chromatin structure of fatty acid treated MCF10A cells by employing partial wave spectroscopic (PWS) microscopy, which quantifies chromatin packing scaling (*D*) in live cells (*32*). *D* represents the power-law scaling relationship between the 1-D size of the chromatin polymer i.e. the number of nucleotides and the 3-D space the chromatin polymer occupies. Recent evidence indicates that higher chromatin packing scaling is associated with increased intercellular and intra-network transcriptional heterogeneity as well as increased malignancy and chemoresistance in cancer cells (*27, 33, 34*). PWS was used to evaluate the effect of LA on chromatin packing scaling in live MCF10A cells. Images were obtained every 6 hours over a 24-hour period. Our results showed significant increases in chromatin packing scaling upon exposure

to LA, in a manner similar to that of octanoate, suggesting that there is an increase in the dynamic range of gene expression and transcriptional gene network heterogeneity following lipid treatment (**Fig. 2A-B**). Thus, LA treatment results in changes in chromatin packing structure which are associated with a more malignant phenotype. Such significant changes in chromatin packing behavior also indicate significant changes in chromatin accessibility, which is directly associated with chromatin structure (*35*).

ATAC Sequencing reveals increased chromatin accessibility in regulatory regions of genes in the MAPK and cAMP signaling pathways in lipid treated mammary cells To acquire more detailed insight into the specific regions of open chromatin that were made accessible by LA treatment, we proceeded with ATAC sequencing on LA treated MCF-10A cells. We examined the genomic locations of ATAC-seq peaks, representing open chromatin sites, and discovered 1704 open chromatin sites. Open chromatin regions were overrepresented within 1 kb of transcription start sites (TSSs) by 40-fold relative to the whole genome **(Fig. 2C**). Further, KEGG pathway analysis revealed 326 open chromatin regions with a log fold change >= 1.5 and FDR < 0.05 compared to vehicle treated cells. Among the top pathways that were upregulated significantly upon LA treatment are MAPK signaling pathway, PI3K-AKT signaling pathway, and the cAMP adenylate cyclase pathway. Additionally, motif analysis conducted using 'HOMER'(*36*) showed that chromatin regions made accessible/inaccessible by LA treatment have binding motifs for a number of transcription factors (**Fig. 2E**). These data reveal that linoleic acid affects chromatin heterogeneity and increases/decreases the accessibility of specific regions that include transcription factor binding sites.

Notch pathway genes are overexpressed in patients at high risk of ER- disease

Next, we sought to determine whether the genes, or sets of genes/pathways that we identified in our *in vitro* study were also differentially expressed *in vivo* in tissue of patients at risk for ER- and ER+ breast cancer. We took advantage of RNA from the contralateral unaffected breast (CUB) of breast cancer cases utilized in our previous studies, which revealed the association of LiMe genes in the CUBs of women with unilateral ER- breast cancer (*9, 10*). We combined the data from the RNA and ATAC sequencing experiments and collated a list of 44 genes of interest and 3 housekeeping genes. The list consists of the genes from the *HEDGEHOG*, *NOTCH*, *WNT*, *EMT*, *PPARγ* and adenylate cyclase pathways (**supplementary file S1**). TaqMan low density arrays were utilized to measure expression of these genes in CUBs of ER- and ER+ cases compared with the reduction controls. The study population included 84 women, with participants comprised of 28 matched triplets of women with ER-positive breast cancer, ER-negative breast cancer, and reduction mammoplasty controls. The three groups were matched by age, race and menopausal status as shown in **Fig. S2A**. As noted in our original publication, ANOVA revealed a significant difference in BMI across the three groups with BMI in the reduction mammoplasty control group (30.0 ± 5.8) notably higher than in ER-negative cases (25.3 ± 6.3, p=0.015), but not significantly higher than in the ER-positive group (26.7 ± 5.5, p= 0.136) (*10*). There was no significant difference in HER2 status between ER-positive and ER-negative cases. The majority of the selected genes had higher expression in high risk CUB specimens than the controls, irrespective of the ER status of the index tumor (**Fig. S2B**). The comparison between the ER- and ER+ CUBs revealed that in the ER- CUBS there is increased expression of genes that function in the Notch pathway: *NOTCH1* (1.7-fold, p=0.002, BH_adjP=0.07), *NOTCH4* (1.7-fold, p=0.04, BH_adjP=0.3), *DLL4* (2.5-fold, p=0.7, BH_adjP=0.8) and *HEY 1* (1.5-fold, p=0.05, BH_adjP=0.3), in addition to the *SMO* gene (1.47-fold, p= 0.05, BH_adjP=0.3), which is a key

component of the hedgehog signaling pathway (**Fig. 3**). Altogether, these data reveal upregulation in *NOTCH* signaling in benign breast tissue samples from women at risk for ER- disease, suggesting that dysregulation of these pathways may play a role in the early stages of ER- cancer development.

LA increases the expression of Notch pathway genes and specific genes involved in fatty acid oxidation *in vitro* The increased expression of Notch pathway genes we discovered in the ER- CUBs, along with the similar findings in MCF10A cells exposed to octanoate (described above), led us to test the hypothesis that long chain fatty acids have similar effects on gene expression. We therefore investigated whether an increased LA environment influences the expression of Notch pathway genes and specific genes involved with fatty acid oxidation *in vitro*. We treated MCF-10A cells and mammary organoids from reduction mammoplasty patient samples with LA for 24 hours and then quantified changes in gene expression using RT-qPCR. To begin with, we assayed the genes involved in the activation of fatty acid oxidation. Upon entering cells, free fatty acids are converted into a fatty acyl-CoA molecules by the enzymes of the acyl-CoA synthetase (ACS) family (*37*). Notably, acyl-CoA synthetase long chain (*ACSL3*) is one of the LiME genes found to be upregulated in high risk ER- CUBs samples. Generation of acetyl-CoA occurs through a cyclical series of reactions in which a fatty acid is shortened by two carbons per cycle, eventually generating acetyl co-A. Acetyl co-A is a substrate for ketogenesis, which is initiated by the mitochondrial enzyme 3-hydroxy-3-methylglutaryl-CoA synthase 2 (*HMGCS2*), another of the previously identified LiMe genes. The mechanism for LCFAs oxidation is slightly more complex than for MCFAs, as this is regulated primarily via the enzyme carnitine palmitoyltransferase 1 (*CPT1*), the rate limiting enzyme of FAO which enables transport into the mitochondria. As shown in **Fig. 4A**, the expression of *HMGCS2*, *ACSL3*, and *CPT1B* were increased by LA exposure in

MCF-10A cells and mammary organoids. Additionally, we observed a significant increase in *DLL4* expression followed by *HEY1*, *HEY2* and *NOTCH1* in the lipid treated mammary cells (**Fig. 4A**). We revisited the ATAC sequencing data to examine the effect of LA on chromatin architecture near key genes in the *DLL4*/*NOTCH* signaling pathway, and observed increased accessibility around the transcription start sites of *DLL4*, *NOTCH1* and *HEY1* showing significant lowered chromatin density with p-values of 1.62e-17, 0.017 and 0.03 respectively (**Fig. 4B & Fig. C**).

Fatty acids drive flux through metabolic reactions resulting in increased histone methylation

While most of the experiments reported by McDonnell et al. were performed in AML 12 liver cells, these investigators also demonstrated increased H3K9 acetylation in octanoate-exposed MCF7 and MDAMB-231 breast cancer cells (*24*). Therefore, we sought to determine if these same experimental conditions would lead to H3K9 acetylation in a non-malignant MCF-10A cells. We exposed MCF-10A non-transformed ER - breast epithelial cell line to 5mM octanoate (C8) for 24 hours in medium containing both glucose (1.441 g/L) and glutamine (0.292 g/L). Western blot analysis demonstrated that octanoate exposure of MCF-10As resulted in increased acetylation at both H3K9 and H3K14 (**Fig. 5A**). To demonstrate that this was a fatty acid-specific effect, we treated the cells with 1,4-Cyclohexanedimethanol (1,4-CHDM), an alcohol with the same formula as octanoate; no acetylation was observed consequent to the alcohol exposure (**Fig. S3A**). To validate the specificity of the antibody against the acetylated histone lysines, we treated MCF-10A cells with sodium butyrate, a histone deacetylase (HDAC) inhibitor. Sodium butyrate treatment increased the acetylation of H3K9 and H3K14 as shown in **Fig. S3B**.

To exhaustively explore the impact of octanoate treatment on metabolic pathways, we used flux balance analysis (FBA) (*38*). FBA makes use of genome-scale metabolic network models that

contain all known metabolic reactions in a cell or tissue based on evidence from the published literature (*39*). Genome-scale metabolic models have been widely used to predict the metabolic behavior of various mammalian cell types (*40-44*). Here we used the Recon1 human network model that maps the relationship between 3744 reactions, 2766 metabolites, 1496 metabolic genes, and 2004 metabolic enzymes (*45*). This model was augmented with biochemical reactions corresponding to histone acetylation and methylation (*40, 46*), allowing us to predict the consequences of octanoate-induced metabolic changes on histone modifications by tracking the flux through the substrates for the histone modifications. These models were previously used to predict bulk histone acetylation levels in various cell lines based on the nuclear flux of acetyl-coA directed towards histone acetylation (*46*). Similarly, bulk histone methylation levels can be predicted based on the nuclear flux of S-adenosyl-L-methionine (SAM) (*40*).The model predicted octanoate treatment would result in increased histone methylation levels, with a more modest increase in histone acetylation levels (**Fig. 5C**). As a comparison, we repeated this analysis with immortalized hepatocyte cells used by McDonnel et al; they found a significant increase in histone acetylation after octanoate treatment (*24*). We calculated metabolic flux in these hepatocytes using the transcriptomics data from McDonnel *et al* and found a much larger increase in histone acetylation after octanoate treatment (**Fig. S3C**). These results suggest that the impact of metabolic alterations on histone acetylation is cell-type specific, as observed in prior studies (*47, 48*). Overall, out of the 3759 reactions in the model, we identified 38 that showed significant increased activity after octanoate treatment (p-value < 0.01; **Fig. S3C**). As expected, reactions involved in lipid and fatty acid metabolism, specifically triacyl glycerol synthesis and glycerophospholipid metabolism were upregulated. Interestingly, among the up-regulated reactions were several reactions related to the one-carbon metabolic pathway, which links folate, SAM, methionine, glycine and serine

metabolism (**Fig. 5D**). The reactions methionine adenosyltransferase, methionine synthase, adenosyl homocysteinase, 5,10-methylene-tetrahydrofolatereductase, glycine N-methyltransferase, and formyltetrahydrofolate dehydrogenase were all predicted to have increased activity after treatment (p-value < 0.01). Increased activity of the one-carbon pathway is associated with increased H3K4 trimethylation in stem cells and cancer cell lines (*40, 49*). These reactions likely support increased histone methylation by providing one carbon units.

Lipid exposure eventuates in histone methylation. In order to profile the specific histone marks significantly changed by the octanoate treatment we performed liquid chromatography/mass spectrometry on tryptic peptides isolated from the nuclei of treated and control MCF10A cells. Increased methylation was observed in various histone proteins including H3K9me1/2/3, H3.1K27me2/3, H3.3K36me2/3, H3K79me1/2 and H3K4 (**Fig. 5F**) together with increased acetylation of H3K14 and H4K16 (Fig. 5E). Notably, the GSEA analysis showed a significant correlation of H3K27 methylation (NES = 2.47, FDR q-value =0.05) and H3K4 methylation (NES= 1.24, FDR q-value = 0.1) with octanoate treatment (**Fig. S3 D-E**) suggesting this lipid rich environment eventuates in histone methylation in mammary epithelial cells.

Discussion

The known determinants of risk for ER-negative breast cancer are genetic (either specific racial inheritance, germline mutations in genes such as BRCA1) or systemic/behavioral factors (premenopausal obesity (*50*), absence of a breastfeeding (*51*)). In contrast, few if any local factors in the breast environment serve to identify women at risk for ER negative tumors. Local in-breast factors are of great interest however, since they may be more specifically targetable for breast cancer prevention than systemic factors. Of note, the two strongest risk factors for breast cancer overall (other than high penetrance germline mutations) are local: atypical proliferative lesions ,

and (*52*) extremely dense breast tissue (*53*) . This reasoning motivated us to investigate the local breast biology that may promote the development of ER negative rather than ER positive breast cancer, using the contralateral unaffected breast (CUB) of women undergoing surgery for a unilateral primary breast cancer as a model for ER-specific breast cancer risk (*7, 54*). In our initial study, we identified a highly correlated lipid metabolism (LiMe) gene signature, which was enriched in the CUBs of women with ER- breast cancer.

To explain the biologic basis for this association, we developed an *in vitro* model wherein we exposed either MCF10A, an ER negative, non-tumorigenic epithelial cell line, or breast organoids derived from reduction mammoplasty samples to an extracellular milieu rich in medium or long chain fatty acids. This model system has now enabled us to demonstrate that the exposure of breast epithelial cells to these fatty acids results in a dynamic and profound change in gene expression, accompanied by changes in chromatin packing density, chromatin accessibility and histone PTMs. The histone modifications, in turn, are the result of both the lipid-engendered increased expression of the requisite enzymes and the increased production of their substrates. Our metabolic flux analysis revealed the upregulation of several reactions related to the one-carbon metabolic pathway, which links folate, SAM, methionine, glycine and serine metabolism. This insight was not evident upon analysis of differential gene expression, which is not surprising as gene expression changes often do not reflect the flux of metabolic reactions (*40*).

Our proteomics data reveal increased methylation at H3K27me2/3, H3K36me3 and H3K9me2/3 in cells treated with octanoate; GSEA analysis showed that genes with ontologies related to histone methylation at H3K27 and H3K4 exhibit changes in expression in the lipid-treated cells. Methylation of H3K27 is carried out by *EZH2*, which showed a 1.65-fold increase in expression

(p=0.001) following exposure to octanoate. *EZH2* expression is sensitive to the level of oxidative phosphorylation. Our metabolic flux data demonstrated increased oxidative phosphorylation following exposure to octanoate; specifically flux through the electron transport chain (ETC). Inhibition of oxidative phosphorylation via complex I of the ETC by the biguanide phenformin markedly reduces EZH2 and SUZ12 protein expression (*55*). This suggests that increased H3K27 methylation may be a consequence of increased flux through the ETC increasing EZH2 expression in concert with increased production of its substrate SAM. Several studies have revealed a significant association of *EZH2* overexpression with ER negative breast cancer (*56*) or ER negative luminal progenitor cell expansion (*57*). EZH2 is the enzymatic subunit of Polycomb Repressive Complex 2, which catalyzes the trimethylation of H3K27. However, EZH2's actions are not be limited to its methyltransferase activity. EZH2 has been shown to bind to the *NOTCH 1* promoter resulting in increased *NOTCH1* transcription, stem cell expansion and accelerated tumor initiation (*58*). The effect of *NOTCH1* expression on mammary cell-lineage fate determination was recognized shortly after the identification of the mammary stem cell (*59*). Mammary stem cell differentiation is a hierarchical organization, and lineage tracing experiments have determined that *NOTCH1* expression exclusively generates ER- luminal cells (*60*). A subsequent study by these investigators revealed that during mammary embryogenesis Notch signaling prevents the generation of basal precursors, and cells expressing active *NOTCH1* exclusively give rise to the ER- (Sca1-/CD133-) lineage at any developmental stage from mouse embryonic day 13.5 to postpartum day 3 (*61*). Even more interesting given our focus on the origins of ER negative breast cancer was their observation that pubertal cells retain plasticity. Ectopic activation of Notch1 in basal cells at puberty was able to completely switch their identity to ER negative luminal cells.

Additional clues regarding the association of our experimental findings with ER negative breast cancer comes from GWAS data. A study that included 21,468 ER-negative cases and 100,594 controls identified independent associations of ten single nucleotide polymorphisms (SNPs) with the development of ER- breast cancer (*62*). Pathway analysis was performed by mapping each SNP to the nearest gene. This identified several pathways implicated in susceptibility to ER-negative, but not ER+ breast cancer. Included among these was the adenylate cyclase (AC) activating pathway. One of the significantly altered biologic processes that we identified by RNA sequencing of the octanoic acid treated cells is adenylate cyclase-activating adrenergic receptor signaling. Adenylate cyclase signals via cyclic AMP. Regions of chromatin with increased accessibility are associated with increased gene expression; our ATAC-Seq results show that linoleic acid exposure significantly increased accessibility to genes in the cAMP signaling pathway. In their discussion of ER- GWAS results, Milne et al. suggest that stimulation of the beta 2 adrenergic-adenylate cyclase-cAMP-β-arrestin–Src–ERK pathway may play a role in the genesis of ER- breast cancer. MetaCore analysis of our RNA-sequencing data reveals similar pathway activation, however, it is the beta1 adrenergic receptor that demonstrates increased expression in the octanoate treated cells. In addition, our ATAC-seq data showed increased RAP1 signaling pathway accessibility. Adenylate cyclase signaling also functions via Epac-Rap1-B-raf-MEK-ERK, with this signaling shown to be responsible for sustained *ERK* activation that occurs at a later time points (10-30 minutes) after cAMP activation (*63*). The MAPK (ERK) pathway can be stimulated by means other than adrenergic receptor ligand binding. Activation of this pathway by overexpression of *EGFR+EGF*, *c-erbB-2*, *RAF1* or *MEK* in MCF7 cells leads to estrogen-independent growth and down-regulation of ERα expression (*64*). These results suggest that hyperactivation of the MAPK(ERK) pathway plays a role in the generation of the ER- phenotype

in breast cancer. We observed *MAPK* activation in our analysis of differentially expressed genes, i.e., "positive regulation of the *MAPK* cascade," and in the analysis of regions of chromatin with significantly increased chromatin.

Using stratified LD score regression, a statistical method for identifying functional enrichment from GWAS summary statistics, SNPs associated with the H3K4me3 histone mark were determined to be contributing to the heritability of ER-negative breast cancer, (2.4-fold, P = 0.0005) (*62*). Increased activity of the one-carbon pathway is associated with increased H3K4 trimethylation in stem cells and cancer cell lines (*40, 49*). Restriction of methionine with consequent modulation of SAM and S-Adenosyl-L-homocysteine (SAH) levels affects methylation at H3K4me3, H3K27me3 and H3K9me3, with H3K4me3 exhibiting the largest changes (45). Interestingly, this restriction leads to loss of H3K4me3 at the promoters of colorectal cancer (CRC)-associated genes, with resulting decreased expression ($p = 0.02$, Fisher's exact test). A computational model developed to identify the direct influences on methionine concentrations in humans suggests that dietary intake explains about 30% of the variation in methionine concentration, and fats (arachidic acid in this model) are among the foods contributing to higher methionine levels (*49*).

One-carbon metabolism has multiple other functions in addition to producing SAM; one of which is to maintain redox homeostasis by producing NADPH. One of the earliest steps in breast tumorigenesis is the filling of the duct/acinar lumen with malignant cells. The viability of ECM-detached cells is dependent on combating the generation of reactive oxygen species (ROS) (*65*). For example, shuttling flux through the pentose phosphate pathway (PPP) promotes NADPH production and consequent reduction of ECM detachment-induced reactive oxygen species. It appears, however, that the process that produces the reducing equivalent is immaterial as cells can

also utilize NADPH-regenerating enzymes in the folate pathway as in metastasizing melanoma (*66*). Therefore, the one-carbon metabolism initiated by FAs may facilitate early tumorigenesis and the survival of matrix detached cells by the production of NADPH.

In conclusion, we have demonstrated in the present study that exposure of breast epithelial cells *in vitro* to fatty acids results in epigenetic effects that produce dynamic and profound changes in the expression of genes that have been associated with the development of ER- breast cancer. Next steps include demonstrating that these same changes are observed *in vivo*. As mentioned in the introduction, polyunsaturated fatty acids are present in normal breast tissue. Although we measured lipid species in the serum of the donors of the CUB specimens, fatty acids can also be mobilized from adjacent adipose tissue; adipocytes have been shown to be a reservoir of lipids for breast cancer stem cells (*67*). We hypothesize that the expression of genes associated with the development of ER- breast cancer is consequent to lipid stimulation of one-carbon metabolism with resultant changes in histone methylation. Important roles for glycolysis, glutaminolysis, lipogenesis and mitochondrial activity have been demonstrated in oncogenesis; the one-carbon pathway has comparatively received less attention and the insights we provide here generate new questions regarding lipid metabolism and ER negative breast cancer, to be pursued in future investigations.

Materials and Methods

Cell culture

MCF10A cell line was obtained from American Type Culture Collection (ATCC) and cultured in mammary epithelial cell growth basal medium with single quots supplements and growth factors (#Lonza CC-4136). Cells were treated with the medium-chain fatty acids (Sigma) sodium

447 octanoate (C8) dissolved in PBS, and long-chain fatty acids (Sigma) Linoleic acid (C18)
448 complexed with fatty acid free BSA (Roche 10775835001). PBS and BSA were used as the vehicle
449 control in experiments containing C8 and C18 respectively. Cells were counted using an Invitrogen
450 Countess automated cell counter using Trypan blue exclusion method and seeded at the indicated
451 densities. All experiments were done in complete MEBM media with fatty acids or vehicle.

452 **CUB Samples**

453 Patients diagnosed with unilateral breast cancer and undergoing contralateral prophylactic
454 mastectomy at Prentice Women's Hospital of Northwestern Medicine were recruited under an
455 approved protocol (NU11B04), with exclusions for neoadjuvant treatment, prior endocrine therapy
456 or pregnancy/lactation during the prior 2 years. A group of reduction mammoplasty (RM) patients
457 were also recruited as standard risk controls. The fresh tissues were frozen and stored in liquid
458 nitrogen. Tissue samples from 56 bilateral mastectomy cases (28 ER+ and 28 ER–) and 28 healthy
459 RM controls were used in this study. The ER+ cases, ER– cases and controls were matched by
460 age, race, and menopausal status.

461 **Mammary Organoids Preparation**

462 Tissues were collected from the non-obese, premenopausal women coming for the reduction
463 mammoplasty. Transfer the breast tissue to be processed into a sterile petri dish. Chop big breast
464 tissue mass into small pieces. Transfer the minced tissue to a sterile 50ml tube and add 30ml of
465 Kaighn's Modification media (Gibco #21127022) containing collagenase from Clostridium
466 histolyticum (Sigma Aldrich, catalog no. C0130), final collagenase concentration is 1 mg/mL.
467 Media containing collagenase is filtered using 0.22 μm filter. The falcon is sealed with parafilm
468 and tissue is gently dissociated on a shaker at 100 rpm and 37°C, overnight (16 hours). Following

day, organoids are collected by the centrifugation of the suspension at 800 rpm for 5 min. Discard the supernatant and wash the organoid pellet two-three times with PBS. Organoids with a size between 40-100uM are collected and resuspended in the fresh media (3mL) and added to a 6 well plate (Ultra-Low Attachment Surface plate, Corning # CLS3471**)**. Organoids are allowed to stabilize for 24 hours before using it for the experiments.

Fatty acid preparation

Sodium octanoate (C8) was dissolved in PBS. To bind linoleic acid (Sigma # L8134) to BSA, they were initially dissolved in water to yield a 50 mM final concentration. Dissolve 0.12g of BSA in 1.2 ml of water resulting a 10% BSA solution. Combine 0.2 ml aliquot of the Na linoleate solution to the 10% BSA solution. After 15 min of slow stirring at 37°C, 0.6 ml of water was added to bring the final concentration of Na linoleate to 5 mMol/L (Pappas et al, 2001).

Lipid analysis

LC-MS grade methanol, dichloromethane, and ammonium acetate were purchased from Fisher Scientific (Pittsburgh, PA) and HPLC grade 1-propanol was purchased from Sigma-Aldrich (Saint Louis, MO). Milli-Q water was obtained from an in-house Ultrapure Water System by EMD Millipore (Billerica, MA). The Lipidyzer isotope labeled internal standards mixture consisting of 54 isotopes from 13 lipid classes was purchased from Sciex (Framingham, MA).

Sample Preparation

Frozen plasma samples were thawed at room temperature (25 °C) for 30 min, vortexed; 25 uL of plasma was transferred to a borosilicate glass culture tube (16 x 100 mm). Next, 0.475 mL of water, 1.45 mL of 1:0.45 methanol:dichloromethane, and 25 uL of the isotope labeled internal standards mixture were added to the tube. The mixture was vortexed for 5 sec and incubated at

491 room temperature for 30 min. Next, another 0.5 mL of water and 0.45 mL of dichloromethane
492 were added to the tube, followed by gentle vortexing for 5 sec, and centrifugation at 2500 g at 15
493 °C for 10 min. The bottom organic layer was transferred to a new tube and 0.9 mL of
494 dichloromethane was added to the original tube for a second extraction. The combined extracts
495 were concentrated under nitrogen and reconstituted in 0.25 mL of the mobile phase (10 mM
496 ammonium acetate in 50:50 methanol:dichloromethane).

497 **Mass Spectrometry**

498 Quantitative lipidomics was performed with the Sciex Lipidyzer platform consisting of Shimadzu
499 Nexera X2 LC-30AD pumps, a Shimadzu Nexera X2 SIL-30AC autosampler, and a Sciex
500 QTRAP® 5500 mass spectrometer equipped with SelexION® for differential mobility
501 spectrometry (DMS). 1-propanol was used as the chemical modifier for the DMS. Samples were
502 introduced to the mass spectrometer by flow injection analysis at 8 uL/min. Each sample was
503 injected twice, once with the DMS on (PC/PE/LPC/LPE/SM), and once with the DMS off
504 (CE/CER/DAG/DCER/FFA/HCER/LCER/TAG). The lipid molecular species were measured
505 using multiple reaction monitoring (MRM) and positive/negative polarity switching. Positive ion
506 mode detected lipid classes SM/DAG/CE/CER/DCER/HCER/DCER/TAG and negative ion mode
507 detected lipid classes LPE/LPC/PC/PE/FFA. A total of 1070 lipids and fatty acids were targeted
508 in the analysis.

509 **Data Processing**:

510 Data was acquired and processed using Analyst 1.6.3 and Lipidomics Workflow Manager 1.0.5.0.
511 For statistical analysis, we evaluated the lipid species enrichments in the ER+, ER-, and control
512 groups. The different groups were compared in pair-wise and the log-fold changes of lipid

enrichment were derived, along with the effect sizes and p-values inferred from the regression models using the lipid measurement as an input variable and group information as the output variable.

Library preparation and RNA Sequencing:

RNA was isolated with Qiagen RNeasy Plus Mini Kit (# 74134) as per the manufacturer's protocol.

The concentration and quality of total RNA in samples were assessed using Agilent 2100 Bioanalyzer. RNA Integrity Number (RIN) of the vehicle and octanoate sample was 9.9 and 9.8 respectively. Sequencing libraries were prepared from a total of 100ng of RNA using KAPA RNA HyperPrep Kit. Single-Indexed adapters were obtained from KAPA (Catalog# KK8701). Library quality was assessed using the KAPA Library Assay kit. Each indexed library was quantified and its quality accessed by Qubit and Agilent Bioanalyzer, and 6 libraries were pooled in equal molarity. 5µL of 4nM pooled libraries were denatured, neutralized and a final concentration of 1.5 pM of pooled libraries was loaded to Illumina NextSeq 500 for 75b single-read sequencing. Approximately 80M filtered reads per library was generated. A Phred quality score (Q score) was used to measure the quality of the sequencing. More than 88% of the sequencing reads reached Q30 (99.9% base call accuracy). Single-end FASTQ reads from RNA-seq measurements were aligned and mapped to hg38 ENSEMBL genome using STAR alignment (*68*).

Gene Ontology Analysis of Differentially Expressed Genes

Transcriptions per million (TPM) from mapped reads were estimated using RSEM from the STAR aligned reads (*69*). The *DESeq2* R package (*70*) was employed to determine differentially expressed genes for the octanoate treatment group compared to the vehicle-treated controls with FDR cutoff = 0.01 and $|\log_2 FC| \geq 2$ to identify a reasonable number of differentially expressed

genes, on the order of several thousands of genes total, for subsequent analysis. Gene ontology pathway analysis for biological processes was performed on each set of differentially expressed genes using *Metascape* (*71*).

GSEA Analysis

Raw counts were first estimated using *HTSeq* from STAR aligned reads *(72)*. Next, replicates for control cells and treated cells were merged and normalized using modules from the GenePattern software package (*73*). Gene set enrichment analysis (GSEA) (*74, 75*) was performed on these DESeq-normalized reads using annotations from online databases, including KEGG, Hallmark, Reactome, BioCarta, and Canonical Pathways. The normalized enrichment score (NES) of these top 20 pathways associated with the control and the octanoate-treated condition are shown with nominal p-value = 0.0. *Metascape* was employed to perform network analysis on these top 20 pathways associated with each treatment condition.

ATAC Seq Library preparation and sequencing

1×10^6 cells were pelleted and lysed in ATAC-resuspension buffer as described (*76*). Extracted nuclei was processed for TN-5 mediated tagmentation using the Illumina Tagment DNA Enzyme and buffer kit (Nextera Illumina # 20034210) : Transposon reaction mix as 2X TD Buffer-25 µl, Tn5 Transposase – 2.5µl, 1X PBS containing nuclei- 16.5µl, 10% Tween-20- 0.5µl (Sigma # P9416), 1% Digitonin-0.5µl (Promega # G9441) and water at 37°C, 1000rpm for 30mins. Tagmented DNA was isolated by Nucleospin PCR clean-up (Takara Bio USA, Inc # 740609.250). Libraries were amplified for 8 cycles and purified using AMPure XP (Agencourt # A63880). Fragment sizes were determined using 106 LabChip GXII Touch HT (PerkinElmer), and 2×50 paired-end sequencing performed on NovaSeq S1 6000 flow cell (Illumina) flow to yield 100M reads per sample.

559 **ATAC-seq data sequencing and peak calling**

560 Illumina adapter sequences and low-quality base calls were trimmed off the paired end reads with
561 Trim Galore v0.4.3. Sequence reads were aligned to human reference genome hg38 using bowtie2
562 with default settings. Duplicate reads were discarded with Picard. Reads mapped to mitochondrial
563 DNA together with low mapping quality reads were excluded from further analysis. MACS2 was
564 used to identify the peak regions with options -f BAMPE -g hs –keep-dup all -B -q 0.01. Peaks for
565 samples in the same condition were merged using the function 'merge' of bedtools and peaks for
566 samples in different conditions were intersected using the function of 'intersect' of bedtools.

567 **Differential chromatin accessibility analysis**

568 The number of cutting sites of each samples were counted using the script dnase_cut_counter.py
569 of pyDNase. The raw count matrix was normalized by CPM. R package edgeR was used to conduct
570 the differential accessibility analysis for all 66,853 common peaks. Significant different accessible
571 chromatin regions under different conditions were defined as the threshold 0.05 for FDR. With the
572 cutoff 1 for the absolute value of fold change, comparing treatment group with vehicle control
573 group, we got 1,704 significant increased peaks and 3,340 significant decreased peaks.

574 **Motif analysis**

575 Motif analysis were conducted for significant changed chromatin regions using
576 'findMotifsGenome.pl' script of HOMER with default settings. The principal component analysis
577 was conducted to detect the important motifs using the relative enrichment of motifs. Biplot was
578 used to visualize the principal component analysis results.

579 **Genomic distribution of open chromatin regions**

We calculated the overall genomic distribution of open chromatin regions, comparing the treatment to the vehicle, based on the methods as described (*77*). We used the hg38 refseq genes annotation from UCSC genome browser to define the genomic features. All TSSs were considered in the analysis if a gene had multiple TSSs. The formula for reported enrichment is (a/b)/(c/d). a is the number of peaks overlapping a given genomic feature, b is the number of total peaks, c is the number of regions corresponding to the feature, and d is the estimated number of discrete regions in the genome where the peaks and feature could overlap. Specifically, d is equal to (genome size)/ (mean peak size + mean feature size), following the implementation in the bedtools fisher.

Pathway analysis for open chromatin regions

For the 326 open chromatin regions with logFC $\geq$ 1.5 and FDR < 0.05 comparing the treatment with the vehicle, we used R package 'clusterProfile' to conduct KEGG pathway analysis.

Validation of candidate genes qRT-PCR:

Treated cells and organoids were washed with PBS and RNA was isolated with Qiagen RNeasy plus mini Kit (# 74134) as per the manufacturer's protocol. cDNA was synthesized using the SuperScript VILO cDNA synthesis kit (#11755250). Real-time qPCR was performed using Applied biosystem Quant studio 5 real time PCR System (Thermo Scientific). Expression data of the studied genes was normalized to RPLP1 to control the variability in expression levels and were analyzed using the $2^{-\Delta\Delta CT}$ method described by Livak and Schmittgen (*78*). TaqMan gene expression assays were purchased from ThermoFisher Scientific and the list of the assays is provided in supplemental file **S1**.

qRT-PCR based TaqMan low density array assays

Based on histological diagnosis atypical hyperplasia benign breast epithelium was identified and captured by laser capture microdissection (LCM). RNA was isolated with Qiagen RNeasy plus mini Kit (# 74134) as per the manufacturer's protocol. RNA quality was checked for integrity using Bioanalyzer 2100 by Agilent. 100ng RNA was reverse transcribed using High Capacity RNA-to-cDNA Master Mix (#4388950) and preamplified for 14 cycles using TaqMan PreAmp Master Mix 2X((#4488593) and pooled assay mix for the genes in which we were interested. Pre-amplified cDNA were diluted to 1:20 with 1X TE buffer and mixed with Fast advanced master mix (# 4444965) Each sample was loaded in duplicate in 384- well microfluidic cards customized with 47 genes of interest including three housekeeping genes (GAPDH, RPLP0 and RPLP1). TaqMan assays with best coverage attribution were used for the TLDA study as recommended by the manufacturer. A list of the genes and the Assay ID for the primers obtained from ThermoFisher is provided in supplemental file S2. Real Time PCR reactions were carried out in Quant studio 7 Flex system for 40 cycles using comparative Ct (ΔΔCt) method. Results were analyzed using Expression suite software.

Live cell PWS Imaging:

Before treatment and imaging, MCF-10A cells were seeded in 6 wells black culture plate at least 35% confluency and allowed to adhere overnight before the treatment with 500µM LA (C18:2) and 5mM Octanoate. We based the concentration of LA used in the experiment on the range in human plasma: 0.2 to 5.0 mmol/L (*79*). For chromatin study experiments, live-cell PWS images were acquired at room temperature (22 °C) and in trace CO2 (open air) conditions. Imaging was performed using the commercial inverted microscope (Leica DMIRB) Hamamatsu Image-EM CCD camera C9100-13 coupled to a liquid crystal tunable filter (LCTF; CRi Woburn, MA) to acquire mono-chromatic spectrally resolved images that range from 500–700 nm at 1 nm intervals

produced by a broad band illumination provided by an Xcite-120 LED Lamp (Excelitas, Waltham, MA) as previously described (33, 34). Briefly, PWS measures the spectral interference resulting from internal light scattering structures within the cell, which captures the mass density distribution. To obtain the interference signal directly related to refractive index fluctuations in the cell, we normalized measurements by the reflectance of the glass medium interface, i.e., to an independent reference measurement acquired in an area without cells. PWS measures a data cube (spatial coordinates of a location within a cell and the light interference spectrum recorded from this location). The data cube then allow to measure spectral SD (Σ), which is related to the spatial variations of refractive index within a given coherence volume. The coherence volume was defined by the spatial coherence in the transverse directions (~200 nm) and the depth of field in the axial direction (~1 mm). In turn, the spatial variations of refractive index depended on the local autocorrelation function (ACF) of the chromatin refractive index. Finite-difference time-domain simulations have shown that PWS is sensitive to ACF within the 20- to 200-nm range. According to the Gladstone-Dale equation, refractive index is a linear function of local molecular crowding. Therefore, S depends on the ACF of the medium's macromolecular mass density. Small molecules and other mobile crowders within the nucleus are below the limit of sensitivity of PWS, and PWS is primarily sensitive to chromatin conformation above the level of the nucleosome. To convert S for a given location within a nucleus to mass fractal dimension D, we modeled ACF as a power law B¥(r)= ðrÞ ¼ s2ϕrrmi_ _D_3, where ϕ is the variance of CVC (60). In general, S is a sigmoidal function of D. However, for fractal structures such as a chromatin packing domain where within physiological range $2 < D < 3$, S can be approximated as a linear function of D by the relationship $D \approx D0 + aS$, where $D0 = 1.473$ and is comparable to the minimal fractal dimension that an unconstrained polymer can attain and constant $a \sim 7.6$. The measured change in chromatin packing

scaling between treatment conditions was quantified by first averaging D within each cell's nucleus and then averaging nuclei from over 100 cells per condition.

Flux based analysis (FBA)

We calculated the relative activity of reactions in MCF-10A cells by interpreting gene expression data using the Recon1 human metabolic model augmented with histone modifications (*46, 80*). We then identified a metabolic flux state that is most consistent with gene expression data in control and octanoate treatment. This was achieved by maximizing the activity of reactions that are associated with up-regulated genes and minimizing flux through reactions that are down-regulated in a condition, while simultaneously satisfying the stoichiometric and thermodynamic constraints embedded in the model using linear optimization (*46, 80*). The glucose, fatty acid, and glutamine levels in the simulations were adjusted based on the growth media used for culturing the cells. All p-values were corrected for multiple comparisons.

Statistical analysis

Prior to performing the analyses, the log2-transformed relative (log2RE) amounts of mRNA expression normalized to GAPDH and expressed as $\log_2 2^{-(CtX- CtGAPDH)} = -(CtX- CtGAPDH)$ where Ct is threshold cycle. Mann-Whitney test was performed to identify genes with pairwise differences between ER+ and ER- samples. The analyses were adjusted for multiple testing, 34 genes, using the Benjamini-Hochberg (BH) adjustment in order to control the false discovery rate at the two-sided 0.05 level. Boxplots were used to visualize differences in log2RE by group. The log2RE analyses were conducted using the R statistical environment [R] version 3.5.1.

Supplementary Materials

S1 List of primers used for qRT-PCR validation of candidate genes. List of Assay ID (ThermoFisher) of primers utilized in qRT-PCR of candidate genes (Fig. 4A).

S2 List of genes assayed by TaqMan low density array (TLDA) and their corresponding primers. List of genes and the corresponding Assay ID (ThermoFisher) of primers used for TLDA assays.

References

1. B. Weigelt, J. S. Reis-Filho, Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nature reviews. Clinical oncology* **6**, 718-730 (2009).
2. M. H. Gail *et al.*, Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* **81**, 1879-1886 (1989).
3. B. Fisher *et al.*, Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* **90**, 1371-1388 (1998).
4. J. Cuzick *et al.*, Long-term results of tamoxifen prophylaxis for breast cancer--96-month follow-up of the randomized IBIS-I trial. *J Natl Cancer Inst* **99**, 272-282 (2007).
5. J. Cuzick *et al.*, Use of anastrozole for breast cancer prevention (IBIS-II): long-term results of a randomised controlled trial. *Lancet* **395**, 117-122 (2020).
6. C. Bouchardy *et al.*, Risk of second breast cancer according to estrogen receptor status and family history. *Breast Cancer Res Treat* **127**, 233-241 (2011).
7. A. W. Kurian *et al.*, Second primary breast cancer occurrence according to hormone receptor status. *J Natl Cancer Inst* **101**, 1058-1065 (2009).
8. S. M. Swain *et al.*, Estrogen receptor status of primary breast cancer is predictive of estrogen receptor status of contralateral breast cancer. *J Natl Cancer Inst* **96**, 516-523 (2004).
9. J. Wang *et al.*, Lipid metabolism genes in contralateral unaffected breast and estrogen receptor status of breast cancer. *Cancer prevention research (Philadelphia, Pa.)* **6**, 321-330 (2013).
10. J. Wang *et al.*, Overexpression of lipid metabolism genes and PBX1 in the contralateral breasts of women with estrogen receptor-negative breast cancer. *International journal of cancer* **140**, 2484-2497 (2017).
11. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
12. W. H. Koppenol, P. L. Bounds, C. V. Dang, Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer* **11**, 325-337 (2011).
13. A. Carracedo, L. C. Cantley, P. P. Pandolfi, Cancer metabolism: fatty acid oxidation in the limelight. *Nat Rev Cancer* **13**, 227-232 (2013).
14. K. M. Nieman *et al.*, Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nature medicine* **17**, 1498-1503 (2011).

15. R. Mitra, T. T. Le, P. Gorjala, O. B. Goodman, Jr., Positive regulation of prostate cancer cell growth by lipid droplet forming and processing enzymes DGAT1 and ABHD5. *BMC cancer* **17**, 631 (2017).
16. K. B. Singh *et al.*, Prostate cancer chemoprevention by sulforaphane in a preclinical mouse model is associated with inhibition of fatty acid metabolism. *Carcinogenesis* **39**, 826-837 (2018).
17. Huang *et al.*, HIF-1-mediated suppression of acyl-CoA dehydrogenases and fatty acid oxidation is critical for cancer progression. *Cell reports* **8**, 1930-1942 (2014).
18. R. Camarda *et al.*, Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. *Nature medicine* **22**, 427-432 (2016).
19. J. H. Park *et al.*, Fatty Acid Oxidation-Driven Src Links Mitochondrial Energy Reprogramming and Oncogenic Properties in Triple-Negative Breast Cancer. *Cell reports* **14**, 2154-2165 (2016).
20. K. M. Havas *et al.*, Metabolic shifts in residual breast cancer drive tumor recurrence. *J Clin Invest* **127**, 2091-2105 (2017).
21. V. Cappelletti *et al.*, Metabolic Footprints and Molecular Subtypes in Breast Cancer. *Dis Markers* **2017**, 7687851 (2017).
22. S. You *et al.*, Raman Spectroscopic Analysis Reveals Abnormal Fatty Acid Composition in Tumor Micro- and Macroenvironments in Human Breast and Rat Mammary Cancer. *Sci Rep* **6**, 32922 (2016).
23. M. A. Reid, Z. Dai, J. W. Locasale, The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat Cell Biol* **19**, 1298-1306 (2017).
24. E. McDonnell *et al.*, Lipids Reprogram Metabolism to Become a Major Carbon Source for Histone Acetylation. *Cell reports* **17**, 1463-1472 (2016).
25. M. Grabacka, M. Pierzchalska, M. Dean, K. Reiss, Regulation of Ketone Body Metabolism and the Role of PPARalpha. *International journal of molecular sciences* **17**, (2016).
26. P. Schönfeld, L. Wojtczak, Short- and medium-chain fatty acids in energy metabolism: the cellular perspective. *Journal of lipid research* **57**, 943-954 (2016).
27. R. K. A. Virk *et al.*, Disordered chromatin packing regulates phenotypic plasticity. *Sci Adv* **6**, eaax6232 (2020).
28. S. Yadav *et al.*, Abstract P2-02-06: Deregulated lipid metabolism fuels the genesis of estrogen receptor negative breast cancer. *Cancer Research* **79**, P2-02-06 (2019).
29. J. Whelan, K. Fritsche, Linoleic acid. *Adv Nutr* **4**, 311-312 (2013).
30. D. Ahel *et al.*, Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325**, 1240-1243 (2009).
31. S. J. Guyenet, S. E. Carlson, Increase in adipose tissue linoleic acid of US adults in the last half century. *Adv Nutr* **6**, 660-664 (2015).
32. L. M. Almassalha *et al.*, Macrogenomic engineering via modulation of the scaling of chromatin packing density. *Nature biomedical engineering* **1**, 902-913 (2017).
33. L. M. Almassalha *et al.*, The Global Relationship between Chromatin Physical Topology, Fractal Structure, and Gene Expression. *Scientific reports* **7**, 41061-41061 (2017).
34. L. M. Almassalha *et al.*, The Greater Genomic Landscape: The Heterogeneous Evolution of Cancer. *Cancer Res* **76**, 5605-5609 (2016).
35. S. Yadav *et al.*, Abstract P3-02-07: Fatty acid metabolism is associated with chromatin remodeling in mammary epithelial cells. *Cancer Research* **80**, P3-02-07 (2020).

36. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
37. S. Steensels, B. A. Ersoy, Fatty acid activation in thermogenic adipose tissue. *Biochimica et biophysica acta. Molecular and cell biology of lipids* **1864**, 79-90 (2019).
38. E. J. O'Brien, J. M. Monk, B. O. Palsson, Using Genome-scale Models to Predict Biological Capabilities. *Cell* **161**, 971-987 (2015).
39. J. D. Orth, I. Thiele, B. O. Palsson, What is flux balance analysis? *Nat Biotechnol* **28**, 245-248 (2010).
40. S. Chandrasekaran *et al.*, Comprehensive Mapping of Pluripotent Stem Cell Metabolism Using Dynamic Genome-Scale Network Modeling. *Cell reports* **21**, 2965-2977 (2017).
41. C. Frezza *et al.*, Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225-228 (2011).
42. A. Mardinoglu *et al.*, Personal model-assisted identification of NAD(+) and glutathione metabolism as intervention target in NAFLD. *Mol Syst Biol* **13**, 916 (2017).
43. T. Shlomi, M. N. Cabili, M. J. Herrgard, B. O. Palsson, E. Ruppin, Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* **26**, 1003-1010 (2008).
44. K. Yizhak, B. Chaneton, E. Gottlieb, E. Ruppin, Modeling cancer metabolism on a genome scale. *Mol Syst Biol* **11**, 817 (2015).
45. N. C. Duarte *et al.*, Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* **104**, 1777-1782 (2007).
46. F. Shen, L. Boccuto, R. Pauly, S. Srikanth, S. Chandrasekaran, Genome-scale network model of metabolism and histone acetylation reveals metabolic dependencies of histone deacetylase inhibitors. *Genome Biol* **20**, 49 (2019).
47. A. Carrer *et al.*, Impact of a High-fat Diet on Tissue Acyl-CoA and Histone Acetylation Levels. *J Biol Chem* **292**, 3312-3322 (2017).
48. J. T. Li *et al.*, BCAT2-mediated BCAA catabolism is critical for development of pancreatic ductal adenocarcinoma. *Nat Cell Biol* **22**, 167-174 (2020).
49. S. J. Mentch *et al.*, Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism. *Cell Metab* **22**, 861-873 (2015).
50. K. Kerlikowske *et al.*, Risk Factors That Increase Risk of Estrogen Receptor-Positive and -Negative Breast Cancer. *J Natl Cancer Inst* **109**, (2017).
51. K. Li *et al.*, Risk prediction for estrogen receptor-specific breast cancers in two large prospective cohorts. *Breast Cancer Res* **20**, 147 (2018).
52. M. J. Worsham *et al.*, Multiplicity of benign breast lesions is a risk factor for progression to breast cancer. *Clin Cancer Res* **13**, 5474-5479 (2007).
53. O. Alonzo-Proulx, J. G. Mainprize, J. A. Harvey, M. J. Yaffe, Investigating the feasibility of stratified breast cancer screening using a masking risk predictor. *Breast Cancer Res* **21**, 91 (2019).
54. A. S. Reiner *et al.*, Hormone receptor status of a first primary breast cancer predicts contralateral breast cancer risk in the WECARE study population. *Breast Cancer Res* **19**, 83 (2017).
55. H. W. Smith *et al.*, An ErbB2/c-Src axis links bioenergetics with PRC2 translation to drive epigenetic reprogramming and mammary tumorigenesis. *Nat Commun* **10**, 2901 (2019).
56. M. E. Gonzalez *et al.*, Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. *Oncogene* **28**, 843-853 (2009).

57. J. Wu, D. L. Crowe, The histone methyltransferase EZH2 promotes mammary stem and luminal progenitor cell expansion, metastasis and inhibits estrogen receptor-positive cellular differentiation in a model of basal breast cancer. *Oncol Rep* **34**, 455-460 (2015).
58. M. E. Gonzalez *et al.*, EZH2 expands breast stem cells through activation of NOTCH1 signaling. *Proc Natl Acad Sci U S A* **111**, 3098-3103 (2014).
59. G. Dontu *et al.*, Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells. *Breast Cancer Res* **6**, R605-615 (2004).
60. V. Rodilla *et al.*, Luminal progenitors restrict their lineage potential during mammary gland development. *PLoS Biol* **13**, e1002069 (2015).
61. A. M. Lilja *et al.*, Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nat Cell Biol* **20**, 677-687 (2018).
62. R. L. Milne *et al.*, Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* **49**, 1767-1778 (2017).
63. Y. Li, T. J. Dillon, M. Takahashi, K. T. Earley, P. J. Stork, Protein Kinase A-independent Ras Protein Activation Cooperates with Rap1 Protein to Mediate Activation of the Extracellular Signal-regulated Kinases (ERK) by cAMP. *J Biol Chem* **291**, 21584-21595 (2016).
64. C. J. Creighton *et al.*, Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors. *Cancer Res* **66**, 3903-3911 (2006).
65. Z. T. Schafer *et al.*, Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109-113 (2009).
66. E. Piskounova *et al.*, Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186-191 (2015).
67. T. Wang *et al.*, JAK/STAT3-Regulated Fatty Acid beta-Oxidation Is Critical for Breast Cancer Stem Cell Self-Renewal and Chemoresistance. *Cell Metab* **27**, 136-150 e135 (2018).
68. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
69. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
70. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
71. Y. Zhou *et al.*, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
72. S. Anders, P. T. Pyl, W. Huber, HTSeq – A Python framework to work with high-throughput sequencing data. *bioRxiv*, 002824 (2014).
73. M. Reich *et al.*, GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006).
74. V. K. Mootha *et al.*, PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267-273 (2003).
75. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545 (2005).
76. A. Ackermann, <ATAC-seqProtocolOmni-KaestnerLab.pdf>. (2019).

77. K. Miyamoto *et al.*, Chromatin Accessibility Impacts Transcriptional Reprogramming in Oocytes. *Cell reports* **24**, 304-311 (2018).
78. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).
79. S. A. Abdelmagid *et al.*, Comprehensive profiling of plasma fatty acid concentrations in young healthy Canadian adults. *PloS one* **10**, e0116195 (2015).
80. F. Shen, C. Cheek, S. Chandrasekaran, Dynamic Network Modeling of Stem Cell Metabolism. *Methods Mol Biol* **1975**, 305-320 (2019).

Acknowledgments

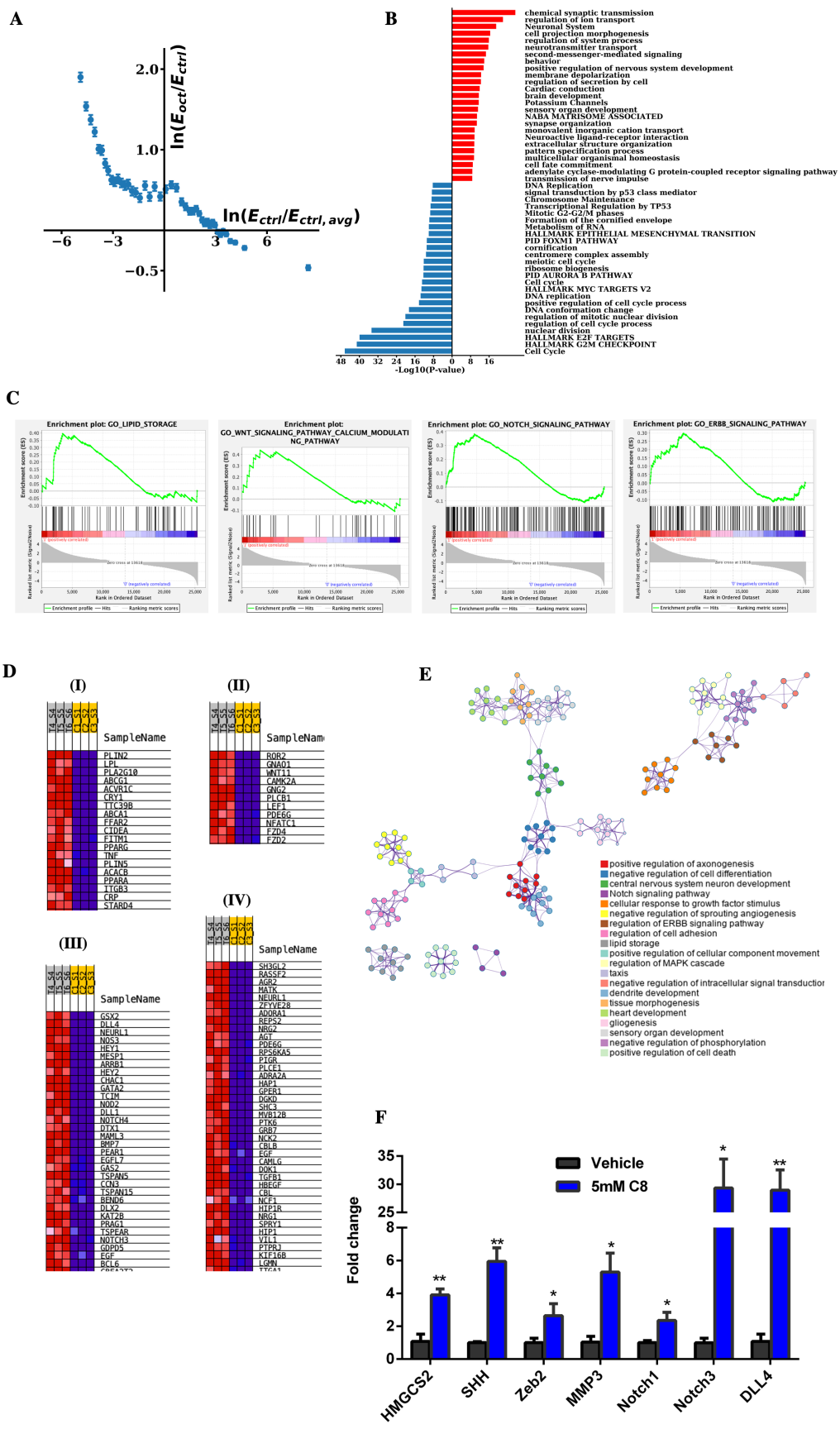
General: We would like to thank Professors Matthew D Hirschey and Neil Kelleher for advice regarding histone proteomics and, Jeannie Camarillo and the Northwestern Proteomics Core for conducting the histone proteomic analysis; The Northwest Metabolic Research Center (NW-MRC) at University of Washington for performing the lipidomics analysis; The Center for Medical Genomics at the Indiana University School of Medicine for RNA library preparation and RNA sequencing, and ATAC sequencing; The NU Seq Core facility for providing The Quant Studio 7 Flex system; Natalie Pulliam for consenting patients and collecting tissue, and our many lab colleagues for feedback.

Funding: This work was supported by The Breast Cancer Research Foundation and the Bramsen-Hamill Foundation (S.A.K), and NIH grant # 1S10OD021562-01 (NW-MRC).

Author contributions: Conceptualization and project design: S.Y., S.E.C., and S.A.K.; experiments: S.Y ; Sequencing: X.X and H.G; RNA seq analysis : R.K.A.V., Z.Z., G.C and M.R.; ATAC Seq analysis: D.C.; PWS microscopy: D.V.D.; Metabolic flux analysis: C.H.C and S.C; Statistical Analysis: K.L.B ; writing (original draft): S.Y., S.E.C., and S.A.K.; writing (review and editing): V.B., S.C., and R.C.

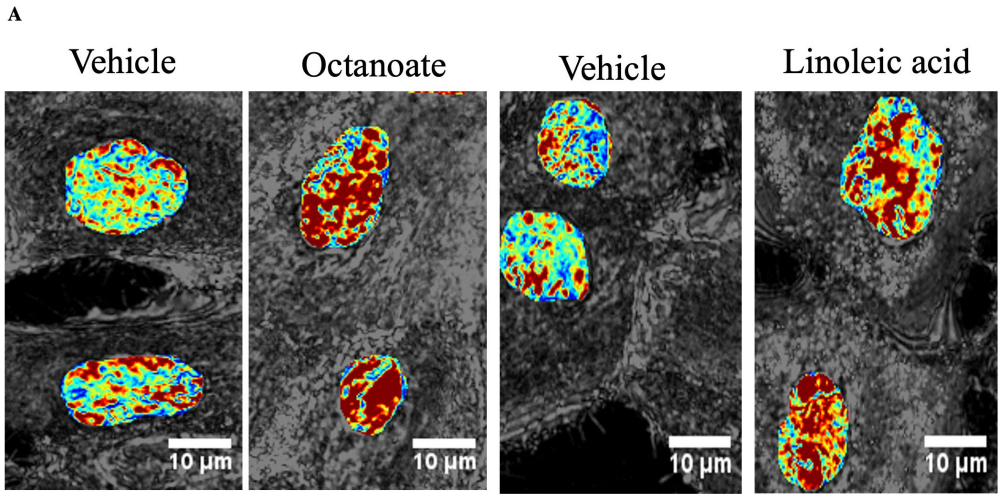
Competing interests: The authors declare that they have no conflict of interests.

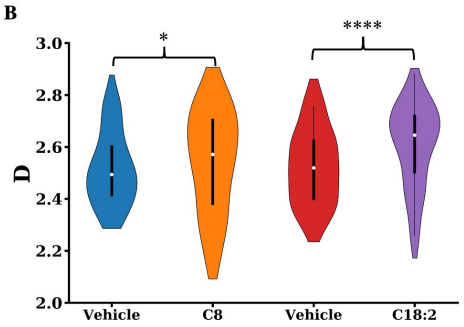
868 **Data and materials availability**: The datasets generated and analyzed during the current study
869 are publicly available in the Gene Expression Omnibus: accession number GSE126799 (RNA-
870 seq) and XXX (ATAC-seq)

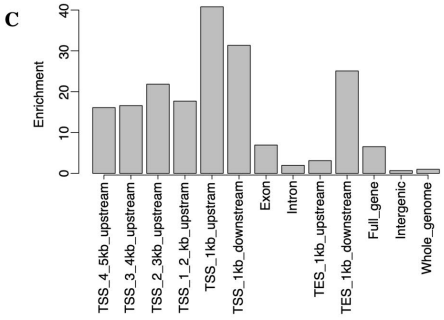


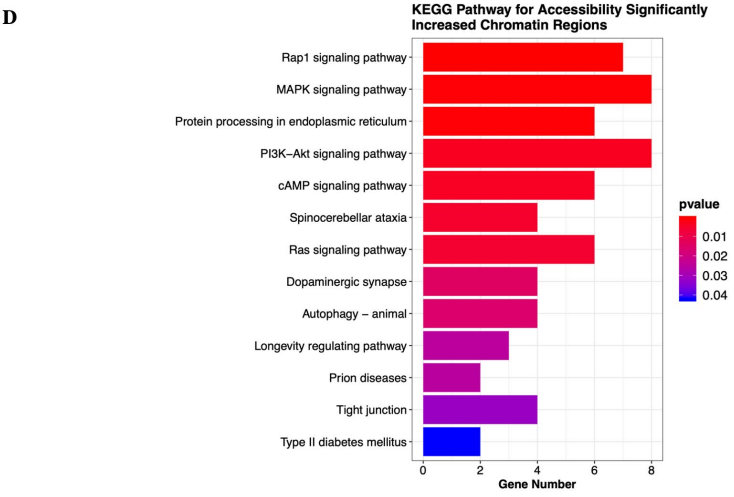
872 **Fig. 1. Lipid-rich environment enables transcriptional reprogramming in mammary**
873 **epithelial cells.** (A) 24-hour treatment of MCF10A cells with 5mM octanoate results in a
874 completely distinct transcriptional profile compared to untreated controls. E_{ctrl} is the expression
875 of genes in the control condition across all 3 control replicates, $E_{ctrl,avg}$ is the average expression
876 for the control condition across all genes and replicates, E_{oct} is the expression of genes across all
877 3 octanoate replicates. $E_{ctrl}/E_{ctrl,avg}$ represents the ratio of expression of a particular gene to the
878 average expression across all control cells. Thus, a positive value of $\ln(\frac{E_{ctrl}}{E_{ctrl,avg}})$ corresponds to
879 genes that are highly expressed in the control conditions while a negative value of $\ln(\frac{E_{ctrl}}{E_{ctrl,avg}})$
880 corresponds to genes that have an initial lower expression in the control condition. E_{oct}/E_{ctrl}
881 represents the ratio of expression of a particular gene for octanoate-treated versus vehicle control-
882 treated cells. Genes with initially low expression are upregulated while genes with initially high
883 expression are downregulated upon octanoate treatment. (B) Gene ontology analysis of
884 differentially expressed genes induced by octanoate treatment. Upregulated and downregulated
885 genes were first identified using DESeq2 (FDR < 0.01, |logFC| = 2) for 5mM octanoate treated
886 cells compared to vehicle-treated control cells. Pathway enrichment analysis was performed on
887 identified differentially expressed genes with annotations from online pathway databases (KEGG,
888 Hallmark, Canonical Pathways, Reactome, BioCarta) and Gene Ontology Biological Processes.
889 Pathway enrichment was ranked by p-value on a -Log_{10} scale and a selection from the top 25
890 pathways associated with upregulated genes (in red) and downregulated genes (in blue) are shown.
891 (C) GSEA analysis of Gene Ontology Biological Processes showing top pathways associated with
892 octanoate treatment with FDR < 0.1 related to differentiation, cell signaling, and metabolic
893 processes. (D) List of core enrichment genes differentially expressed in treated replicates -T4, T5,
894 T6 versus control replicates- C1, C2, C3 (I) Lipid storage pathways (II) Wnt pathway (III) Notch

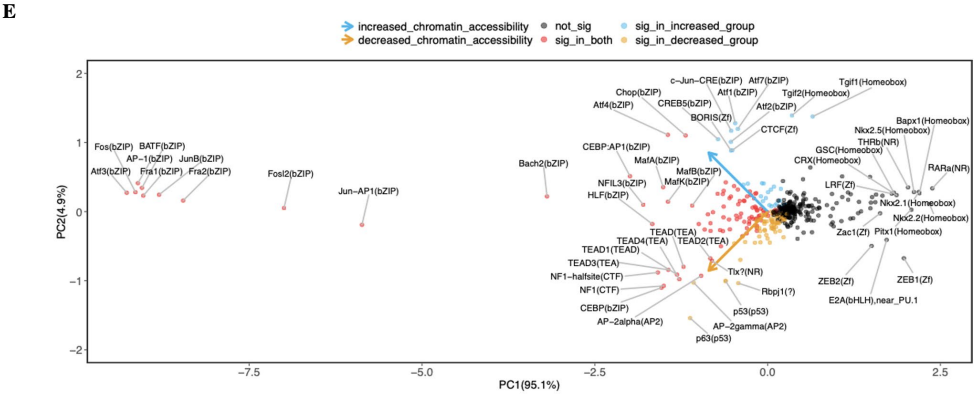
pathway (IV) ERBB pathway each pathway as identified by GSEA leading edge analysis. (E) Network analysis of pathways associated with the octanoate phenotype in GSEA analysis of Gene Ontology Biological Processes. (F) qPCR analysis of genes associated with the NOTCH pathway. Two genes, NOTCH3 and DLL4 show significant upregulation upon 5mM octanoate treatment compared to other identified genes such as NOTCH1. Statistical significance was determined by the unpaired t-test with Welch's correction.











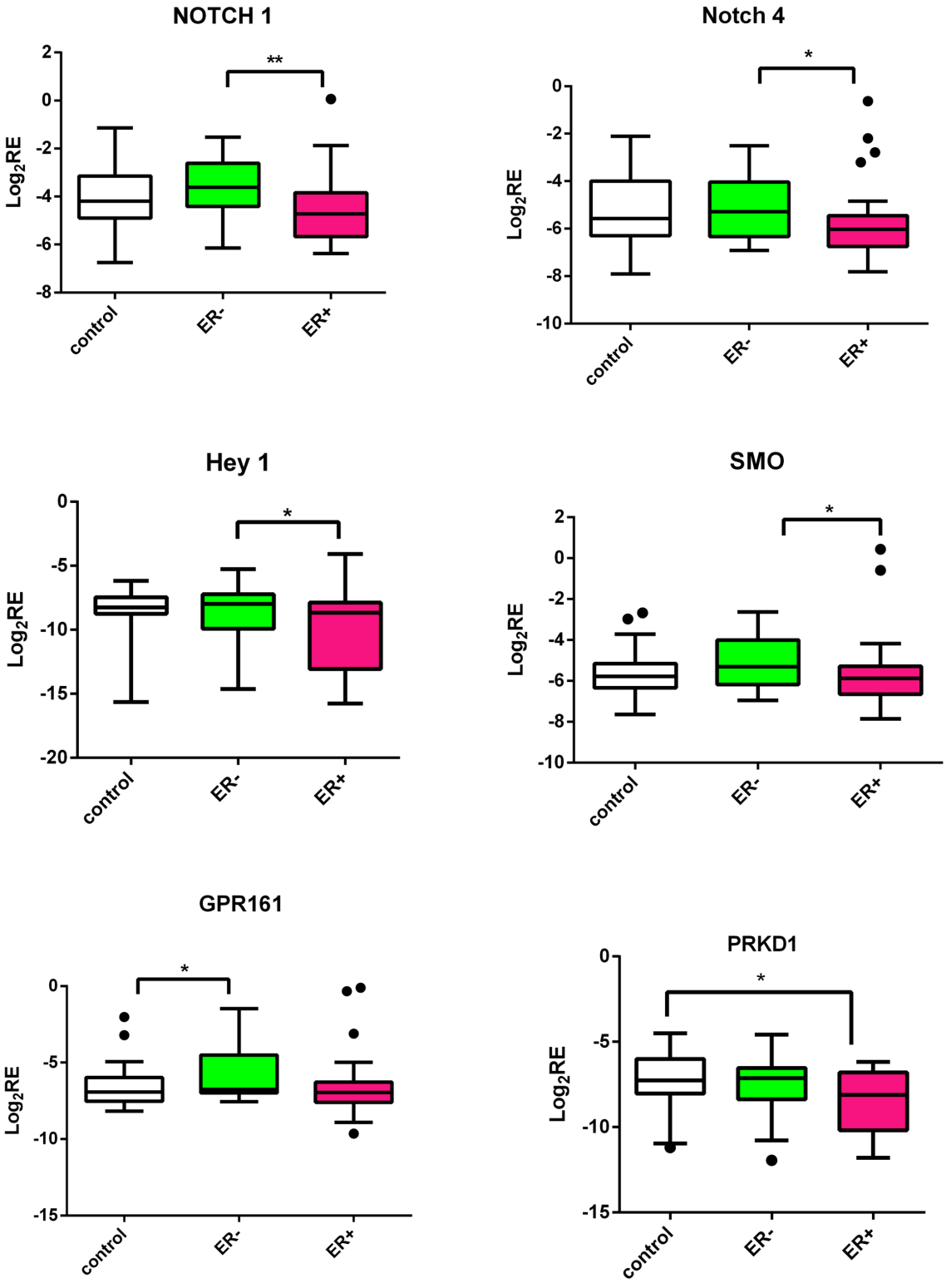
902 **Fig. 2. Linoleic acid alters large-scale chromatin packing behavior in MCF-10A cells.**

903 (A) Representative PWS microscopy images of MCF-10A cell nuclei at 24 hours after
904 treatment with vehicle controls and lipids – octanoate and linoleic acid. Scale bars, 10μm.
905 Chromatin packing scaling *(D)* map of nuclei shows an increase in chromatin packing
906 scaling upon lipid treatment as demonstrated by an increase in red regions.

907 (B) Changes in average chromatin packing scaling among MCF-10A cells upon treatment with
908 vehicle controls and lipids compared to untreated cells. Significance was determined using
909 unpaired Kolmogorov-Smirnov *t*-test (****$P < 0.0001$, *$P < 0.05$). Bar graphs show the
910 mean change in intranuclear *D* across cell populations for N = 88 cells PBS (vehicle for
911 octanoate), N = 110 cells Octanoate (C8), N = 103 cells BSA (vehicle for linoleic acid),
912 and N = 94 Linoleic acid (C18:2). (C) Enrichment of genomic locations for 1704 open
913 chromatin regions (FDR < 0.05, logFC > 1) in LA treated MCF-10A cells. The enrichment
914 of peaks in each type of genomic region relative to the whole genome is shown on the y-
915 axis. Two ATAC-seq libraries were used for the analysis. (D) Pathway analysis for the
916 regions with increased chromatin accessibility in linoleic acid treated cells identified using
917 the KEGG database. (E) Biplot showing changes in chromatin accessibility for specific
918 regions identified by HOMER analysis. Motifs with a significant increase in the chromatin
919 accessibility are shown in blue and those with a significant decrease in accessibility are
920 shown in yellow (FDR<0.05 and |logFC| > 1).

921

922



923

924

925 **Fig. 3. Notch pathway is overexpressed in CUB samples of patients at high risk of ER- disease**

926 Expression of genes from various pathways in matching CUBs from ER negative, ER positive
927 patients and controls. The log2-transformed relative (log2RE) amounts of mRNA expression
928 normalized to the housekeeping gene and expressed as $\log_2 2^{-(CtX-CtGAPDH)} = -(CtX - Ct\ GAPDH)$
929 where Ct is threshold cycle and X is gene of interest. IGF2 and GPR161 were significantly higher
930 in ER negative versus normal whereas ER positive showed significant increase in PRKD1 versus
931 normal. Genes from the Notch pathway were significantly higher in ER negative CUBs in
932 comparison to ER positive patients. Mann-Whitney test was used to test the pairwise differences
933 between the samples (ER+, ER-, Control) * $p < 0.05$; ** $p < 0.01$.

934

935

936

937

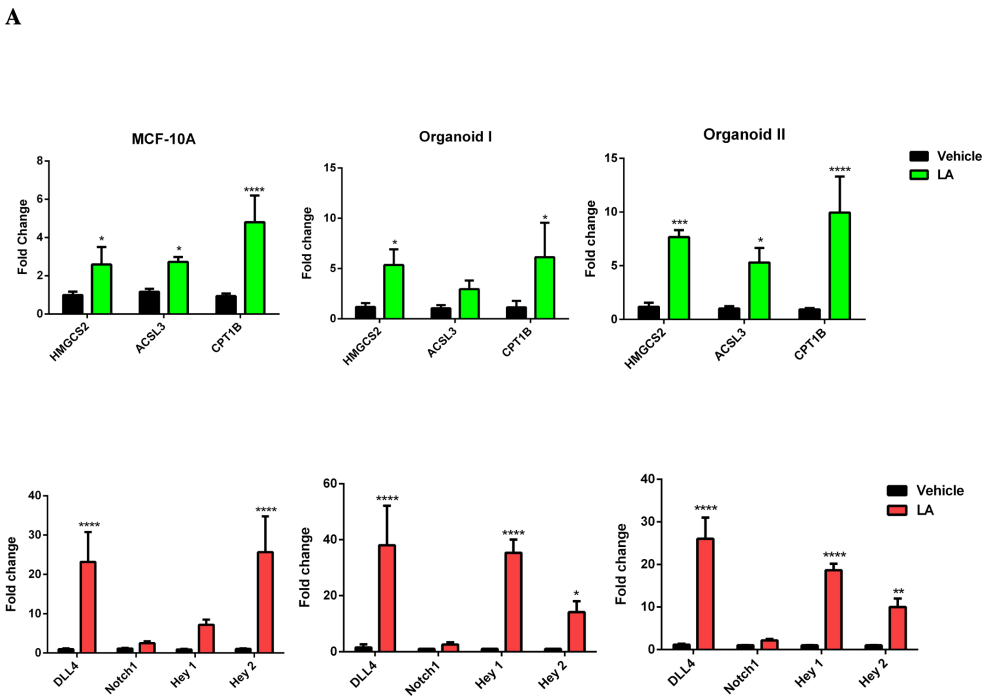
938

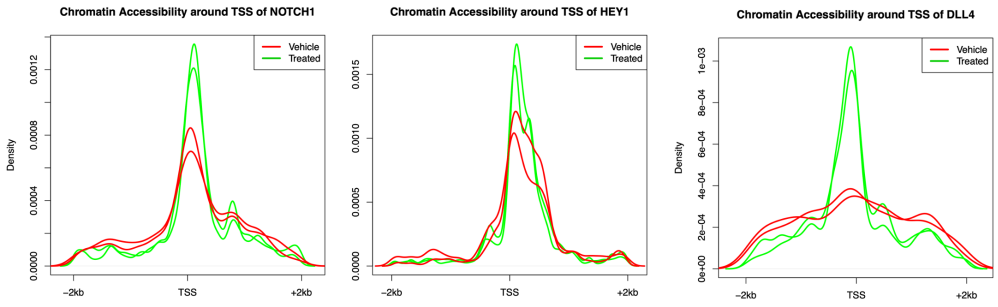
939

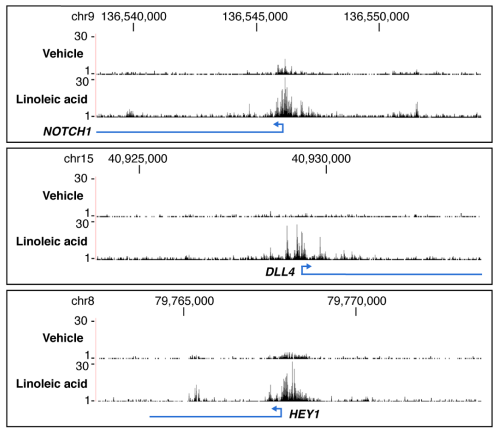
940

941

942







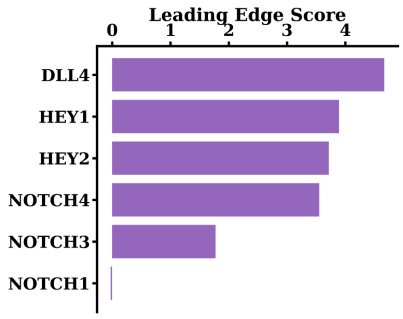
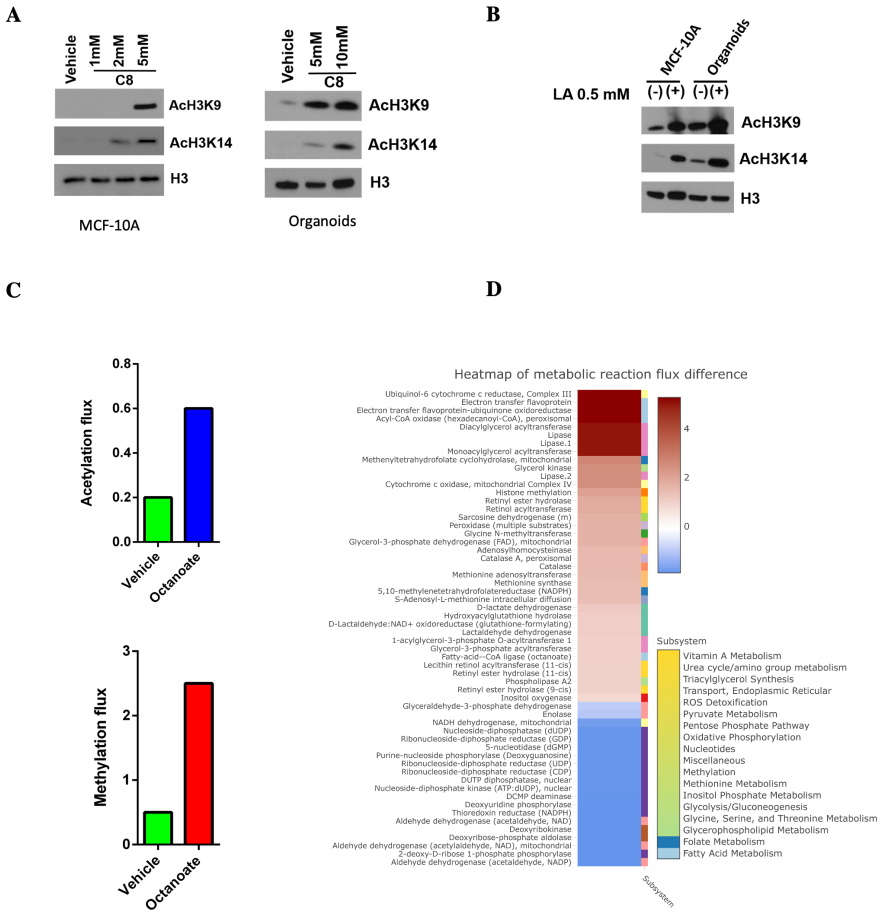


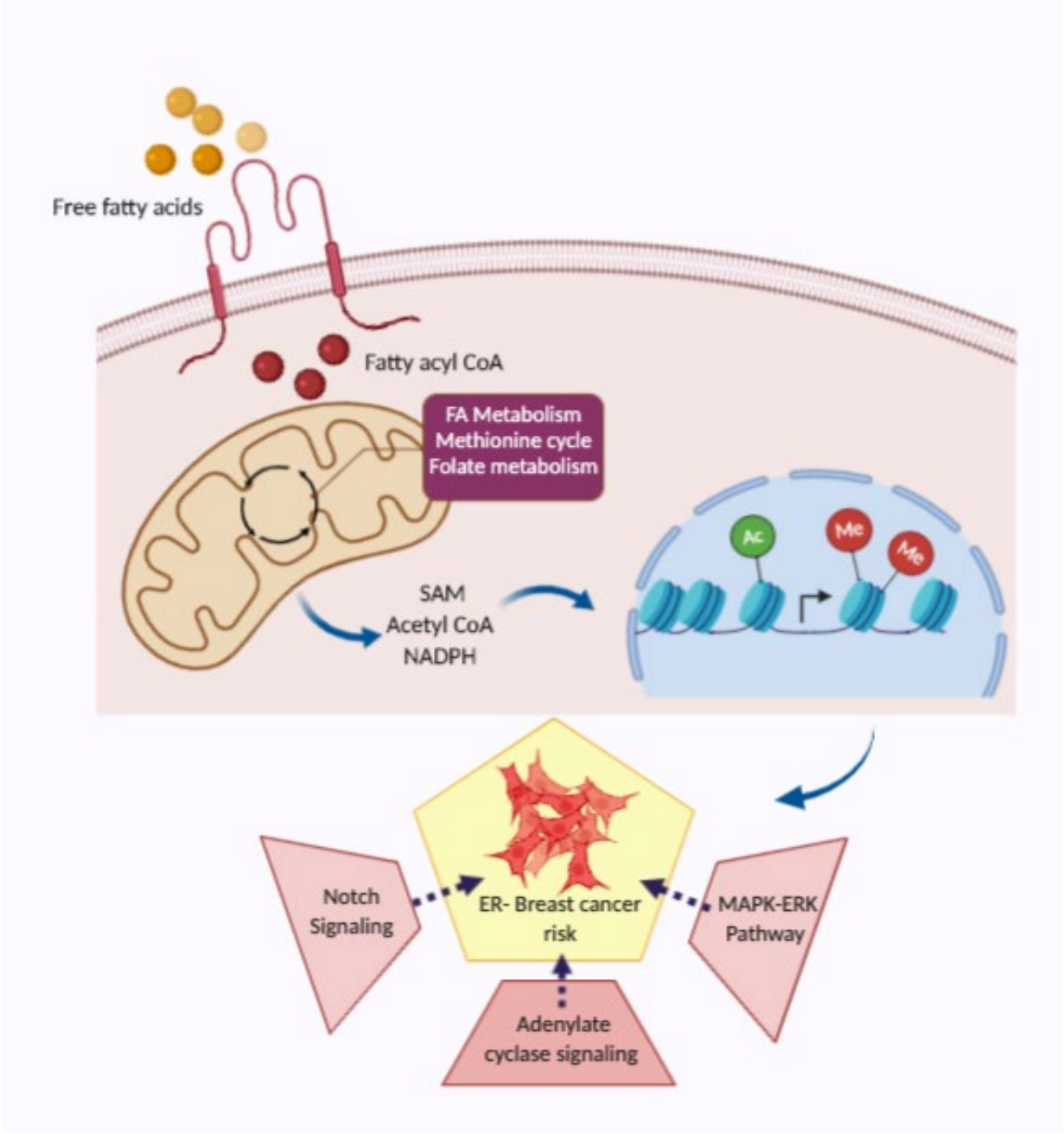
Fig. 4. Increased DLL4/Notch signaling is associated with the stimulated fatty acid oxidation

(**A**) qPCR data showing increase in lipid metabolism genes (green) and Notch pathway genes (red) after 24-hour linoleate treatment in MCF-10A and mammary organoids. Statistical significance was determined by the unpaired t-test with Welch's correction. (**B**) Chromatin accessibility in the lipid treated cells around the transcription start site (TSS) of NOTCH1, HEY1 and DLL4 (FDR<0.001). (**C**) Gene tracks and increase in peaks for the Notch genes in LA treated cells with the exact location on the chromosome. (**D**) Leading edge scores for genes of interest associated with the NOTCH signaling pathway as determined by GSEA leading edge analysis. DLL4, HEY1, HEY2, NOTCH3, and NOTCH4 were identified as core enrichment genes in the NOTCH pathway.



955 **Fig. 5**. **Fatty acids drive histone modifications and metabolic flux** Western blot of histone
956 acetylation at H3 lysine K9 and K14 in MCF-10A cells and organoids treated with (A) octanoate
957 and (B) linoleic acid. **(C)** The effect of octanoate treatment on histone acetylation and methylation
958 flux in MCF-10A cells predicted using genome-scale metabolic modeling. (D) Heatmap of
959 reaction flux differences predicted by metabolic modeling to be differentially active (p-value <
960 0.01) between control and treatment. The corresponding pathways (subsystem) that each reaction
961 belongs to is listed in the legend. Proteomic acetylation (E) and methylation (F) profiling measured
962 by mass spectrometry of MCF-10A cells treated in triplicate with 5mM octanoate for 24 hours in
963 a complete media compared to vehicle. Two-way ANOVA was performed to determine the
964 statistical significance and corrected for multiple comparisons using Sidak test.

965



966

967 **Fig. 6. Proposed model illustrating the orchestration of lipid induced molecular changes**

968 *Sensors*: Senses the fatty acid rich environment and perturb cellular metabolism providing the

969 essential substrate for histone modifications and thereby turning on the *Mediators*- histone PTMs,

970 which consequently activates the *Effectors*- Notch, adenylate cyclase and MAPK-ERK the key

971 protein signaling associated with ER- breast cancer.

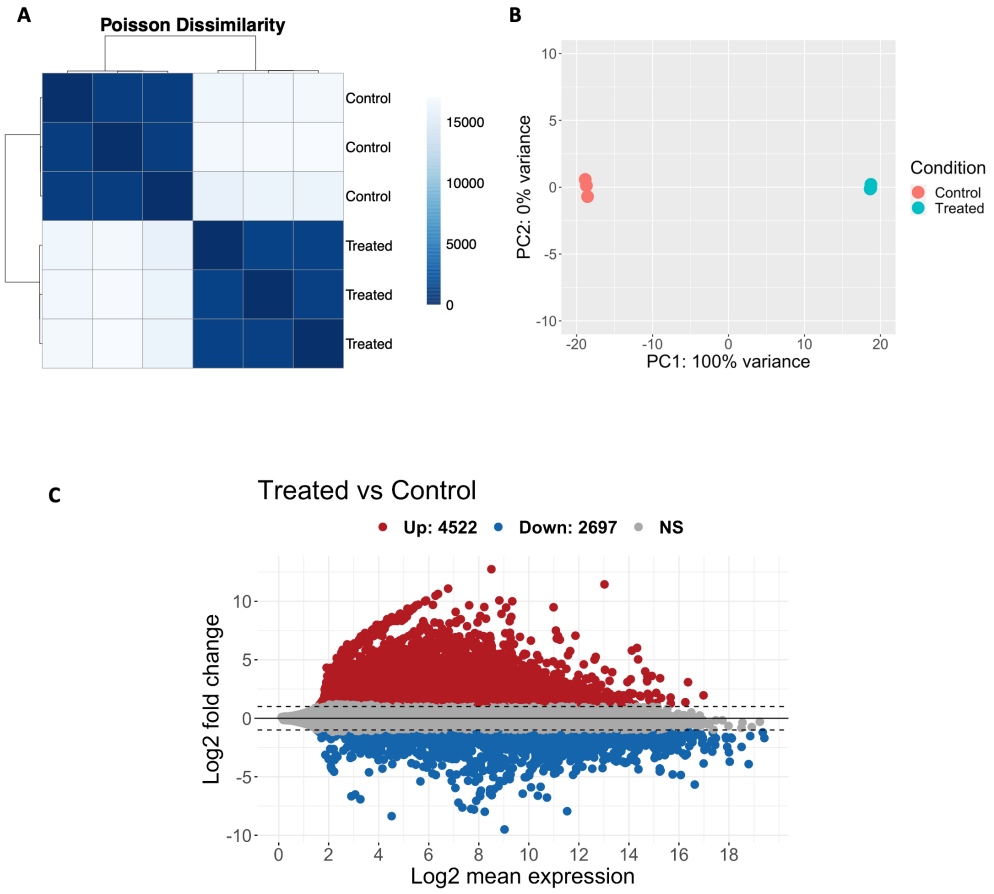


Fig. S1 (A)Poisson distance clustering to overview the distribution of counts and clustering in the treated versus untreated group. Scalebar represents Poisson distance between samples. (B) Principal component analysis (PCA) of the DESeq2 analysis showing two distinct populations of control and treated group. PCA dimensionality reduction was performed on all samples. Almost 100% of the variance is associated with the first principle component, which separates replicates in the vehicle and octanoate treatment conditions. (C) DESeq2 analysis showing 2131 upregulated genes and 632 downregulated genes for octanoate group compared with the vehicle with FDR cutoff = 0.01 and $|\log_2 FC| \geq 2$.

A

	N	Overall Pop: (n=84)	ER+ Pop: (n=28)	ER- Pop: (n=28)	Control Pop: (n=28)
Age					
	Median	50	51	50	50
	Range	(34-65)	(38-65)	(34-63)	(36-63)
BMI					
	Median	26	26	23	29
	Range	(17-41)	(18-40)	(17-41)	(21-41)
ER (%)					
	0	28 (33.33)	0	28 (100)	0
	1	28 (33.33)	28 (100)	0	0
	n/a	28 (33.33)	0	0	28 (100)
MenopauseStatus (%)					
	post	36 (42.86)	10 (35.71)	15 (53.57)	11 (39.29)
	post (hysterectomy)	13 (15.48)	7 (25)	1 (3.57)	5 (17.86)
	post (medication)	5 (5.95)	1 (3.57)	2 (7.14)	2 (7.14)
	pre	1 (1.19)	0	1 (3.57)	0
	pre(1)	7 (8.33)	1 (3.57)	1 (3.57)	5 (17.86)
	pre(2)	11 (13.1)	4 (14.29)	5 (17.86)	2 (7.14)
	pre(3)	11 (13.1)	5 (17.86)	3 (10.71)	3 (10.71)
MensmenoCode (%)					
	Early_Follicular	7 (8.33)	1 (3.57)	1 (3.57)	5 (17.86)
	Late_Follicular	11 (13.1)	4 (14.29)	5 (17.86)	2 (7.14)
	Luteal	11 (13.1)	5 (17.86)	3 (10.71)	3 (10.71)
	NA	1 (1.19)	0	1 (3.57)	0
	Post_menopausal	54 (64.29)	18 (64.29)	18 (64.29)	18 (64.29)
obcode (%)					
	NA	4 (4.76)	0	2 (7.14)	2 (7.14)
	NW	32 (38.1)	12 (42.86)	15 (53.57)	5 (17.86)
	OB	27 (32.14)	8 (28.57)	7 (25)	12 (42.86)
	OW	21 (25)	8 (28.57)	4 (14.29)	9 (32.14)
Race (%)					
	AA	13 (15.48)	3 (10.71)	3 (10.71)	7 (25)
	Cauc	70 (83.33)	24 (85.71)	25 (89.29)	21 (75)
	Other	1 (1.19)	1 (3.57)	0	0

B

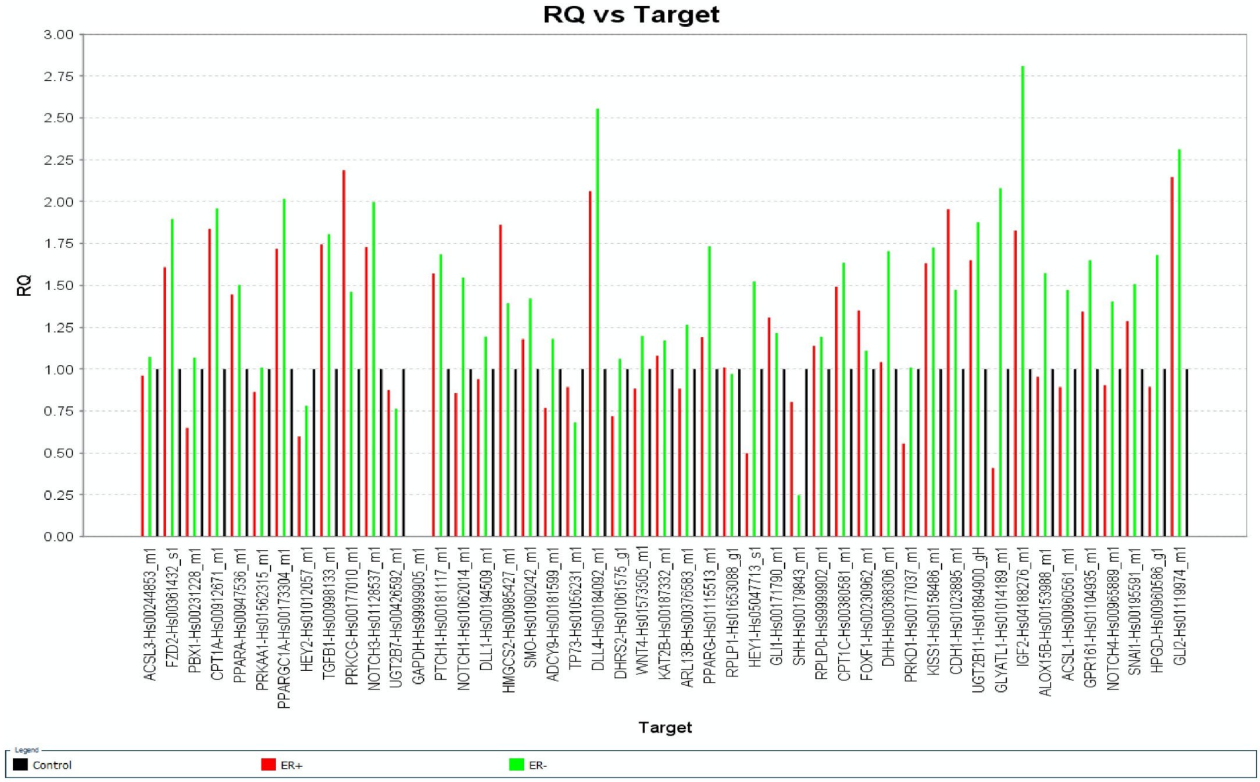
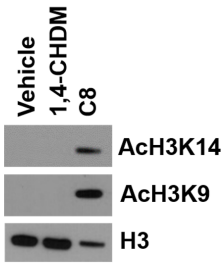
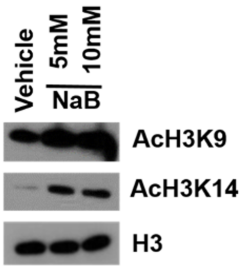


Fig. S2 (A) The difference in age and BMI among three groups was analyzed by ANOVA with Sidak adjustment on pairwise comparison. The difference in menopausal status and race among three groups were analyzed using X^2 test. The difference in HER2 status between ER1 and ER– group was analyzed using X^2 test. (B) Histogram showing fold change or relative quantitation (RQ) for all genes of interest in the ER + (red) and ER- (green) in reference to the controls (black).

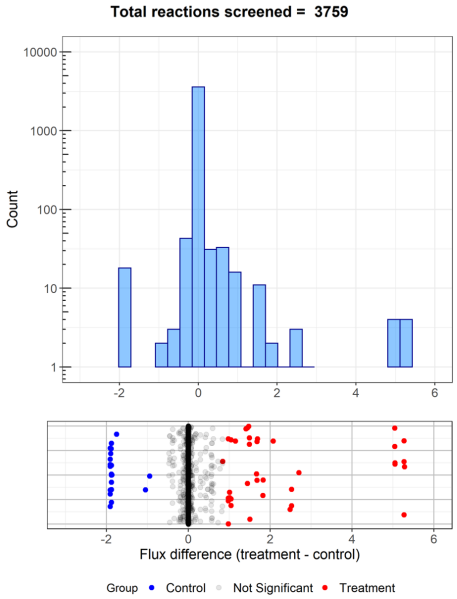
A



B



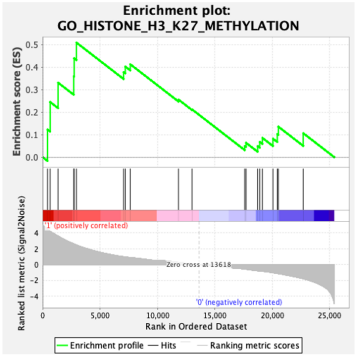
C

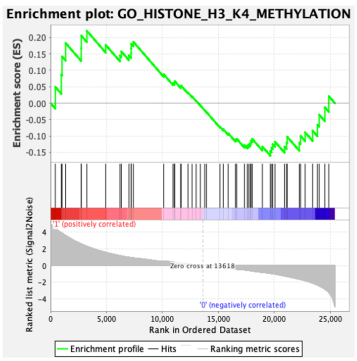


Subsystem	Reaction example	n
Triacylglycerol Synthesis	Diacylglycerol acyltransferase	7
Vitamin A Metabolism	Retinyl ester hydrolase	5
Fatty Acid Metabolism	Electron transfer flavoprotein	4
Pyruvate Metabolism	D-lactate dehydrogenase	4
Methionine Metabolism	Adenosylhomocysteinase	3
Folate Metabolism	Methenyltetrahydrofolate cyclohydrolase, mitochondrial	2
Glycerophospholipid Metabolism	Glycerol kinase	2
Miscellaneous	Peroxidase (multiple substrates)	2
Oxidative Phosphorylation	Ubiquinol-6 cytochrome c reductase, Complex III	2
Glycine, Serine, and Threonine Metabolism	Glycine N-methyltransferase	1
Glycolysis/ Gluconeogenesis	Glycerol-3-phosphate dehydrogenase (FAD), mitochondrial	1
Inositol Phosphate Metabolism	Inositol oxygenase	1
Methylation	Histone methylation	1
ROS Detoxification	Catalase	1
Transport, Endoplasmic Reticular	S-Adenosyl-L-methionine intracellular diffusion	1
Urea cycle/amino group metabolism	Sarcosine dehydrogenase (m)	1

Subsystem	Reaction example	n
Nucleotides	Nucleoside-diphosphatase (dUDP)	12
Glycolysis/ Gluconeogenesis	Glyceraldehyde-3-phosphate dehydrogenase	5
Pentose Phosphate Pathway	Deoxyribokinase	2
Oxidative Phosphorylation	NADH dehydrogenase, mitochondrial	1

D





E

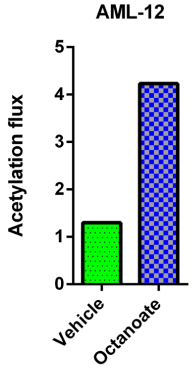


Figure S3: (A) To verify that the acetylation was specific to exposure to a fatty acid, MCF10A cells were exposed to 1,4-Cyclohexanedimethanol (1,4-CHDM), an alcohol with the same number of carbons, hydrogens and oxygens as octanoic acid. (B) Western blot of MCF-10A cells treated with HDAC inhibitor- sodium butyrate (NaB) 10mM for 24 hours to validate the specificity of the histone antibodies against the acetylated H3K9 and HK14.

(C) The histogram and scatter plot show the distribution of flux differences of all 3759 metabolic reactions in the model between octanoate treatment and control. The horizontal x-axis shows the difference in flux of each reaction, while the y-axis of the histogram shows the total number of reactions in each bin. Metabolic pathways and representative reactions that showed the greatest differences in flux (p-value < 0.01) between the treatment and control are highlighted in the scatter plot and listed in the table. (D) GSEA analysis showing H3K27 and H3K4 enrichment in octanoate treated cells with corresponding leading-edge genes. (E) Predicted acetylation flux in octanoate treated AML-12 cells using FBA model.