**SUPPLEMENTARY MATERIALS AND METHODS**

## 1. MOLECULAR, CELLULAR, AND IN VIVO STUDIES

### Cell lines

SUM159, SUM149T, HCC38, and BT-549 TNBC cell lines were grown according to manufacturer's protocols. Mammary epithelial hTERT-HME (HME) and MCF-10A cell lines were cultured as described (Cardone et al., 2012).

### Lentiviral infection of SUM159-derived mammospheres

Mammospheres formation. Adherent cells were replated in ultralow attachment flasks (Corning) and cultured in mammospheres-forming medium (MFM) composed of serum-free DMEM/F12 modified medium containing 20 ng/ml EGF (Peprotech, 100-15), 10 ng/ml bFGF (Peprotech, 100-18B), 50 µg/L insulin (Sigma, 91077C) and supplemented as described (Ricci-Vitiani et al., 2008). Mammospheres derived from SUM159 TNBC cells and obtained by three weeks culture in mammospheres-forming medium, were transduced with lentiviral particles produced in 293T cells co-transfected with lentiviral TWEEN-Luc-EGFP vector and both envelope (pMD2.G) and packaging (psPAX) vectors by using CaCl2 transfection method in complete IMDM media. Concentrated lentiviruses were added to the cells and media supplemented with 1 µg/ml of polybrene. The lentiviral reporter vector TWEEN-Luc-EGFP was constructed by using pRRL-CMV-PGK-GFP- WPRE (pTWEEN) as a backbone for subcloning a firefly luciferase NheI/XbaI cDNA fragment extracted from pGL3 (Promega) into the XbaI site at the 3' of the CMV promoter.

### Generation and propagation of mammospheres-forming cells (MFCs) with stem-like cancer cells from primary tumors (FP-MFCs) and derivatives metastases (met-MFCs).

Third generation SUM159 mammospheres were infected with an EGFP-Luciferase reporter and were orthotopically injected into mammary fat pad. After reaching a volume of ~300 mm3, primary tumors were removed, and appearance of metastasis monitored in time. After surgical removal, tumor cells from fat pad (FP) or corresponding metastases (met) were isolated by cell sorting for the expression of EGFP and cultured in mammospheres-forming medium to generate FP-MFCs- or met-MFCs-derived spheroids enriched in CSCs. Three independent fat pad ortothopic tumors have been isolated (FP#1, FP#2 and FP#6) with their relative nodal or lung metastases. To propagate both FP-MFCs or met- MFCs, spheroids were single cells dissociated by mild incubation with Accutase

(Euroclone), followed by mechanical dissociation, counted and replated in mammosferes-forming medium at concentration of $10x10^3$ MFCs/ml in ultralow attachment flasks (Corning). Medium was replaced every three days through gentle centrifugation for spheroids collection. The self-renewal potential of MFCs allowed the propagation of CSCs for several passages.

**ALDH activity by Flow cytometry**

For the quantification of $ALDH^{high}$ cellular subpopulation, single-cell suspensions ($1x10^6$ cells) from FP#1, SUM159-2D or SUM159-3D were resuspended in assay buffer and 5 µl of AldeRed 588-A (AldeRed™ ALDH Detection Kit., Merk) were added to the suspension. In each case a 0.5 ml aliquot of the suspension was immediately added to the tube containing 5 µl of DEAB. Cells were incubated at 37 °C for 40 minutes followed by centrifugation and washing with 0.5 ml of cold assay buffer. Resuspended cells were stored on ice until analysed. Stained cells were analysed by FACS Aria or FACS LSR II (BD Biosciences) equipped with three lasers (405, 488, and 633 nm) and 8 wavelengths detected. AldeRed 588-A was also tested on a FACSCalibur (BD Biosciences) with 488-nm blue laser and phycoerythrin (PE) filter.

**Flow Cytometry analysis**

For cytofluorimetric analysis, adherent cells or mammosphere were dissociated as single cells, washed and $1X10^5$ cells were incubated with the appropriate dilution of control or specific antibody. Primary antibodies used: APC-conjugated mouse anti-Human CD44 and PE/Cy7-conjugated mouse anti-Human CD24 from BD Biosciences (San Jose, CA). APC Mouse IgG2b κ Isotype and PE-Cy™7 Mouse IgG2a, κ Isotype were employed as control antibodies. Cell analysis was performed using either a FACScan or an LSRII flow cytometer (Becton Dickinson, Franklin Lakes, NJ).

**mRNA and Quantitative Real-time-PCR (qRT-PCR)**

Total RNA extraction was carried out using RNeasy kit (Quiagen) according to the manufacturer's instructions. Total RNA was then reverse transcribed into cDNA by QuantiTect Reverse Transcription Kit (Qiagen) with random hexamers oligo. The cDNA was diluted 1:3 and subjected to quantitative real time PCR analysis by using Light Cycler (Applied Biosystem) with SYBR Green PCR Master MIX Kit (Applied Biosystem).

**Antibodies and Reagents**

The following antibodies were purchased from Cell Signalling Technology and used according manufacturer's instructions: anti PARP (ref.: #9542), anti-cleaved caspase-3 (Asp175) (ref.: #9661). Anti-β-actin (A2228) was purchased from Sigma-Aldrich. Pyrvinium pamoate, 2-deoxyglucose (2-DG), etomoxir, BPTES, UK-5099, succinate, pyruvate, palmitoyl-carnitine, rotenone, metformin, antimycin A, pamoate salt, ICRT3, betulin, 6-Fluoromevalonate, and atorvastatin were purchased from Sigma-Aldrich. All Seahorse consumables were purchased from Seahorse Biosciences (North Billerica, MA, USA).

**Cell viability assay and Cell-cycle analysis**.

10 x $10^3$cells were cultured as mammospheres and treated as indicated for 72hrs. At the end of the incubation, cells were collected and both cell viability and cell proliferation were determined by trypan blue dye exclusion assay. Cell proliferation was calculated as Fold Increase over the number of cells plated at time 0 (T=0). Flow cytometric cellular DNA content evaluation was performed on the same samples by PI staining. Briefly, after trypsinization, cells were fixed in 70% ethanol overnight at -20°C. Samples were then washed in phosphate-buffered saline 1X (PBS 1X) and incubated for 30 min at 37°C with a PI solution (0.1% TritonX-100, 50 mg/ml propidium iodide, and 200mg/ml RNAse A). Twenty thousand events were acquired (BD Accuri Instrument, BD Biosciences, San Jose, CA, USA) and analysed by using the FlowJoX software.

**Colony formation assay**

Soft agar colony-forming assays were carried out for CSCs treated with either vehicle (DMSO) or drugs. 100 (or 500) single cells were plated in the top agar layer in each well of a 24-well culture plate with 0.3% top agar layer and 0.4% bottom agar layer (SeaPlaque Agarose, Cambrex, NJ, USA). Cultures were incubated at 37 °C for about 10-20 days. Colonies from triplicate wells were stained with crystal violet (0.01% in 10% MetOH), visualized and counted under microscope and photographed.

**Protein extracts and Western blot analysis.**

Cells were washed in PBS1X containing 1X protease inhibitor cocktail and 1mM PMSF, were lysated in modified RIPA buffer: 50 mM Tris-HCl [pH 7.8], 150 mM NaCl, 5 mM EDTA, 15 mM MgCl2, 1% Nonidet P-40, 0.5% sodium deoxycholate, 1 mM dithiothreitol, 1X protease inhibitors, 1mM PMSF, 50 mM NaF, 10 mM β-glycerophosphate and 1 mM $Na_3VO_4$, and then immediately frozen in liquid nitrogen. Cleared protein extracts were quantified by using the Bradford assay (Bio- Rad). For

western blotting analysis, protein samples were separated on 8%–12% SDS-PAGE and transferred to nitrocellulose membrane (Amersham). Membranes were blocked in TBS containing 5% non-fat milk, incubated with primary antibodies according to the antibody manufacturer's instructions, followed by incubation with horseradish peroxidase-conjugated goat anti-rabbit or anti-mouse IgG (Amersham) and enhanced chemiluminescence detection (Pierce).

## 2.IN SILICO STUDIES

### BIO-COMPUTATIONAL METHODS for GSMM-based studies

**A constraint-based model of metabolism.** A metabolic network consisting of $m$ metabolites and $n$ reactions can be represented by a *stoichiometric matrix* $S$, where the entry $S_{ij}$ represents the stoichiometric coefficient of metabolite $i$ in reaction $j$. A Constraint-Based Model (CBM) imposes mass balance, directionality and flux capacity constraints on the space of possible fluxes in the metabolic network's reactions through a set of linear equations

$$S \cdot v = 0 \qquad (1)$$

$$v_{min} \leq v \leq v_{max} \qquad (2)$$

Where $v$ is the flux vector for all reactions in the model (i.e. the *flux distribution*). The exchange of metabolites with the environment is represented as a set of *exchange (transport) reactions*, enabling a pre-defined set of metabolites to be either taken up or secreted from the growth media. The steady- state assumption represented in Equation (1) constrains the production rate of each metabolite to be equal to its consumption rate. Enzymatic directionality and flux capacity constraints define lower and upper bounds on the fluxes and are embedded in Equation (2). In the following, flux vectors satisfying these conditions will be referred to as feasible steady-state flux distributions. Gene knockdowns are simulated by constraining the flux through the corresponding metabolic reaction to zero. Similarly, environmental perturbations are simulated by constraining the flux through the associated exchange reaction to zero.

For each of the dataset analyzed here, we simulated the same media that was used in the experiment (DMEM). For modelling human metabolism, we have used Recon1.

### Metabolic Transformation algorithm (MTA)

MTA receives as input the gene expression measurement of two distinct metabolic states, termed source and target states. The algorithm then executes the following steps: (1) to determine the flux distribution that corresponds to the source state using integration Metabolic Analysis Tool (iMAT); (2) to identify the set of genes whose expression have significantly elevated or reduced between the source and targets states, and the set of genes whose expression remained relatively constant between the states. Next, the algorithm searches for perturbations that can alter the fluxes of the

changed reactions in the observed direction, while keeping the fluxes of the unchanged reaction as close as possible to their predicted source state. Finally, MTA outputs a ranked list of candidate perturbations according to their ability to transform from the source to the target metabolic state.

**The Transformation Score**

Relying on the optimization value obtained by MTA to rank the transformations induced by different perturbations is suboptimal, since the integer-based scoring of the changed reactions is coarse-grained and does not distinguish between solutions achieving large flux alterations and those obtaining flux changes barely crossing the $\varepsilon$ threshold. Therefore, we chose to quantify the success of a transformation by a scoring function based on the resulting flux distributions rather than on the optimization objective values themselves. First, we denoted the resulting flux distribution obtained in a given MIQP solution (for a given reaction knock-out) as $\mathrm{v}^{\mathrm{res}}$. Second, reactions found in $\mathrm{R_F}$ and $\mathrm{R_B}$ are classified into two groups $\mathrm{R_{success}}$ and $\mathrm{R_{unsuccess}}$, denoting whether they achieved a change in flux rate in the required direction (forward or backward) or not. The following scoring function is then used to assess the global change achieved by the employed perturbation:

$$\frac{\sum_{i\epsilon R_{success}} abs\left[\left(v_i^{ref}-v_i^{res}\right)\right]-\sum_{i\epsilon R_{unsuccess}} abs\left[\left(v_i^{ref}-v_i^{res}\right)\right]}{\sum_{i\epsilon R_S} abs(v_i^{ref}-v_i^{res})} \qquad (*)$$

The numerator of this function is the sum over the absolute change in flux rate for all reactions in $R_{success}$, minus a similar sum for reactions in $R_{unsuccess}$. The denominator is then the corresponding sum over reactions in $R_S$ (the reactions which should stay untransformed). Following, perturbations achieving the highest scores under this definition are the ones most likely to perform a successful transformation by both maximizing the change in flux rate for significantly changed reactions, and minimizing the corresponding change in flux of unchanged reactions. Using an alternative scoring function based on the Euclidean distance instead of absolute values yielded similar results. While we believe that the TS score (Equation (*)) is the right one to pursue from a biological point of view, optimizing it directly is a very difficult mathematical task. To accomplish that it would necessary to develop a novel optimization algorithm for solving a mixed *non-linear* programming problem, whose objective function is non-smooth and non-differentiable, requiring non-smooth optimization tools. Attempting such a solution directly would greatly complicate the problem as one would need

to add many variables and constraints. Furthermore, the specific form of this ratio is actually dependent on the solution itself (as it evaluates $R_{!"\#\#\$!!}$ and $R_{!"\#!\$\$\%\#\#}$ separately) making the entire task infeasible. In light of these evident difficulties, we have chosen to take a two-steps approach in this study that is sub-optimal but yet tractable. While the wild-type solution always achieves maximal values in terms of the original proxy objective function used in step 3 (by definition), it does not necessarily achieve high transformation scores (step 4). This is because the wild type solution is the least constrained, and hence most of the solutions found in step 3 can be satisfied by achieving only a minimal epsilon change. Those are obviously non-optimal from a biological standpoint as they do not really come close to the desired objective, and hence their TS score (in step 4) is sub-optimal in many of the cases, correctly ruling them out as biologically viable solutions. MTA analysis is established upon learning the regulatory effects of the knockdown of metabolic genes via the direct stoichiometric flux coupling of the reactions they encode to other reactions in the human metabolic network (which are inherently embedded in the reactions stoichiometric matrix it includes).

**Reactions selected by MTA analysis.**

MTA was applied from CSS (source state) to metastases (target state) four times, for four different p-value thresholds used for classification of reaction to $R_F$, $R_B$ and $R_S$ (FDR corrected P<0.5,0.1,0.05 and 0.01). Selected reactions are reactions that in all four runs were scores higher than random (MTA score when no perturbation is simulated).

**Kaplan-Meier survival analysis.**

Kaplan-Meier survival analysis is applied to examine the association of genes/pathways with poor patient survival. We use METABRIC survival and gene expression data (http://www.nature.com/nature/index.html) and separate the expression of each gene/pathway to 'high' and 'low' bins by its median level (pathway expression is assigned with the median expression of all genes in a pathway). We calculated the $\Delta AUC$ and log-rank P-value resulting from the two Kaplan-Meier curves. The association between SQLE and MVD gene expression with BC metastases occurrence and disease relapse was evaluated by PROGgeneV2 tool: (http://watson.compbio.iupui.edu/chirayu/proggene/database/index.php) (Goswami and Nakshatri, 2013, 2014) and by the Kaplan-Meier plotter tool (Gyorffy B, et al., 2010),(http://kmplot.com/analysis/index.php?p=service&cancer=breast).

In each dataset, for the selected gene, survival information in terms of survival status (overall or metastasis free or recurrence free survival), and survival time (time to death or time to metastasis or time to recurrence) are retrieved along with gene expression as continuous variable. Using median gene expression value as bifurcating point, samples are divided into High and Low gene expression groups. Using survival data and continuous expression variable, survival analysis is done by fitting cox proportional hazards model using function “coxph” of library survival. Hazard ratio (HR) as 'exp(coef)’ and log rank p value are retrieved from the fitted model

## 3. METABOLISM STUDIES

### Mitochondrial substrate specific oxidation assay in permeabilized Cells

To measure the different specific respiratory complexes, FP-MFCs treated for 12 hours as indicated were permeabilized by using the XF Plasma Membrane Permeabilizer (XF PMP- Agilent Seahorse Bioscience) and specific substrate oxidation was examined by measuring OCR changes using a Seahorse-XF Analyzer. To measure the oxidation of NADH-linked substrates, OCR was measured prior to and after sequential injections of 10 mM pyruvate with 1mM malate and rotenone (Measurements 4-7) at 2 µM working concentration. To measure complex II OxPhos activity, OCR was measured prior to and after sequential injections of 10 mM succinate with 2 µM rotenone (Measurements 4-7) and antimycin A at 2 µM working concentration. To analyse palmitoyl-carnitine oxidation, OCR was measured prior to and after sequential injections of 40 µM palmitoyil-carnitine (Measurements 4-7) and antimycin A at 2 µM working concentration. To assess cytochrome oxidase activity, OCR was measured prior to and after injection of 100 µM TMPD with 10 mM of ascorbate (Measurements 5-7). All chemicals necessary for this assay have been purchased from Sigma Aldrich. Catalogue' number of each compound is available on request.

### Metabolite extraction and derivatization method

The metabolites extraction from quenched mammospheres was performed as described before (Elia et al., 2017; van Gorsel et al., 2019). Briefly, samples were resuspended with 800 µl of 60 % methanol containing 90 ng/ml of glutaric acid precooled in a mixture of dry and wet ice and were lysed with a tissue lyser for 3 min in a -40 °C dry ice-ethanol bath. Subsequently, were added 500 µl of precooled chloroform containing 10 µg/ml of C17 internal standard and samples were vortexed for 10 min at 4°C followed by a centrifugation for other 10 min (max. speed, 4 °C). After centrifugation, polar metabolites in the methanol/water (upper) phase and the lipid fraction in the chloroform (lower) phase were separated by a protein layer. The lipid fraction was divided in two Eppendorf with equal volumes for the analysis of fatty acids and cholesterol. Following metabolites separation, every phase was dried at 4 °C overnight using a vacuum concentration.

The samples were derivatized and measured as described before (Lorendeau et al., 2017). Briefly, polar metabolites were derivatized in 13 µl of 20 mg/ml methoxyamine in pyridine per sample for 90 min at 37 °C. Subsequently, 15 µl of N-(tert-butyldimethylsilyl)-N-methyl-trifluoroacetamide,

with 1 % tert-butyldimethylchlorosilane were added to 7.5 µl of each derivative and incubated for 60 min at 60 °C. Fatty acids were esterified with 500 µL of 2 % sulfuric acid in methanol buffer per sample and incubated overnight at 50 °C. Subsequently, fatty acids were extracted with 600 µl of MS-grade hexane and 100 µl of saturated NaCl. Hexane fraction was dried in a vacuum concentration at room temperature for 30 min and was resuspended in 30 µl of hexane. The dried fatty acid phase for cholesterol analysis was saponified with 20 % w/v of sodium hydroxide in 50 % ethanol at 97 °C for 5 min, and the unsaponifiable fraction was extracted with hexane and dried. Cholesterol samples were then resuspended in 20 µl of pyridine and derivatized with trimethylsilyl (TMS) for 1 hour at 40 °C.

**Gas chromatography–mass spectrometric analysis for isotope-based fluxomic studies**

The metabolites were analyzed by gas chromatography (7890A GC system) coupled to mass spectrometry (5975C Inert MS system) from Agilent Technologies. The inlet temperature was set at 270 °C and 1 µL of sample was injected with a split ratio 1 to 3. The separation of the metabolites was performed with a DB35MS column (30 m, 0.25mm, 0.25 µm) with a carrier gas flow of helium fixed at 1.3 ml/min for the analysis of fatty acids and at 1 ml/min for the analysis of cholesterol and polar metabolites. Different gradients of temperature were applied for the analysis of polar metabolites, fatty acids and cholesterol. For the detection of fatty acids, the oven was set at 140 °C for 2 min and increased at the ramping rate of 1°C/min to 185°C following by a ramping rate of 20°C/min to 300°C. For the measurement of cholesterol, the initial temperature was set at 160°C for 1 min and ramped to 320°C at 20°C/min. Finally, polar metabolites were analyzed with a gradient set at 100°C for 1 min ramping to 105°C at 2.5°C/min, then to 240°C at 3.5°C/min and finally to 320°C at 22°C/min.

For the measurement of metabolites by mass spectrometry, the temperatures of the source and the quadrupole were set at 230°C and 150°C, respectively. A scan mode ranging from 100 to 600 a.m.u (mass) with an electron impact ionization fixed at 70 eV was applied for the analysis of cholesterol and fatty acids, while a selected-ion monitoring (SIM) mode was used for the measurement of polar metabolites.

**Data analysis – Matlab**

After the acquisition of raw ion chromatograms, a Matlab M-file was used to extract mass distribution vectors, then the different metabolites were integrated and corrected by the naturally isotopes distribution using the method developed by Fernandez et al, 1996 (Fernandez, Charles A., et al.

"Correction of 13C mass isotopomer distributions for natural stable isotope abundance." Journal of Mass Spectrometry 31.3 (1996): 255-262.). The peak area was subsequently normalized to the protein content and to the internal standards, C17 was used for fatty acids and cholesterol peak area correction, while glutaric acid internal standard was used for the correction of polar metabolites. Finally, the total contribution of carbon was calculated as previously published by Buescher et al. 2015 (Buescher et al., 2015).

**ATP measurement**

The intracellular ATP content in the cells was determined using the Cell Titer-Glo Luminescent Cell Viability Assay. $50x10^3$ cells plated in 96-well microplates were treated for 12 hrs with PP and then lysed by directly adding Cell Titer-Glo Reagent to each well. Luminescent signals were detected by using Luminometer GLOMAX (Promega). The percentage of intracellular ATP content reduction was calculated by comparing the signal of treated vs. control cells. Three replicates were used for each group, and the experiments were repeated three times to confirm the results.

**Measurement of intracellular NADH/NAD and NADPH/NADP ratios.**

Levels of NAD, NADH, NADP, NADPH were measured using the NADH/NAD (#MAK037) and the NADPH/NADP (#MAK038) Assay Kits from Sigma Aldrich (USA) and according to the manufacturer's instructions.

***1H*** **NMR analyses**

Cells were washed in PBS and centrifuged at 162 rcf for 3 min. The pellet was resuspended in PBS with n20% D2O and 2mM Sodium 3-(trimethylsilyl) propionate -2,2,3,3-d4 (TMSP) as a frequency standard. A 15 µl aliquot of the suspension was transferred into a 1 mm NMR tube and centrifuged to obtain a packed cell volume. Perchloric acid (PCA) extracts were prepared as described in a previous work *(Palma et al., 2011).* All NMR reagents were purchased from Cambridge Isotope Laboratories, Inc. 1D and 2D COSY 1H NMR spectra were run at 400.14 MHz on a digital Avance spectrometer equipped with a 1 mm microprobe (Bruker, AG, Darmstadt, Germany). Both 1D and 2D COSY spectra were acquired at T = 298 K. Water suppression in 1D and 2D 1H experiments were obtained using pre-saturation. Intact cell spectra were acquired with several scans (ns) equal to 1000 (sufficient to obtain a good signal-to-noise ratio), while ns=4000 were used for PCA extract and culture media spectra. 2D COSY spectra were acquired with a 90°-t1–90°-t2 pulse sequence and

ns=32 for cell or ns=128 for PCA extracts samples. Chemical shifts were measured with respect to Lactate methyl signal (Lac) at 1.33 ppm in 1D spectra and to Lac cross peak at 1.33– 4.12 ppm in 2D COSY spectra for both cells and PCA extracts spectra. All NMR parameters were obtained in at least three independent experiments. Metabolite signal assignments were performed according to indications from literature and by comparison with pure compounds. 1D peak deconvolution and integration as well as 2D COSY cross peak integration were performed by the WINNMR software (Bruker). Individual integral values were normalized using the methyl group of cytosolic polypeptides at 0.94 ppm as internal reference for 1D spectra while for 2D spectra signal integrals were normalized to the intensity of the lysine cross-peak at 1.70–3.00 ppm. The analyzed metabolites were determined as following: Fumarate signal was measured in 1D spectra at 6.51 ppm; Aspartate signal intensity was measured in 2D COSY spectra at 2.80-2.67 ppm. Aspartate signal was measured in 1D spectra at 2.68 ppm (Asp1) and 2.79 ppm (Asp2); Succinate signal intensity was measured in 1D spectra at 2.41 ppm.

**LC/GC-MS analysis**

**Chemicals and reagents:** Methanol, acetonitrile, isopropyl alcohol, and ethyl acetate were of HPLC-MS grade and were purchased from Merck (Darmstadt, Germany). HPLC-MS grade ammonium formate was purchased in granular form Fluka (Buchs, St. Gallen, Switzerland). Butylated hydroxytoluene (BHT) and acetic acid were purchased from Sigma–Aldrich (St. Louis, MO, US). Deuterated cholesterol in 2,2,3,4,4,6 (d6-cholesterol), d5glyceryl-d93-trihexadecanoate (d98TG 48:0), and d17-palmitic acid (d17C16:0) were purchased from CDN Isotopes Inc. (Pointe-Claire, Quebec, Canada). Deuterated cholesterol oleate (d7CE18:1) was purchased from Avanti Polar Lipids, Inc. (Alabaster, Alabama, US).

**Lipid extraction:** Lipids were extracted from FPS 105 cells according to the procedure described by DahlhoffM. et al. with slight adaptations. Briefly, 1,5 x 106 FPS cells were suspended in 200 µL of distilled water (18.2 Ω), and cracked by repeated freezing in liquid nitrogen and-thawing. Cell debris were pelleted by spinning at 11000 rpm for 5 min. Protein content was determined by Bradford assay in 5 µL supernatant. The pellet was suspended by vortex-shaking, mixed with 100 µL methanol, and spiked with 5, 1 and 10 nmoles of d6-cholesterol, d98TG 48:0, and d17C16:0, respectively, and 400 pmoles of d7CE(18:1) internal standards (iSTD) dissolved in isopropyl alcohol containing 0.001% of BHT to prevent lipids autoxidation. Liquid-liquid extraction of lipids was performed with 1 mL of absolute ethyl acetate. After vigorous vortex-mixing for 10 minutes, and

spinning at 11000 rpm for 2 minutes, the upper organic phase was transferred to a clean Eppendorf tube. The operation was repeated twice and the pooled organic phases were evaporated to dryness under a nitrogen stream. The dried lipid extract was dissolved in 200 µL isopropanol and used for the analysis of free fatty acids (FFA), and sterols, including cholesterol and its esters (CE), and glycerol lipids, such as triglycerides.

**Gas chromatography-mass spectrometry**

Gas-chromatography coupled to mass spectrometry (GC-MS) was used to quantitate FFA and cholesterol, along with its intermediates. Samples were analyzed with a GC 7890A coupled to the MS 5975 VL analyzer (Agilent Technologies, CA, USA) as previously reported (Singh et al, 2018). Briefly, 50 µL of extract was evaporated to dryness and dissolved in 50 µL bis(trimethylsilyl)-trifluoacetamide (TMS) in pyridine to obtain trimethylsylil (TMS) derivatives of FFA, and sterols. The reaction was carried out at 60 °C for 60 minutes. GC-MS analyses were performed with a gas-chromatographer series 7890A coupled with the mass spectrometer series 5975C (Agilent Technologies, CA, USA). GC separation was performed with an Agilent DB 5ms ultra inert column (30 m x 0.25 mm, 0.25µm film thickness) and the MS acquisition was done in SCAN mode by means of electron impact (EI) MS. Helium was used as the carrier gas. The retention times (RT), the target/qualifier masses, and calibration curves were defined on the basis of reference authentic standards. Compound areas were normalized by the respective iSTD. Cholesterol and intermediates of cholesterol biosynthesis were quantified against d6-cholesterol, whereas FFA were quantified against d17C16:0 with the MassHunter quantitative software (Agilent Technologies, CA, USA). The quantitative data obtained from 6 independent experiments were reported as mean ± standards deviation (SD). Significance was determined by two-tailed unpaired student's t test. Differences were considered statistically significant when p values were ≤ 0.05.

**Liquid chromatography-mass spectrometry**

The chromatographic apparatus consisted of the 1260 Infinity II series LC system (Agilent Technologies, CA, USA). The stationary phase of the high resolution reversed phase LC (RPLC) was a Zorbax SB-C8 Zorbax SB-C8 rapid resolution HT 2.1 x 100 mm 1.8 µm p.s. with a maximal operational backpressure at 600 Bar (Agilent Technologies, CA, USA). Lipid mixtures were eluted a binary gradient of (A) 5 mM ammonium formate in water, (B) methanol, (C) acetonitrile, (D) isopropanol. The mobile phases were filtered through 0.45 µm glass filters and continuously degassed under

vacuum. The elution program was as follows: A/B/C/D 60/28/8/4 at time 0 and held for 1 min, brought to A/B/C/D 1/70/20/9 in 10 min and held up to 20 min. The flow rate was maintained at 400 µL/min during the entire LC run. The column was thermostated at 60 ºC. The injection volume was 0.80 µL. The injector needle was washed with the mobile phase in the wash port during the LC runs. The eluent outlet was connected to two different MS analyzers for the detection and characterization.

Accurate mass measurements in full MS and auto MS/MS were conducted with a G6545B series LC-QTOF (Agilent Technologies, USA) equipped with a JetStream Technology electrospray interface (ESI) interface operating in positive ion mode. Analytes eluted from the LC system were introduced into the Q-TOF apparatus at the operating chromatographic flow rate (see chromatographic conditions). Nitrogen was used as the nebulizing and desolvation gas. The temperature and the flow of the drying gas temperature were 200 °C, and 12 L/min, respectively. The temperature and the flow of the sheath gas were 350°C and 12 L/min, respectively. The nebulizer pressure was 40 psi. The capillary and the fragmentor voltage were 4000 and 180 V, respectively. Full scan mass spectra were acquired in the range from m/z 100 to m/z 1700. To enhance accurate mass measurement for the ion species a reference solution of two compounds with m/z 121.050873 and 922.009798, respectively, was vaporized in continuum in the spray chamber by means of a separate nebulizer.

**Extraction of MS features**

Molecular features, defined by an m/z, RT and signal intensity value, were extracted from the raw RPLC-MS data files using the untargeted or the targeted batch recursive feature extraction in the MassHunter Profinder software (Agilent Technologies, USA). Procedures and details can be found in 'MassHunter Profinder Software Quick Start Guide on the Agilent Technologies webpage. The features extracted were exported into a compound exchange format (CEF) reporting RT, the accurate mass and the absolute abundance for each entity to be processed in the subsequent chemometric analysis as previously reported (Ludovici et al., 2018; Singh et al., 2018).

**Data analysis**

Agilent Mass Profiler Professional (MPP version 15.1) was used to process the RPLC-MS untargeted and targeted data. RT were aligned by setting a window of 0.6 minutes, whereas m/z binning was performed by setting windows at 10 ppm. Absolute abundance of each entity was normalized by the absolute abundance of the d98TG 48:0 iSTD. Data were filtered by frequency of detection, which

reflects the number of samples that presented particular features. A frequency filter was applied to data extracted from MPP and only entities present in 100% of samples belonging to at least one of the investigated groups were retained for the statistical analysis. Fold changes of filtered entities were compared between groups volcano plots in the MPP tools. Fold changes ≥ 1.5 with p values < 0.05 after the Bonferroni's correction were considered as significant. Identification of entities within the MPP workflow was performed based on the METLIN Metabolomics Database (http://metlin.scripps.edu/) and the Lipid Annotator software (Agilent Technologies, CA, USA). Quantitative assessment of cholesterol esters (CE) and TG were performed with the iSTD d7CE(18:1) and d98TG 48:0, respectively.

**Metabolomic analysis of breast cancer samples**

Breast cancer patients were recruited between 1993 and 2003, as described previously (Mishra et al., 2018; Terunuma et al., 2014a). Patients completed a questionnaire and provided biospecimens. Samples of fresh-frozen tumor tissue and adjacent non-cancerous tissue were processed by a pathologist immediately after surgery at the Department of Pathology, University of Maryland. Clinical and pathological information was obtained from medical records and pathology reports. The collection of biospecimens and the clinical and pathological information was approved by the University of Maryland Institutional Review Board for the participating institutions (UMD protocol #0298229). IRB approval of this protocol was then obtained from all institutions (Veterans Affairs Medical Center, Union Memorial Hospital, Mercy Medical Center, and Sinai Hospital, Baltimore, MD). The research was also reviewed and approved by the NIH Office of Human Subjects Research Protections (OHSRP #2248). All patients signed a consent form. The research followed the ethical guidelines set by the Declaration of Helsinki.

Metabolomic profiling of human breast tissues was performed using both an untargeted discovery approach and a targeted approach for validation and absolute quantification. Untargeted metabolic profiling of known and unknown metabolites in the discovery set included 67 human breast tumors and 65 tumor-adjacent noncancerous tissues and was performed by Metabolon Inc, as described previously previously (Mishra et al., 2018; Terunuma et al., 2014a). Clinical and pathological information (e.g., hormone receptor status) was obtained from medical records and pathology reports, with 33 ER positive and 34 ER negative. Basal-like and Her2-positive tumors were defined using both gene expression and immunohistochemistry (IHC) data (e.g., ER- negative, HER2-negative, and cytokeratin 5/6-positive or EGFR-positive IHC for basal-like tumors), as previously

described for these tumors previously (Mishra et al., 2018; Terunuma et al., 2014a). 17 Triple-negative tumors were negative for estrogen, progesterone, and HER2 receptor expression, and most were of the basal-like subtype (16 out of 17). Self-reported race/ethnicity was collected as Black (not of Hispanic origin) for African-Americans and White (not of Hispanic origin) for European-Americans. Race/ethnicity was further evaluated with ancestry informative markers and self-identified African-American patients had African ancestry in the range of 67% to 95% (Terunuma et al., 2014b). Disease staging was performed according to the tumor–node–metastasis (TNM) system of the American Joint Committee on Cancer/ the Union Internationale Contre le Cancer (AJCC/UICC). The Nottingham system was applied to determine tumor grade. Patient survival was followed up with National Death Index data from the USA National Center for Health Statistics.

**Statistical Analysis**

Results were expressed as Mean SD, and the number of individual replicates detailed in figure legends. Statistical significance of the mean values was established by two-tailed distribution Student's t test. Where appropriate, the use of different statistical approaches has been specifically described in the figure legend.

**SUPPLEMENTARY REFERENCES TO MATERIALS AND METHODS**

Buescher, J.M., Antoniewicz, M.R., Boros, L.G., Burgess, S.C., Brunengraber, H., Clish, C.B., DeBerardinis, R.J., Feron, O., Frezza, C., Ghesquiere, B., et al. (2015). A roadmap for interpreting (13)C metabolite labeling patterns from cells. Curr Opin Biotechnol *34*, 189-201.

Cardone, L., Bardelli, A., and Avvedimento, V.E. (2012). Activation of beta-catenin by oncogenic PIK3CA and EGFR promotes resistance to glucose deprivation by inducing a strong antioxidant response. PLoS One *7*, e37526.

Dahlhoff M. at al. 2013 , Volume 1830, Issue 10, October 2013, Pages 4642–4649.

Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M.F., Verfaillie, C., Grunewald, T.G.P., and Fendt, S.M. (2017). Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. Nat Commun *8*, 15267.

Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients, Breast Cancer Res Treatment, 2010 Oct;123(3):725-31.

Goswami, C.P., and Nakshatri, H. (2013). PROGgene: gene expression based survival analysis web application for multiple cancers. J Clin Bioinforma *3*, 22.

Goswami, C.P., and Nakshatri, H. (2014). PROGgeneV2: enhancements on the existing database. BMC Cancer *14*, 970.

Lorendeau, D., Rinaldi, G., Boon, R., Spincemaille, P., Metzger, K., Jager, C., Christen, S., Dong, X., Kuenen, S., Voordeckers, K., et al. (2017). Dual loss of succinate dehydrogenase (SDH) and complex I activity is necessary to recapitulate the metabolic phenotype of SDH mutant tumors. Metab Eng *43*, 187-197.

Ludovici M, Kozul N, Materazzi S, Risoluti R, Picardo M, Camera E. Influence of the sebaceous gland density on the stratum corneum lipidome. Sci Rep. 2018 Jul31;8(1):11500.

Mishra, P., Tang, W., Putluri, V., Dorsey, T.H., Jin, F., Wang, F., Zhu, D., Amable, L., Deng, T., Zhang, S., et al. (2018). ADHFE1 is a breast cancer oncogene and induces metabolic reprogramming. J Clin Invest *128*, 323-340.

Palma, A., Grande, S., Rosi, A., Luciani, A.M., Guidoni, L., and Viti, V. (2011). (1)H-MRS can detect aberrant glycosylation in tumor cells: a study of the HeLa cell line. NMR Biomed *24*, 1099- 1110.

Ricci-Vitiani, L., Pallini, R., Larocca, L.M., Lombardi, D.G., Signore, M., Pierconti, F., Petrucci, G., Montano, N., Maira, G., and De Maria, R. (2008). Mesenchymal differentiation of glioblastoma stem cells. Cell Death Differ *15*, 1491-1498.

Singh K, Camera E, et al. JunB defines functional and structural integrity of the epidermo-pilosebaceous unit in the skin. Nat Commun. 2018 Aug 24;9(1):3425.

Terunuma, A., Putluri, N., Mishra, P., Mathe, E.A., Dorsey, T.H., Yi, M., Wallace, T.A., Issaq, H.J., Zhou, M., Killian, J.K., et al. (2014a). MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. J Clin Invest *124*, 398-412.

Terunuma, A., Putluri, N., Mishra, P., Mathe, E.A., Dorsey, T.H., Yi, M., Wallace, T.A., Issaq, H.J., Zhou, M., Killian, J.K., et al. (2014b). MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. J.Clin Invest *124*, 398-412.

Van Gorsel, M., Elia, I., and Fendt, S.M. (2019). (13)C Tracer Analysis and Metabolomics in 3D Cultured Cancer Cells. Methods Mol Biol *1862*, 53-66.