# SUPPLEMENTAL EXPERIMENTAL PROCEDURES AND METABOLITE SPECTRA

**Tissue Collection and Human Breast Cancer Cell Lines.** Patients were recruited in Baltimore hospitals between 1993 and 2003, as previously described (1). They were identified through surgery lists and enrolled into the study prior to surgery. Tumor and adjacent tissue was collected at time of surgery under an established protocol. Samples of macrodissected tumor tissue and adjacent non-cancerous tissue were prepared by a pathologist immediately after surgery and were frozen within 15 min. Tissue quality of the frozen samples, as judged by RNA integrity analysis, was excellent. Specimens were further evaluated on H/E slides and classification as tumor and non-cancerous tissue was confirmed. Clinical and pathological information (e.g., estrogen and progesterone receptor status) was obtained from medical records and pathology reports. Tumor estrogen receptor (ER) status was determined at the Department of Pathology, University of Maryland, consistent with guidelines for clinical laboratories to evaluate semi-quantitatively receptor expression in formalin-fixed, paraffin-embedded tissue ("CONFIRM Estrogen Receptor" assay by Ventana Medical Systems, Tucson, AZ). The immunohistochemical staining protocol for HER2 followed the DAKO HercepTest$^{TM}$ protocol. Tumors were classified as HER2-positive when the HercepTest™ immunohistochemistry score was 3 or when the test score was 2 and the tumor contained a HER2-enriched gene signature based on gene expression profiling according to published criteria (2;3). In the validation cohort, the tumor HER2 status was determined as previously described (4). Triple-negative tumors were negative for estrogen, progesterone, and HER2 receptor expression. Tumors were classified as basal-like based on their gene expression profiles using the PAM50 classifier (2) and/or immunohistochemistry (ER-negative, HER2-

negative, cytokeratin 5/6-positive or EGFR-positive) according to published criteria (5). All patients signed a consent form and completed an interviewer-administered questionnaire that evaluated socio-economic variables as part of a larger survey. Combined annual household income before taxes and deductions was collected as unknown, under \$15,000, between \$15,000 and \$60,000, and above \$60,000. Education information was collected as highest grade or level of schooling and how many years of school were completed. Self-reported race/ethnicity was collected as Black (not of Hispanic origin) for African-Americans and White (not of Hispanic origin) for European-Americans. 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 used to determine the tumor grade.

**Genetic Ancestry Estimation.** Genomic DNA was isolated from either non-cancerous tissue or a blood sample and was genotyped for ancestry informative markers (AIMs). In the discovery cohort, 105 AIMs were genotyped to determine the proportion of European, West African, and Native American genetic ancestry (6). Genotyping was performed using the Sequenom MassARRAY iPLEX platform (Sequenom, San Diego, CA). Individual ancestry estimates were obtained from the genotype data using the Markov Chain Monte Carlo (MCMC) method implemented in the program STRUCTURE 2.1 (7). Our model assumed admixture using prior population information and independent allele frequencies. The MCMC model was run using K = 3 populations (58 Europeans, 67 Native Americans and 62 West Africans) and a burn-in length of 30,000 iterations followed by 70,000 replications. Ancestry estimates were also available for

39 of the African-Americans in the validation cohort. Here, African ancestry was determined with a set of 29 AIMs, as described previously (8). Both methods yielded very similar average African ancestry estimates for the African-American women in this study (83.2% and 81.8%).

**Metabolome Analysis.** Metabolomic profiling of fresh-frozen bulk human breast tissues was performed for two tissue sets. The amount of sample ranged from 33 mg to 109 mg and 47 mg to 109 mg for each set. The discovery set consisted of 67 tumors and 65 non-cancerous tissues (65 tissue pairs). Thirty-four of the tumors were ER-positive and 33 were ER-negative. Patient characteristics for this tissue set are described in Supplementary Table 1. The second tissue set consisted of 70 ER-negative tumors and 36 non-cancerous tissues (36 tissue pairs). Patient characteristics for this validation set are described in Supplementary Table 3. Frozen tissue from 27 tumors and 19 adjacent non-cancerous tissues (total 19 tissue pairs) were analyzed in both the discovery and validation set. Examples of metabolite spectra from these analyses are provided following the reference list for the experimental procedures.

Untargeted metabolic profiling of known and unknown metabolites in the discovery set was performed by Metabolon, Inc., as previously described (9-12). Here, sample preparation was conducted using a proprietary series of organic and aqueous extractions for maximum recovery of small molecules. Median relative standard deviation (RSD) as a measure of instrument and total process variability was 6% and 12%, respectively, for our profiling study. Sample preparation at Metabolon: Samples were prepared using the automated MicroLab STAR® system from Hamilton Company (Reno, NV). A recovery standard was added prior to the first step in the extraction process for quality control (QC) purposes. Sample preparation was conducted using aqueous methanol extraction process to remove the protein fraction while allowing maximum recovery of small molecules. The resulting extract was divided into four fractions: one for analysis by Ultrahigh Performance Liquid Chromatography/Mass Spectroscopy (UPLC/MS/MS) (positive mode), one for UPLC/MS/MS (negative mode), one for Gas Chromatography/Mass Spectroscopy (GC/MS), and one for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. Each sample was then frozen and dried under vacuum. Samples were then prepared for the appropriate instrument, either UPLC/MS/MS or GC/MS. UPLC/MS/MS-based analysis: The LC/MS portion of the platform was based on a Waters ACQUITY UPLC and a Thermo-Finnigan linear trap quadrupole mass spectrometer, which consisted of an electrospray ionization source and linear ion-trap mass analyzer. The sample extract was dried then reconstituted in acidic or basic LC-compatible solvents, each of which contained 8 or more injection standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion optimized conditions and the other using basic negative ion optimized conditions in two independent injections using separate dedicated columns. Extracts reconstituted in acidic conditions were gradient eluted using water and methanol containing 0.1% formic acid, while the basic extracts, which also used water/methanol, contained 6.5 mM ammonium bicarbonate. The MS analysis alternated between MS and data-dependent MS2 scans using dynamic exclusion. GC/MS-based analysis: The samples destined for GC/MS analysis were re-dried under vacuum desiccation for a minimum of 24 hrs prior to being derivatized under dried nitrogen using bistrimethyl-silyl-triflouroacetamide. The GC column was 5% phenyl and the temperature ramp was from 40° to 300° C in a 16 minute period. Samples were analyzed on a

Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization. The instrument was tuned and calibrated for mass resolution and mass accuracy on a daily basis. The information output from the raw data files was automatically extracted as discussed below. Quality assurance/quality control: For QA/QC purposes, additional samples were included with each day's analysis. These samples included extracts of a pool of well-characterized human plasma, extracts of a pool created from a small aliquot of the experimental samples, and process blanks. QC samples were spaced evenly among the injections and all experimental samples were randomly distributed throughout the run. A selection of QC compounds was added to every sample for chromatographic alignment, including those under test. These compounds were carefully chosen so as not to interfere with the measurement of the endogenous compounds. Data extraction and compound identification: Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities.

Metabolic profiling in the validation set was performed for a total of 108 metabolites at the Alkek Center for Molecular Discovery, Baylor College of Medicine and for 6 metabolites at the SAIC-NCI Laboratory of Proteomics and Analytical Technologies. Here, tissues were homogenized in phosphate buffered saline (100 mg tissue/ml PBS) using a micro-homogenizer from Omni International (Kennesaw, GA), and centrifuged (14,000 g/15 minutes). The supernatant was aliquoted and stored at -40° C until use. The pellet was collected for extraction of lipid-based metabolites. Extract aliquots were analyzed with customized assays and MS/MS-based fragmentation for metabolite identification. The LC/MS platform at the Alkek Center has previously been described (13). LC coupled Triple Quadrupole Mass Spectrometry (QQQ; Agilent Technologies, Palo Alto, CA) was used for targeted assessment of compounds with single reaction monitoring strategy. Median RSD for the internal standards and the measured endogenous biochemicals on this platform was 1.8% and 3.7%, respectively. Some compounds, including 2-hydroxyglutarate (2HG), were analyzed using gas chromatography-coupled mass spectrometry (Agilent 7890A and Agilent 7000 QQQ) and a single reaction monitoring strategy with electron-impact ionization. For quantitation of 2HG in tissue extracts and cell lines, a serial dilution of 2HG was mixed with a fixed amount of [D27] myristic acid internal standard so that 0.025, 0.05, 0.5, 2.5, 5 and 50 pmol amounts of 2HG were injected into GC/MS. The peak area ratios of 2HG and [D27] myristic acid were plotted against the amounts of 2HG injected to generate a standard curve. Tissue and cell extracts were mixed similarly with the [D27] myristic acid internal standard, analyzed by GC/MS, and estimated for normalized 2HG levels using this standard curve. We calculated 2HG levels per kg frozen tissue weight and also estimated intracellular 2HG concentrations in breast cancer cell lines (obtained from Rachel Schiff, Baylor College of Medicine) after determining absolute amounts of 2HG in $5\times10^6$ harvested cells and estimating the intracellular 2HG concentration using 1.3 picoliter per cell based on previous estimates for MCF7 cells.

**Proteome Analysis.** Frozen human tissue samples were homogenized in 8 M urea, followed by three rounds of sonication with a 15 second pulse each. After centrifugation of the homogenate at 13,000 rpm for 10 min, the supernatant was loaded on a D-Salt Excellulose desalting column (Pierce Biotechnology, Rockford, IL), eluted with 25 mM $NH_4HCO_3$, pH 8.3, and collected into 0.5 ml fractions. Protein fractions were identified using the Coomassie Plus Assay (Pierce Biotechnology) and were pooled. The total protein concentration of the pooled sample was determined with the BCA protein assay (Pierce Biotechnology). The extracted protein in this sample was digested with trypsin at 37°C overnight using a 1:50 trypsin to protein ratio. The tryptic peptides were desalted using an Empore C18 SPE cartridge (3M). The desalted peptides were fractionated by strong cation chromatography into 45 fractions. The fractions were lyophilized and re-suspended into 0.1% trifluoroacetice acid with a volume that normalized the overall peptide concentration to 0.2 µg/µl. For the LC-MS analysis, fractions 3-43 were combined into 17 samples. MS Analysis: Each combined sample (5 µl) was loaded onto an Agilent 1100 nano-capillary HPLC system (Agilent Technologies) equipped with an in-house made 10 cm integrated nanoRPLC-electrospray ionization emitter column, coupled online with a linear ion trap mass spectrometer (LTQ, Thermo Fisher Scientific, Waltham, MA). After sample injection, a 20 min wash with 98% mobile phase A (0.1% formic acid) was applied, and peptides were eluted using a linear gradient of 2% mobile phase B (0.1% formic acid in acetonitrile) to 42% mobile phase B within 140 min at a constant flow rate of 250 nl/min. The seven most intense molecular ions in the MS scan were sequentially selected for collision-induced dissociation using normalized collision energy of 35%. The mass spectra were acquired over a mass-to-charge (m/z) range of 300–2000. The nano-source capillary voltage and temperature were maintained at 1.75 kV and 200°C, respectively. The MS data were searched against the UniProt Homo sapiens database downloaded from the European Bioinformatics Institute website (http://www.ebi.ac.uk/integr8) using SEQUEST (ThermoElectron/ Thermo Fisher Scientific). Up to two missed tryptic cleavage sites and oxidation of methionyl residues were allowed during the database search. The identifications are considered legitimate with minimum delta correlation (DCn) ≥ 0.08 and charge state dependent cross correlation scores (Xcorr) ≥ 1.9 for [M+H]+1, ≥2.2 for [M+H]+2, and ≥3.1 for [M+H]+3. The database search results were combined for each tissue sample. Peptide counts for each tissue sample were log2 transformed and normalized based on the median value for 626 proteins that were present in more than 90% of all tissue samples. Normalized relative abundance of an identified protein in a sample was used for class comparison analyses.

**RNA Isolation and Gene Expression Microarray Analysis.** Total RNA from 61 tumors and 45 adjacent normal tissues was isolated using TRIzol reagent according to the manufacturer's instructions (Life Technologies, Grand Island, NY). RNA integrity for each sample was confirmed with the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). 250 ng of RNA was converted into cDNA using the Ambion WT Expression Kit for Affymetrix GeneChip Whole Transcript Expression arrays (Life Technologies). After fragmentation and labeling using the GeneChip WT Terminal Labeling Kit from Affymetrix, ssDNA was hybridized onto Gene Chip Human Gene 1.0 ST Arrays according to Affymetrix standard protocols (Santa Clara, CA). The probe cell intensity data was processed by RMA algorithm (14) using Affymetrix Expression Console software, and analyzed using SAM and gene set analysis (GSA) by Tibshirani at Stanford (http://www-stat.stanford.edu/~tibs/GSA/), which is a modified version of gene set enrichment analysis (15). GSA was performed with the collection of gene sets in the Molecular Signatures

Database (http://www.broadinstitute.org/gsea) of the Broad Institute and the human embryonic stem cell gene set by Assou (16). Enrichment for Gene Ontology terms were analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) by National Institute of Allergy and Infectious Diseases (17). Gene expression data from this study were deposited in GEO (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE39004/GSE37751. Affymetrix Platform: GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version].

**Global DNA Methylation Analysis.** To obtain global methylation profiles for genomic DNA, 500 ng genomic DNA was bisulfite converted using the EZ DNA Methylation Gold kit (Zymo Research Corporation, Irvine, CA) and a modified incubation method was employed following the manufacturer's recommendations (Illumina, San Diego, CA). Approximately 120 ng of bisulfite converted DNA were processed through the Infinium HD Methylation Assay (Illumina) and samples were hybridized onto Human Methylation 450 BeadChips (Illumina). The single-base extension and staining portion of the assay was performed on a Tecan Evo robot (Tecan Group Ltd, Mannedorf, Switzerland). The BeadChips were subsequently scanned on an Illumina HiScan SQ and image data extracted and initially analyzed using Genome Studio v2011.1. The data was further analyzed using R packages: *lumi* for 450 BeadChip and *methylumi* for Golden Gate. Beta-values were converted to M-values for statistical analysis. The DNA methylation profiles from this study were deposited in GEO (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE39004/GSE37754. Illumina Platform: GPL13534 Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482_v.1.1).

**Measurement of DNA 5-hydroxymethylcytosine/5-methylcytosine Content by LC/MS.** Experiments were performed as described previously (18). MCF7 and MCF10A cells were cultured in medium with D-2HG or L-2HG (Santa Cruz Biotechnology) and genomic DNA was extracted using a commercial kit (QIAGEN, Valencia, CA) following the vendor's protocol. Human methylated and non-methylated DNA purchased from Zymo Research Corporation (Orange, CA) served as controls for the experiment. For the mass spectrometry-based assessment of cytosine methylation/hydroxy methylation, genomic DNA (1 µg) was denatured by heating at 100°C for 3 minutes followed by chilling on ice. Ammonium acetate (0.1M, one-tenth volume, pH 5.3) and 2 units of nuclease P1 (Sigma Aldrich Corporation, St. Louis, MO) were first added and incubated at 45°C for 2 hrs. Following this, ammonium bicarbonate (1M, one-tenth volume) and 0.002 units of snake venom phosphodiesterase I (Sigma Aldrich) were added and the incubation was continued for an additional 2 hrs at 37 °C. Subsequently, 0.5 units of alkaline phosphatase (Sigma Aldrich) was added and incubated for a further 1 hr at 37 °C. The processed genomic DNA sample was then dried (Genevac EZ-2$^{plus}$, Gardiner, NY), resuspended in identical volume of water:acetonitrile injection solvent (98:2) with 0.1% formic acid, and separated using a RP chromatographic gradient and analyzed for the quantitative content of 5-hydroxymethylcytosine and 5-methylcytosine against standards using a QQQ mass spectrometer.

**Mutational Analyses of IDH1, IDH2, and p53.** Genomic DNA was isolated from fresh-frozen tumor tissue following the manufacturer's protocol for the QIAGEN DNeasy blood and tissue kit (QIAGEN, Valencia, CA). 40 nanogram of DNA was amplified using MyTaq HS Red DNA

Polymerase (Bioline, Tauton, MA), and 10 μM primers for *IDH1* and *2*. The primer sequences for analyzing the *IDH1* and *IDH2* mutational hotspots (covering amino acid residue 41-138 for *IDH1* and 125-226 for *IDH2*) were previously published (19). Samples were loaded on an agarose gel and bands were cut out and purified using the QIAGEN MinElute Gel extraction kit. Sequencing was performed by the NCI Core using the same primers as for amplification, performing reverse only and forward only reactions for each sample. Tumors were screened for p53 mutations as previously described (1).

**siRNA Transfection.** Cells were transfected with either negative control siRNA (*Silencer*® Select Negative Control No. 1), ADHFE1, IDH2, GLS1, or D2HGDG Select siRNAs (Life Technologies, Carlsbad, CA) following the manufacturer's protocol for lipofectamine RNAiMAX (Life Technologies). After 48 hrs, cell pellets were prepared to measure intracellular 2HG ($5x10^6$ cells: trysinized, counted using the TC10 automatic cell counter from Bio-Rad, and pelleted), or protein extracts were prepared for Western blot analysis of either ADHFE1 or IDH2 expression.

**Quantitative Real-Time PCR.** Total RNA was extracted from either breast tissue using the TRIzol method or from siRNA-transfected cell lines using the QIAGEN RNeasy method and subjected to reverse transcription for quantitative PCR (qRT-PCR). qRT-PCR was subsequently performed in triplicate using the TaqMan *ASPA* (aspartoacylase; EC 3.5.1.15), *IDH2* (isocitrate dehydrogenase 2, mitochondrial; EC 1.1.1.42), *ADHFE1* (alcohol dehydrogenase iron containing 1 or hydroxyacid-oxoacid transhydrogenase, mitochondrial; EC 1.1.99.24), *GLS1* (glutaminase; EC 3.5.1.2), and *D2HGDG* (2-D-hydroxyglutarate dehydrogenase; EC 1.1.99.2) expression assays (Applied Biosystems, Foster City, CA), which included pre-optimized probes and primer sets for these genes. Data were collected using the ABI PRISM® 7500 Sequence Detection System. The 18s RNA was used as the internal standard reference. Normalized expression was calculated using the comparative $C_t$ method and fold changes were derived from the $2^{-\Delta\Delta Ct}$ values for each gene.

**Western Blot Analysis.** Whole cell lysates were prepared in radioimmunoprecipitation assay buffer (RIPA; Sigma Aldrich, St. Louis, MO) and 25–100 μg protein per lane were separated on SDS-polyacrylamide gels and transferred to nitrocellulose membranes using an iBlot Dry Blotting System (Invitrogen). Mitochondrial lysates were prepared from isolated mitochondria (Mitochondria/Cytosol Fractionation Kit, Millipore, Billerica, MA). Protein concentrations were determined with the Bio-Rad Protein Assay (Bio-Rad Laboratories, Hercules, CA). The following antibodies were used at concentrations recommended by the manufacturers to detect the membrane bound proteins: rabbit polyclonal anti-ADHFE1 antibody (H-240, #sc-292533) from Santa Cruz Biotechnology; rabbit polyclonal anti-ADHFE1 antibody (#ab102600) from Abcam (Cambridge, MA); mouse monoclonal anti-IDH2 (#ab55271) from Abcam; mouse monoclonal antibodies anti-vinculin, Sigma V9131 (1:4000) and anti-c-myc, Sigma M4439 (1:1000) from Sigma Aldrich; mouse monoclonal anti-IDH2 (#ab55271) from Abcam; mouse monoclonal anti-β-actin (#CP01) from CalbioChem (San Diego, CA). Signals were visualized after incubation with recommended secondary antibody conjugated to peroxidase and protein bands were visualized using Pierce ECL western blotting substrate (Thermo Fisher Scientific, Rockford, IL).

**Migration and Invasion Assay.** Migration and invasion were examined using xCelligence System Technology (Roche Diagnostics, Indianapolis, IN) for real-time monitoring of cellular processes with electronic cell sensor array technology, according to the manufacturer's instructions. Cells were plated on 100 mm dishes and transfected with ADHFE1 Silencer Select siRNA (Life Technologies, Carlsbad, CA) following the manufacturer's protocol for lipofectamine RNAiMAX (Life Technologies). 24 hrs later, cells were trypsinized, counted and plated onto CIM-Plate 16, and the migratory and the invasive capacity of cells were determined using xCelligence. For the assay, the membrane in the top chamber of the CIM plate was coated with 30 µl of a 1:20 dilution of Matrigel (BD Biosciences, San Jose, CA) in serum-free medium and incubated at $37^o$ for 4 hrs. Subsequently, MDA-MB-231 (50,000) or SUM159T (30,000) cells were added to this chamber (in triplicate wells with matrigel for invasion or without matrigel for migration) in serum-free media and the lower chamber was filled with 10% FBS cell culture media, or serum-free medium as a negative control. After 30 minutes at room temperature, the loaded CIM-Plate 16 was placed into the xCelligence analyzer and electrical impedance was measured every 15 minutes over a 48 hr period. Invasion index was calculated from the cell index ratio for Matrigel-coated wells and uncoated wells. Cell index was determined per manufacturer's instructions.

**Statistical Analysis.** All statistical tests were two-sided and an association was considered statistically significant with $P < 0.05$. Statistical analyses were performed using the R software developed by R Development Core Team at R Foundation for Statistical Computing and packages in Bioconductor (20). The abundance of metabolites and transcripts were analyzed globally using significance analysis of microarrays (SAM) (21). Heatmaps were drawn using dChip software (22). Survival analysis was performed using the *survival* package of R. The association between the clinical outcome and a molecular signature (e.g., gene expression or DNA methylation signatures) was investigated using the sum of z scores for the set of molecules with the median as the cutoff. To associate DNA methylation patterns with survival, tumors in the discovery set were divided into three subgroups (I, II and III) by clustering them using the DNA methylation M-values of the highest variation between tumors (standard deviation >1) and of the highest absolute value for each gene. Probes that showed distinct profiles between subgroups III vs. I were used to divide tumor cases in other datasets into two subgroups (subgroup III signature-high and low). Six publicly available datasets of gene expression microarray were included in the survival analysis. The dataset by van de Vijver (23) was obtained from Netherlands Cancer Institute. The datasets by Kao, Pawitan, Miller, and Desmedt (24-27) were obtained from Gene Expression Omnibus (GEO) by National Center for Biotechnology Information with the GEO accession of GSE20685, GSE1456, GSE3494 and GSE7390, respectively. The dataset by Chin (28) was obtained from EMBL-EBI by European Bioinformatics Institute with the accession of E-TABM-158. The datasets by Pawitan, Miller, Desmedt and Chin were combined into a single matrix by scaling the expression values for individual genes using the average in each dataset and the average across all datasets ([a raw value] x [average of all]/[average of the dataset]). A dataset for global DNA methylation in breast tumors, GSE20713 (29), was obtained from GEO. Gene signatures to classify tumors by subtype (PAM50), MYC gene expression (Core MYC gene expression signature), and PIK3CA mutational status (PIK3CA mutation-associated gene signature) were obtained from the corresponding publications (2;30;31). Box plots in figures depict the median value and interquartile range (IQR) by the box, the maximum and minimum values within the range from the first quartile minus IQR x 1.5 to the third quartile plus IQR x 1.5 by whiskers, and extreme values outside that range by open circles.

## References

1. Boersma,B.J., Howe,T.M., Goodman,J.E., Yfantis,H.G., Lee,D.H., Chanock,S.J., and Ambs,S. 2006. Association of breast cancer outcome with status of p53 and MDM2 SNP309. *J Natl Cancer Inst* 98:911-919.

2. Parker,J.S., Mullins,M., Cheang,M.C., Leung,S., Voduc,D., Vickery,T., Davies,S., Fauron,C., He,X., Hu,Z. *et al*. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J.Clin Oncol.* 27:1160-1167.

3. The Cancer Genome Atlas Network 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490:61-70.

4. Glynn,S.A., Prueitt,R.L., Ridnour,L.A., Boersma,B.J., Dorsey,T.M., Wink,D.A., Goodman,J.E., Yfantis,H.G., Lee,D.H., and Ambs,S. 2010. COX-2 activation is associated with Akt phosphorylation and poor survival in ER-negative, HER2-positive breast cancer. *BMC.Cancer* 10:626.

5. Nielsen,T.O., Hsu,F.D., Jensen,K., Cheang,M., Karaca,G., Hu,Z., Hernandez-Boussard,T., Livasy,C., Cowan,D., Dressler,L. *et al*. 2004. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 10:5367-5374.

6. Worsham,M.J., Divine,G., and Kittles,R.A. 2011. Race as a social construct in head and neck cancer outcomes. *Otolaryngol.Head Neck Surg.* 144:381-389.

7. Falush,D., Stephens,M., and Pritchard,J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567-1587.

8. Ruiz-Narvaez,E.A., Rosenberg,L., Wise,L.A., Reich,D., and Palmer,J.R. 2011. Validation of a small set of ancestral informative markers for control of population admixture in African Americans. *Am.J.Epidemiol.* 173:587-592.

9. Sreekumar,A., Poisson,L.M., Rajendiran,T.M., Khan,A.P., Cao,Q., Yu,J., Laxman,B., Mehra,R., Lonigro,R.J., Li,Y. *et al*. 2009. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457:910-914.

10. Evans,A.M., DeHaven,C.D., Barrett,T., Mitchell,M., and Milgram,E. 2009. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal.Chem.* 81:6656-6667.

11. DeHaven,C.D., Evans,A.M., Dai,H., and Lawton,K.A. 2010. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J.Cheminform.* 2:9.

12. Reitman,Z.J., Jin,G., Karoly,E.D., Spasojevic,I., Yang,J., Kinzler,K.W., He,Y., Bigner,D.D., Vogelstein,B., and Yan,H. 2011. Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. *Proc.Natl.Acad.Sci.U.S.A* 108:3270-3275.

13. Putluri,N., Shojaie,A., Vasu,V.T., Vareed,S.K., Nalluri,S., Putluri,V., Thangjam,G.S., Panzitt,K., Tallman,C.T., Butler,C. *et al*. 2011. Metabolomic Profiling Reveals Potential Markers and Bioprocesses Altered in Bladder Cancer Progression. *Cancer Res* 71:7376-7386.

14. Irizarry,R.A., Hobbs,B., Collin,F., Beazer-Barclay,Y.D., Antonellis,K.J., Scherf,U., and Speed,T.P. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 4:249-264.

15. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al*. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A* 102:15545-15550.

16. Assou,S., Le,C.T., Tondeur,S., Strom,S., Gabelle,A., Marty,S., Nadal,L., Pantesco,V., Reme,T., Hugnot,J.P. *et al*. 2007. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* 25:961-973.

17. Dennis,G., Jr., Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C., and Lempicki,R.A. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4:3.

18. Friso,S., Choi,S.W., Dolnikowski,G.G., and Selhub,J. 2002. A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Anal.Chem.* 74:4526-4531.

19. Ward,P.S., Patel,J., Wise,D.R., bdel-Wahab,O., Bennett,B.D., Coller,H.A., Cross,J.R., Fantin,V.R., Hedvat,C.V., Perl,A.E. *et al*. 2010. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* 17:225-234.

20. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al*. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.

21. Tusher,V.G., Tibshirani,R., and Chu,G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc.Natl.Acad.Sci.U.S.A* 98:5116-5121.

22. Li,C. and Wong,W.H. 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc.Natl.Acad.Sci.U.S.A* 98:31-36.

23. Van de Vijver,M.J., He,Y.D., Van't Veer,L.J., Dai,H., Hart,A.A., Voskuil,D.W., Schreiber,G.J., Peterse,J.L., Roberts,C., Marton,M.J. *et al*. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N.Engl.J Med.* 347:1999-2009.

24. Kao,K.J., Chang,K.M., Hsu,H.C., and Huang,A.T. 2011. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC.Cancer* 11:143.

25. Pawitan,Y., Bjohle,J., Amler,L., Borg,A.L., Egyhazi,S., Hall,P., Han,X., Holmberg,L., Huang,F., Klaar,S. *et al*. 2005. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 7:R953-R964.

26. Miller,W.R., Larionov,A.A., Renshaw,L., Anderson,T.J., White,S., Murray,J., Murray,E., Hampton,G., Walker,J.R., Ho,S. *et al*. 2007. Changes in breast cancer transcriptional profiles after treatment with the aromatase inhibitor, letrozole. *Pharmacogenet.Genomics* 17:813-826.

27. Desmedt,C., Piette,F., Loi,S., Wang,Y., Lallemand,F., Haibe-Kains,B., Viale,G., Delorenzi,M., Zhang,Y., d'Assignies,M.S. *et al*. 2007. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin.Cancer Res.* 13:3207-3214.

28. Chin,K., DeVries,S., Fridlyand,J., Spellman,P.T., Roydasgupta,R., Kuo,W.L., Lapuk,A., Neve,R.M., Qian,Z., Ryder,T. *et al*. 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10:529-541.

29. Dedeurwaerder,S., Desmedt,C., Calonne,E., Singhal,S.K., Haibe-Kains,B., Defrance,M., Michiels,S., Volkmar,M., Deplus,R., Luciani,J. *et al*. 2011. DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol.Med.* 3:726-741.

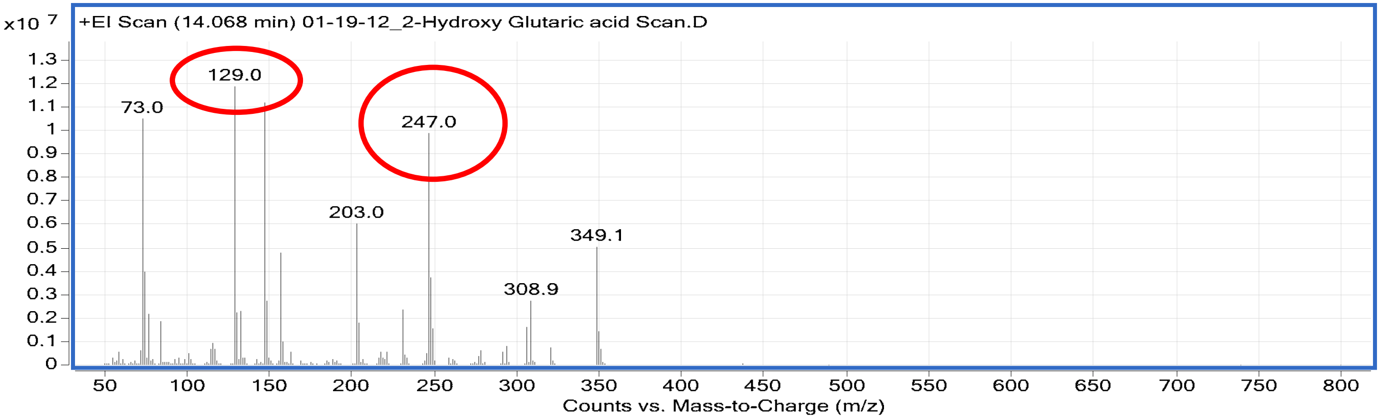
30. Chandriani,S., Frengen,E., Cowling,V.H., Pendergrass,S.A., Perou,C.M., Whitfield,M.L., and Cole,M.D. 2009. A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS.ONE.* 4:e6693.

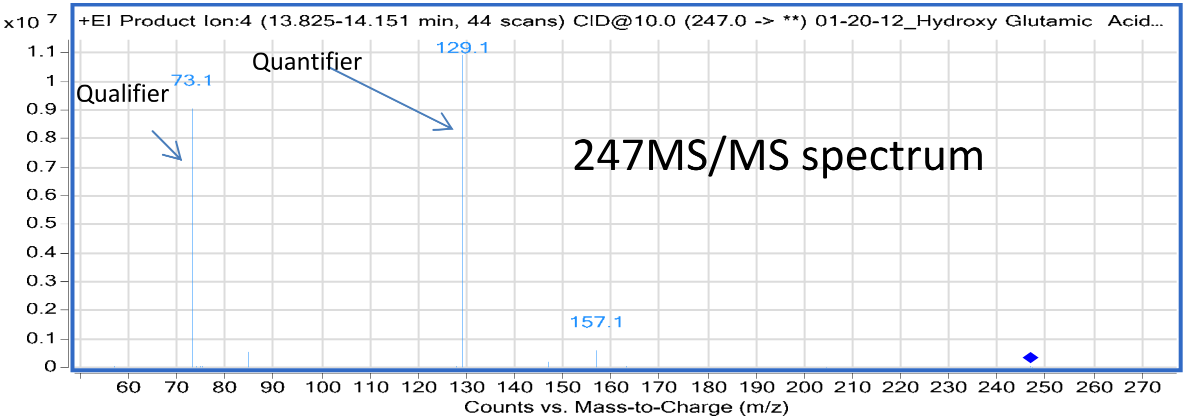
31. Loi,S., Haibe-Kains,B., Majjaj,S., Lallemand,F., Durbecq,V., Larsimont,D., Gonzalez-Angulo,A.M., Pusztai,L., Symmans,W.F., Bardelli,A. *et al*. 2010. PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor-positive breast cancer. *Proc.Natl.Acad.Sci.U.S.A* 107:10208-10213.

# Examples of Metabolite Spectra

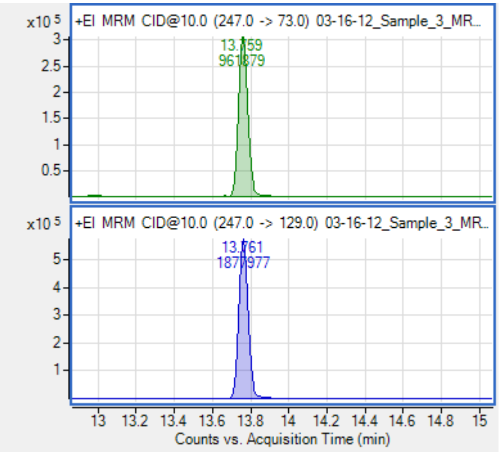
- Alkek Center for Molecular Discovery at Baylor College of Medicine
- Laboratory of Proteomics and Analytical Technologies (LPAT/NCI)
- Metabolon Inc

# 2-Hydroxyglutarate (2HG) EI Spectra (Baylor)



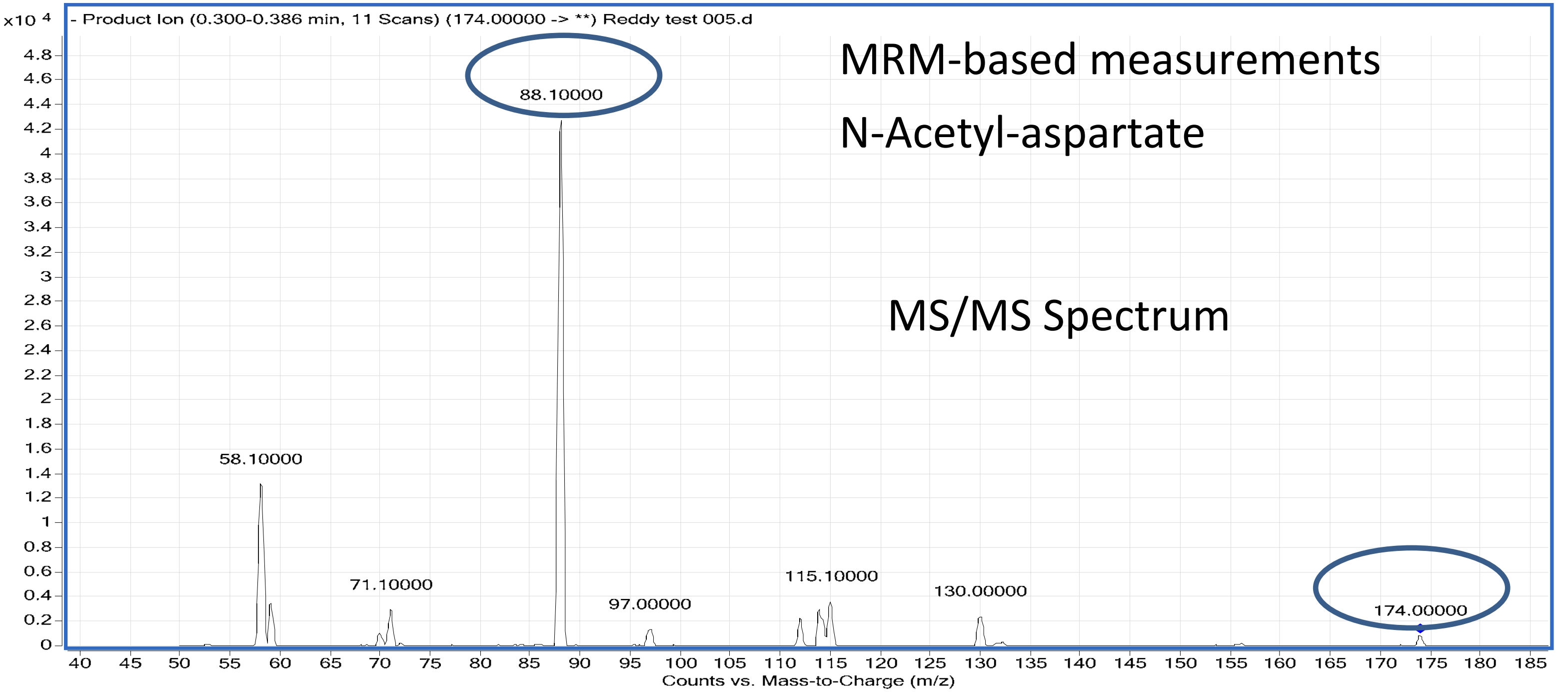


GC/MS RT=13.7



## MRM transitions used to measured 2HG

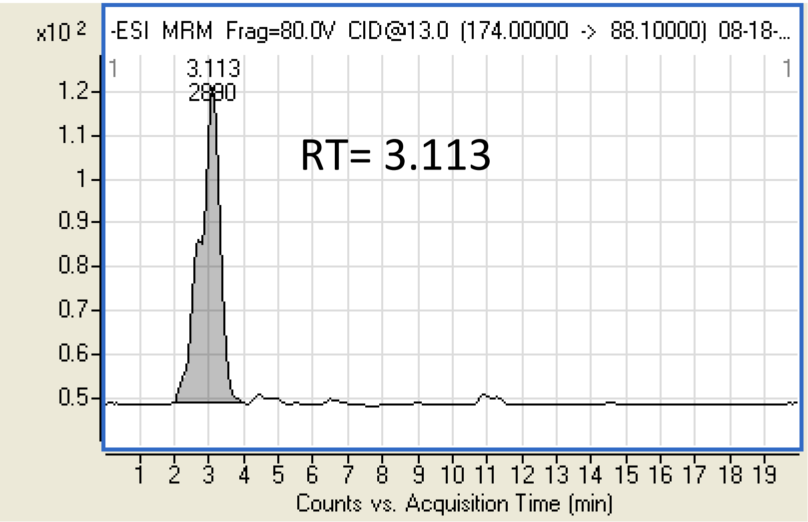
| | | | | |
|---|---|---|---|---|
| **2-Hydroxyglutarate** | **247** | **129** | **100** | **10** |
| **2-Hydroxy glutarate** | **247** | **73** | **100** | **10** |



MRM-based measurements

N-Acetyl-aspartate

MS/MS Spectrum

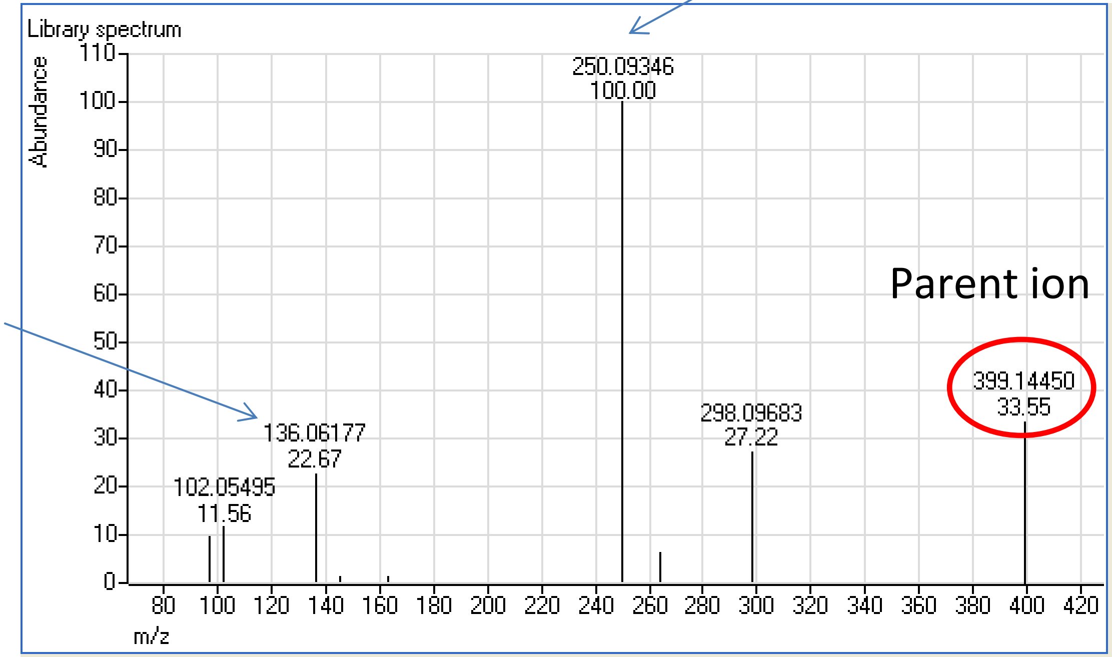


RT= 3.113

**N-Acetyl-aspartate (NAA) spectra (Baylor)**

| Compound Name | ISTD? | Precursor Ion | MS1 Res | Product Ion | MS2 Res | Dwell | Fragmentor | Collision Energy | Cell Accelerator Voltage | Polarity |
|---|---|---|---|---|---|---|---|---|---|---|
| NAA | FALSE | 174 | Unit | 88 | Unit | 75 | 80 | 9 | 7 | Negative |

# S-Adenosine-methionine (SAM) spectrum (Baylor)



Product ion 1 quantifier

Product ion 2 Qualifier

Parent ion

## MRM Transitions

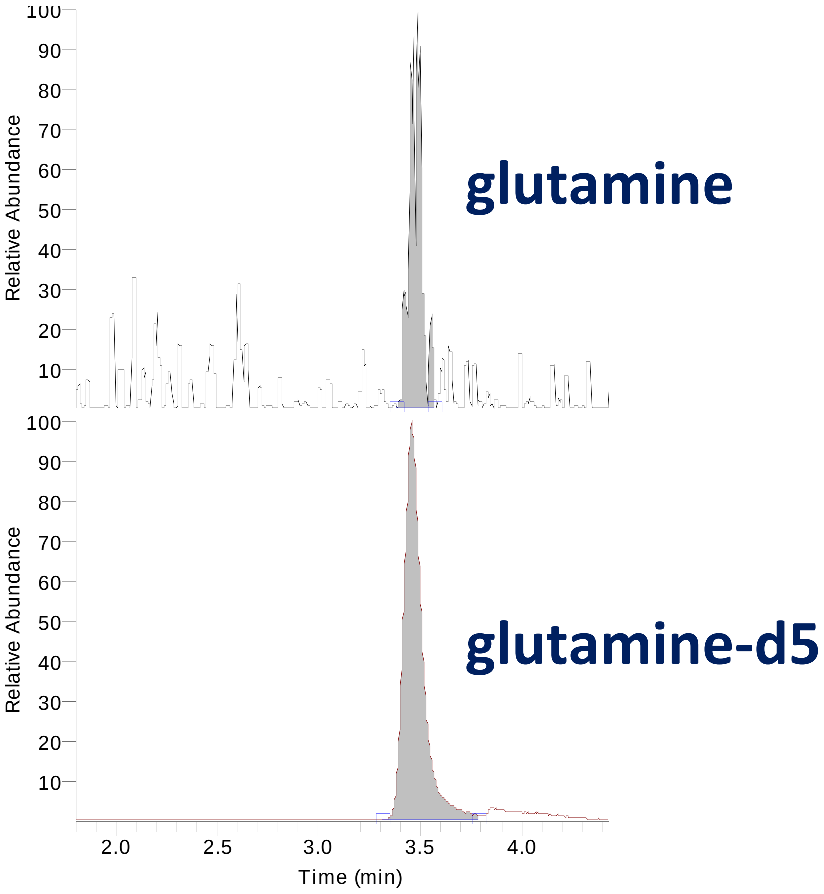
MRM

| Compound Name | ISTD | Precursor Ion | MS1 Res | Product Ion | MS2 Res | Dwell | Fragmentor | Collision Energy | Cell Accelerator Voltage | Polarity |
|---|---|---|---|---|---|---|---|---|---|---|
| SAM | FALSE | 399.2 | Unit | 250.1 | Unit | 75 | 114 | 8 | 7 | Positive |
| SAM | FALSE | 399.2 | Unit | 136.1 | Unit | 75 | 114 | 20 | 7 | Positive |

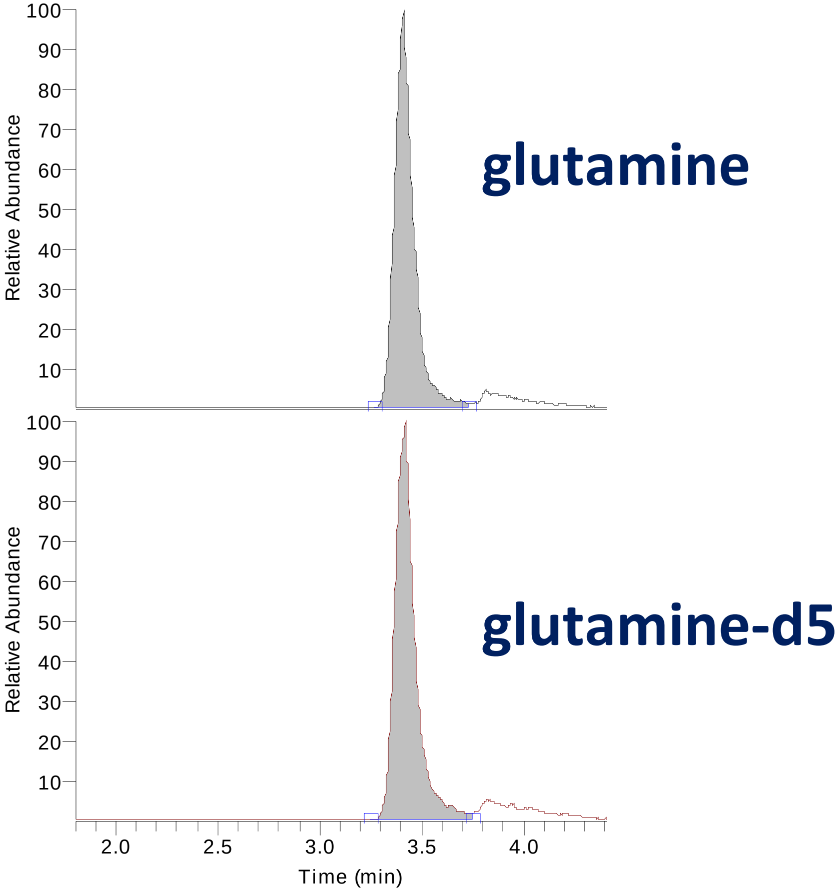
# Glutamine (LPAT/National Cancer Institute)

## Standard

### 0.15 µM Gln



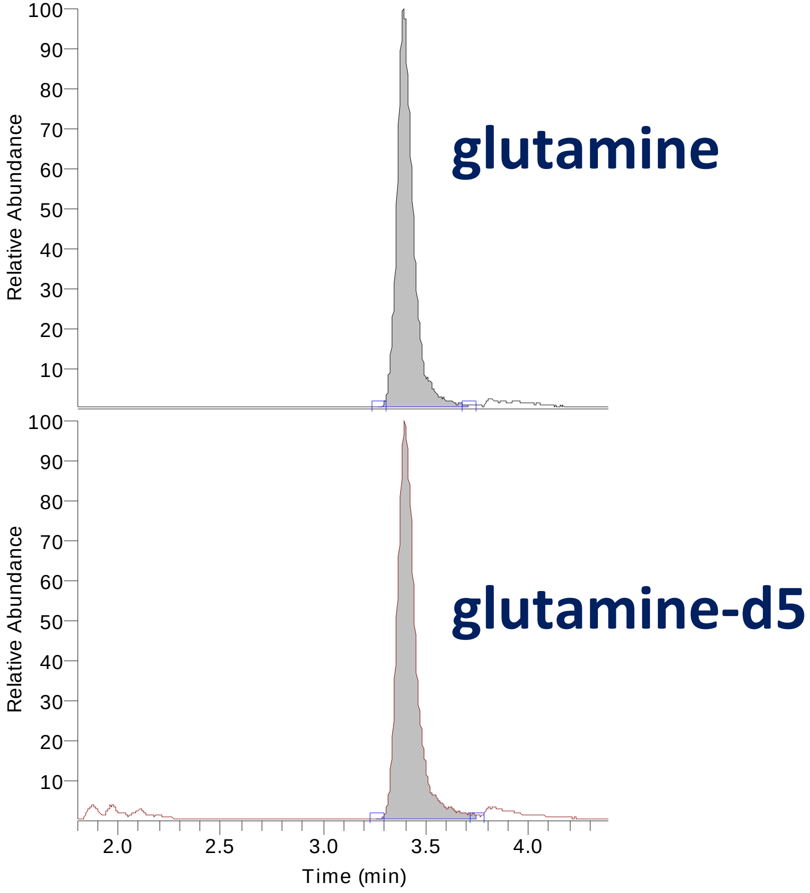
### 111 µM Gln



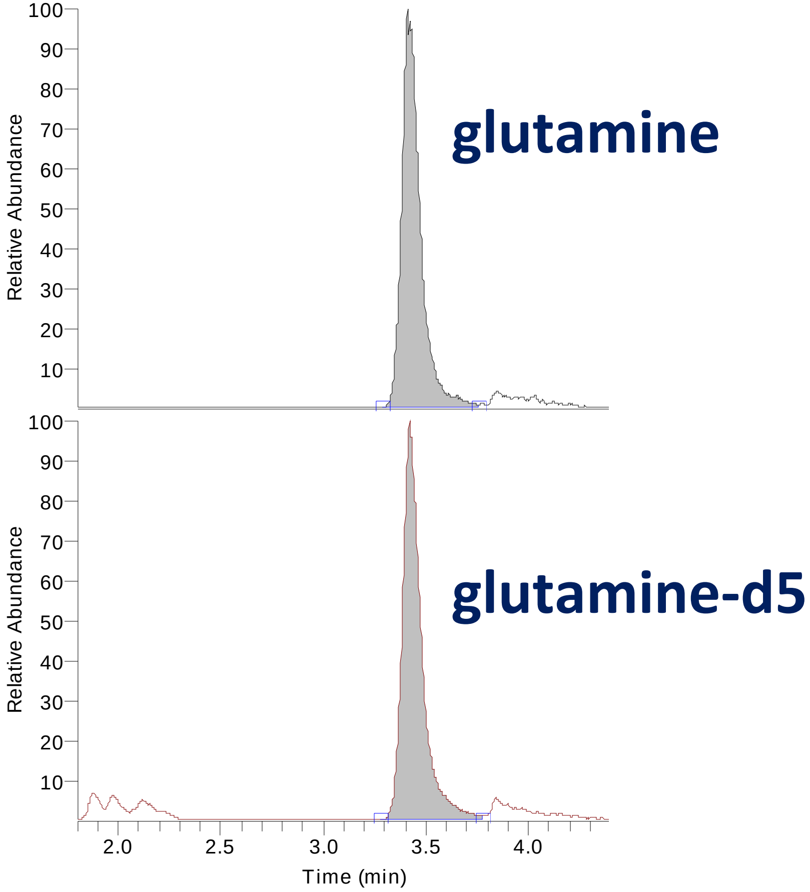
# Glutamine (LPAT/National Cancer Institute)

## Tissue Extracts

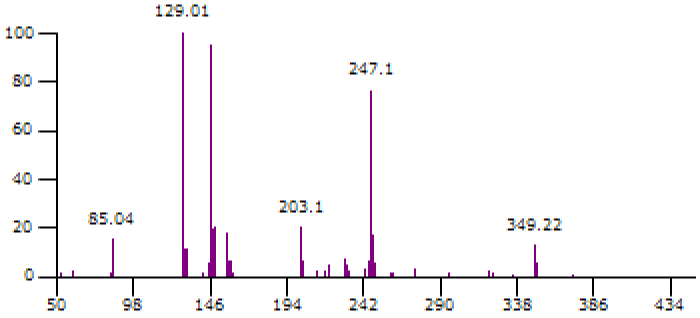
### #45



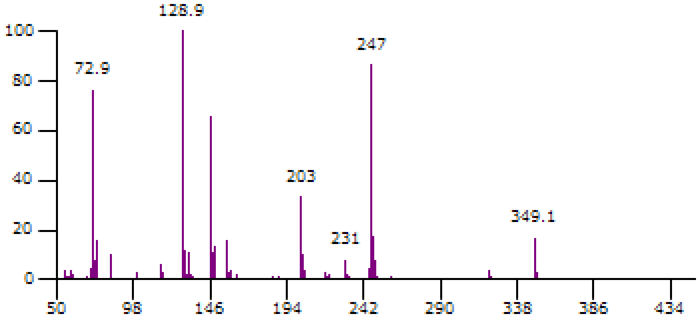
### #91



# 2-Hydroxyglutarate
# (Metabolon Inc)

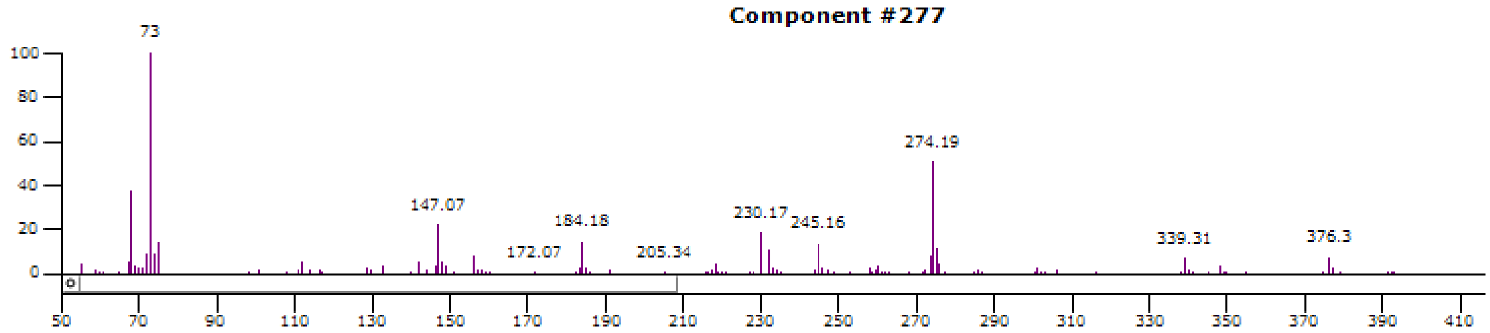


EI (electron ionization)
fragmentation pattern

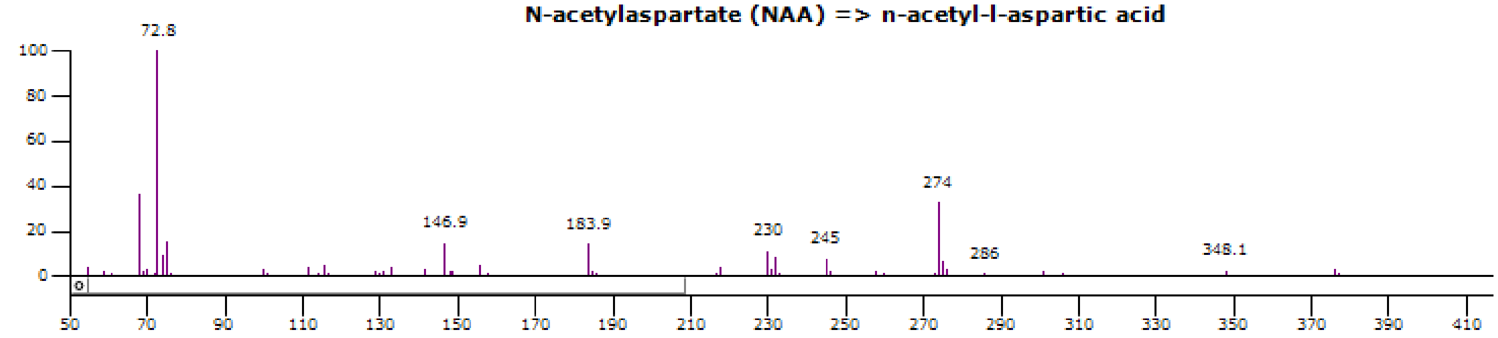


spectrum of authentic standard

# N-Acetyl-aspartate (Metabolon Inc)

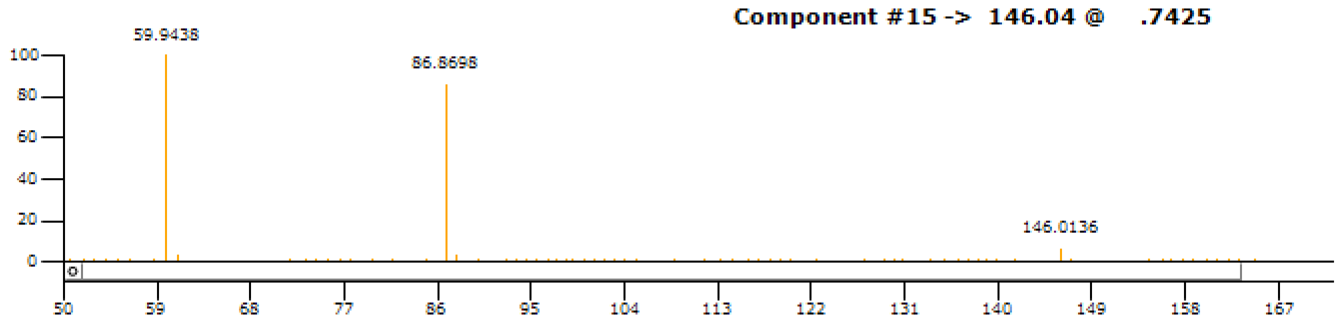

Component #277

EI (electron ionization) fragmentation pattern

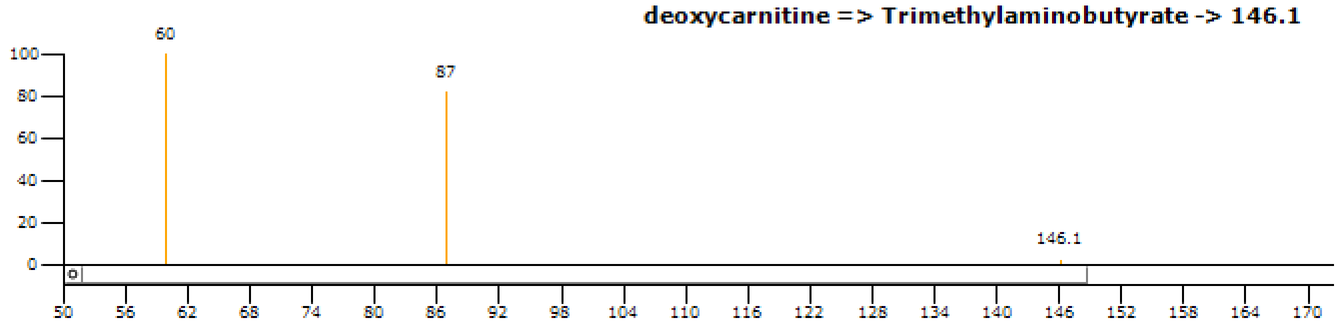

N-acetylaspartate (NAA) => n-acetyl-l-aspartic acid

spectrum of authentic standard

# Deoxycarnitine
# (Metabolon Inc)

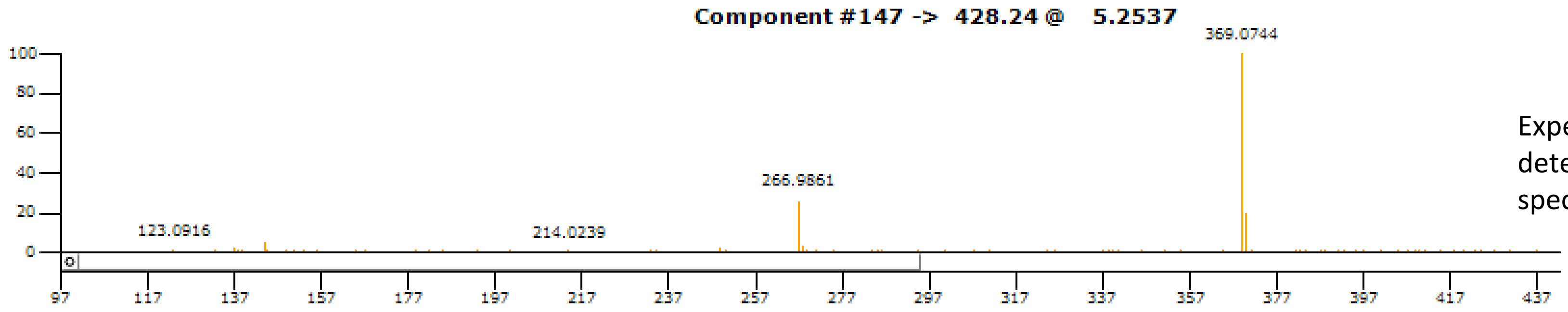


Experimentally detected MS/MS spectrum

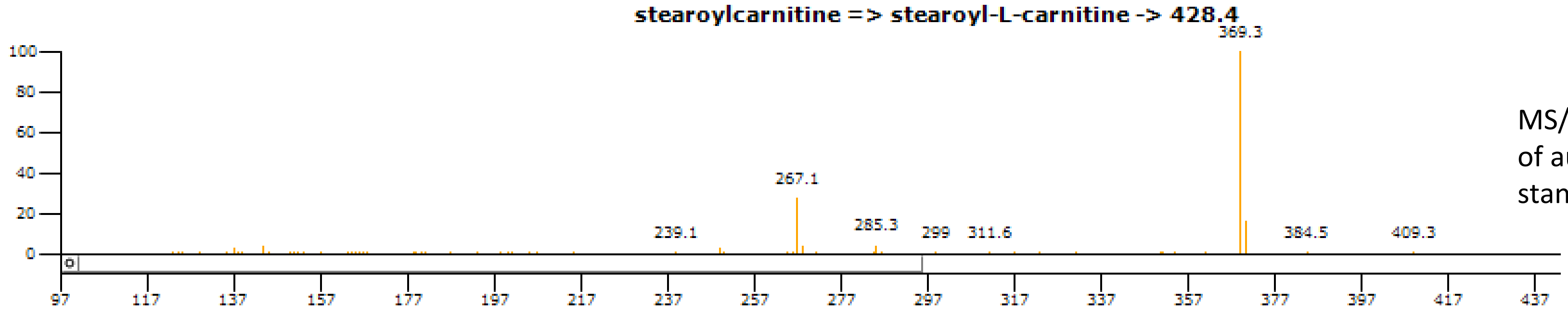


MS/MS spectrum of authentic standard

# Stearoylcarnitine
# (Metabolon Inc)

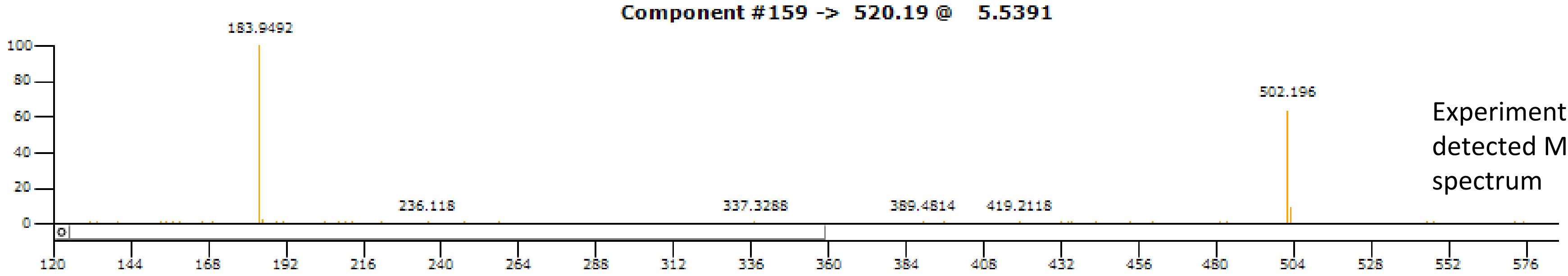


Experimentally detected MS/MS spectrum

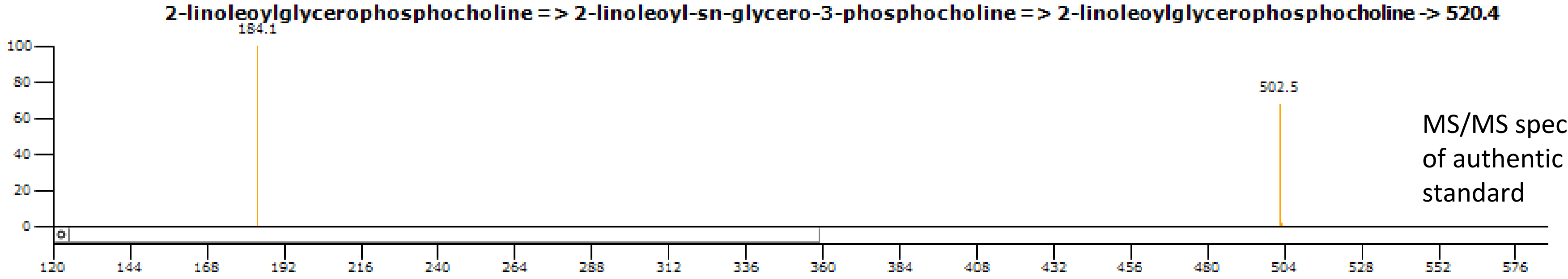


MS/MS spectrum of authentic standard

# 2-Linoleoylglycerophosphocholine (Metabolon Inc)

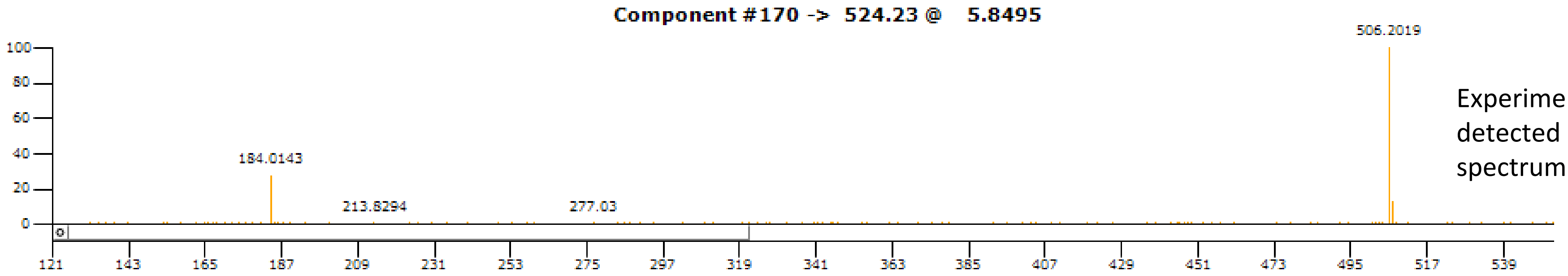

Component #159 -> 520.19 @ 5.5391

Experimentally detected MS/MS spectrum


2-linoleoylglycerophosphocholine => 2-linoleoyl-sn-glycero-3-phosphocholine => 2-linoleoylglycerophosphocholine -> 520.4

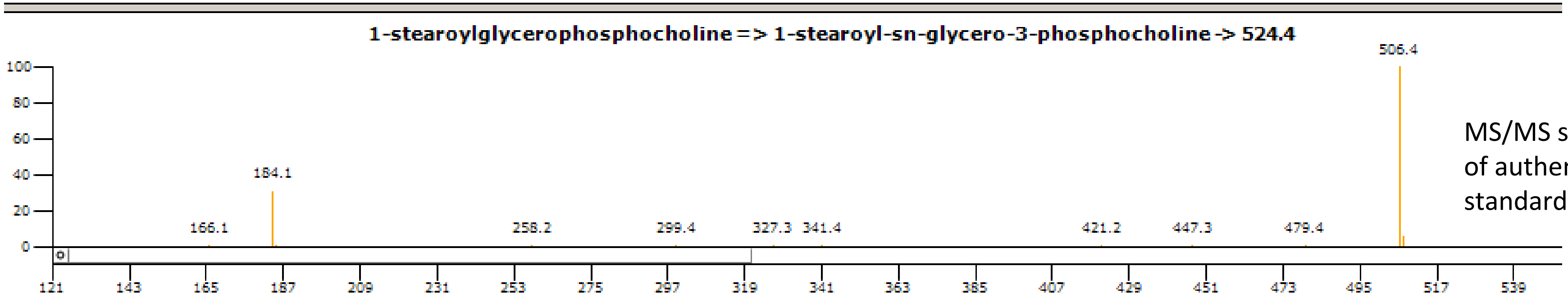
MS/MS spectrum of authentic standard

# 1-Stearoylglycerophosphocholine (Metabolon Inc)



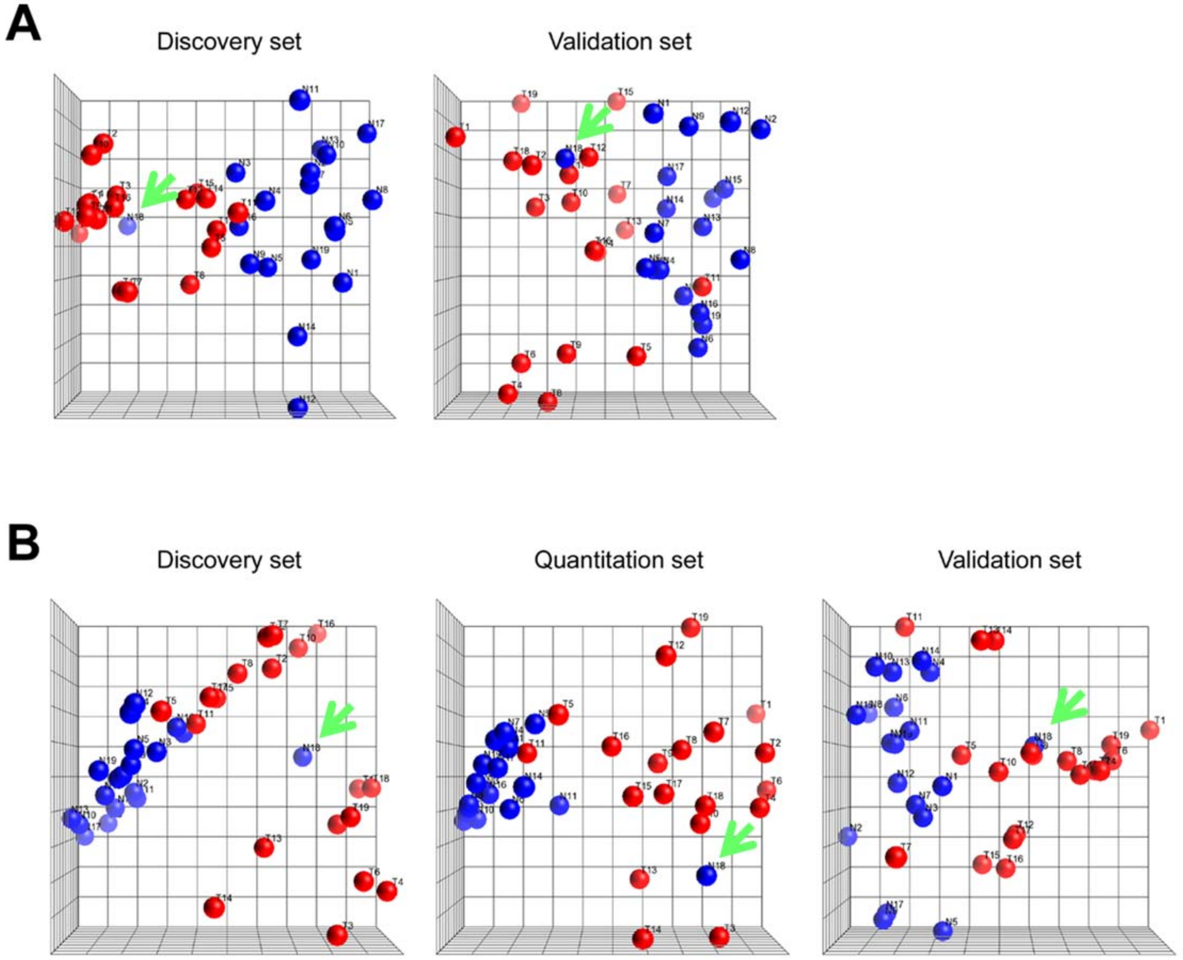
Component #170 -> 524.23 @ 5.8495

Experimentally detected MS/MS spectrum

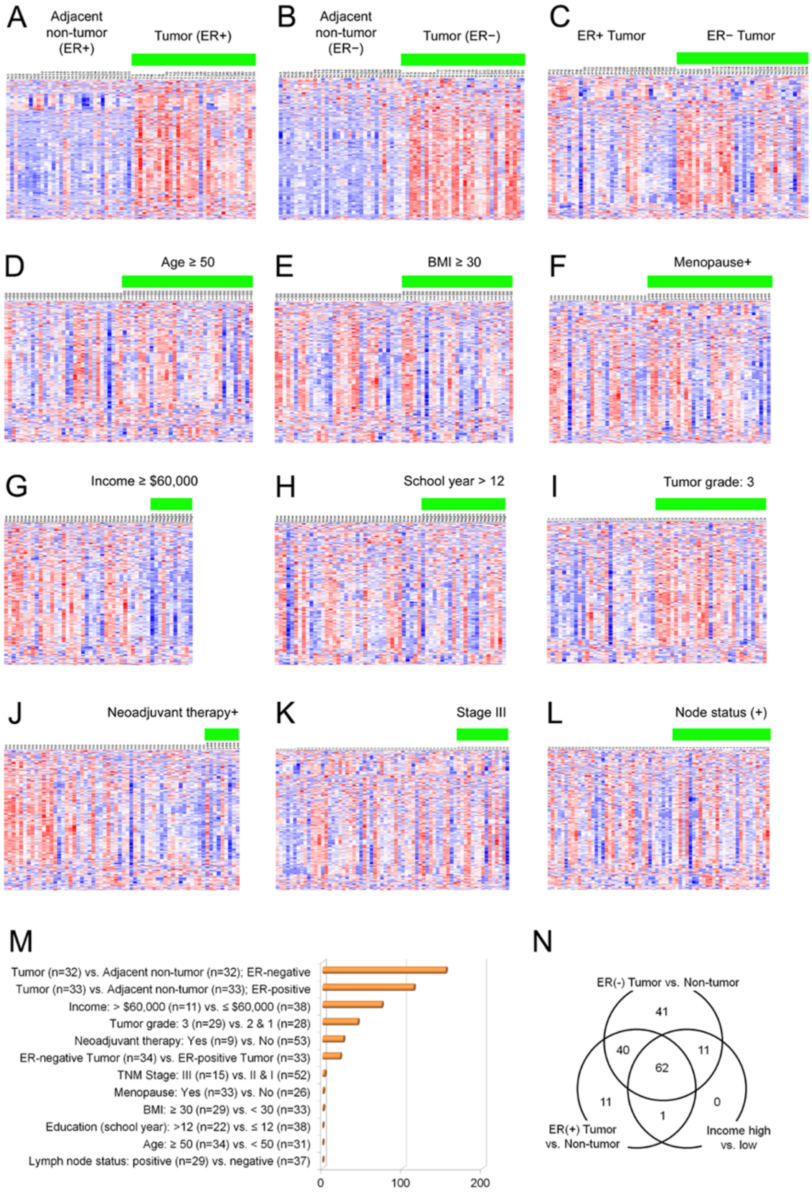


1-stearoylglycerophosphocholine => 1-stearoyl-sn-glycero-3-phosphocholine -> 524.4

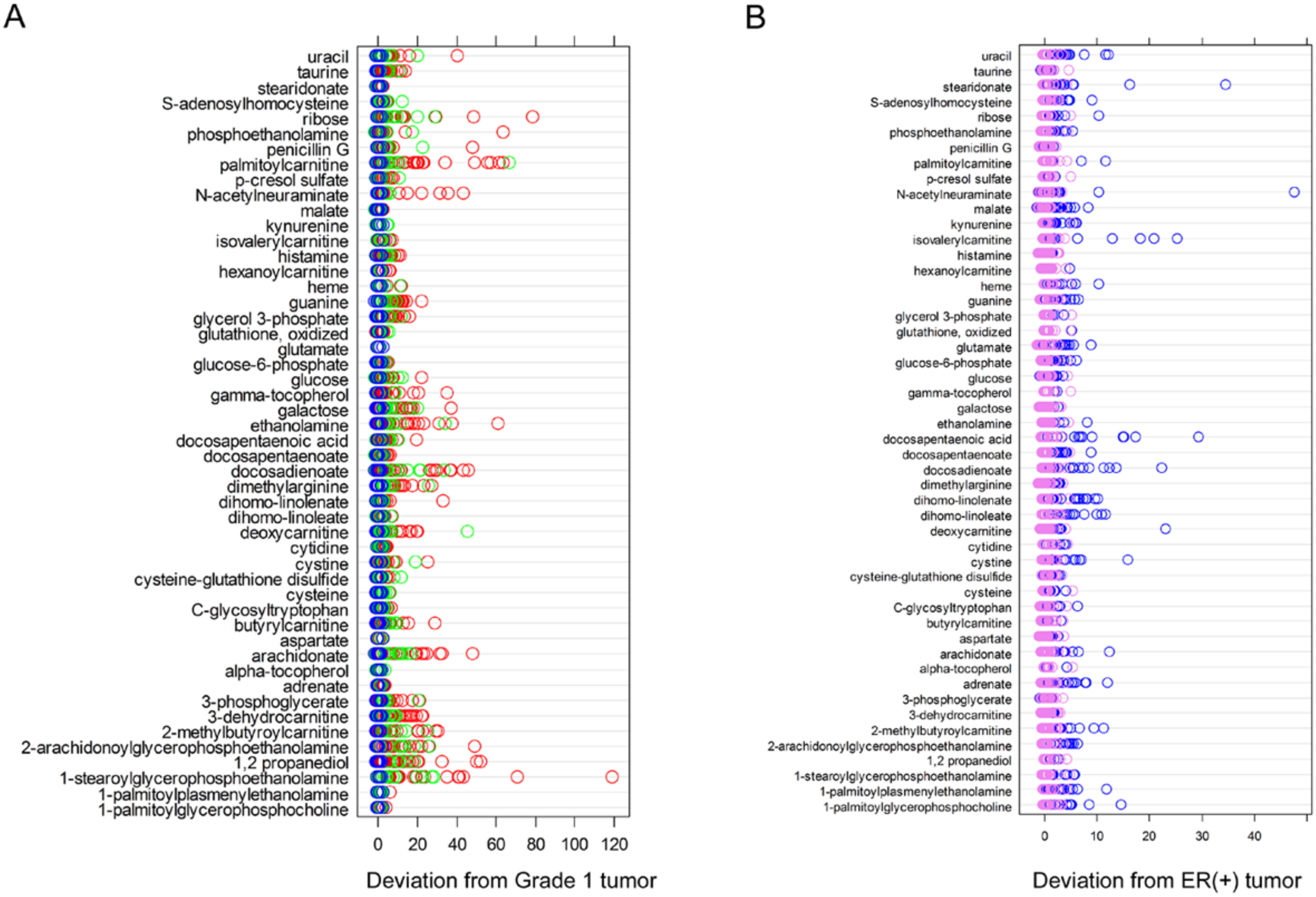
MS/MS spectrum of authentic standard



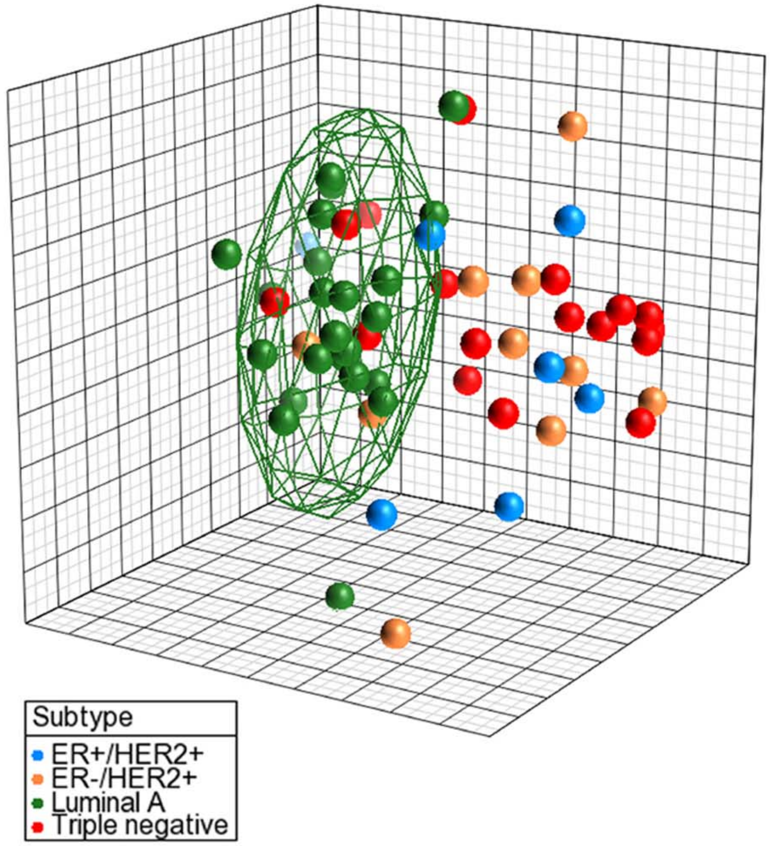
**Supplemental Figure 1. Concordant classification of tumor-adjacent non-cancerous tissue pairs using three independent platforms. A**: Shown is a 50 metabolites-based principal component analysis (PCA) for 19 pairs of estrogen receptor (ER)-negative tumors (red; T1-T19) and adjacent non-cancerous tissue (blue; N1-19) that were analyzed in both the discovery set (left) and the validation set (right). **B**: PCA using 5 metabolites (dimethylarginine, kynurenine, sarcosine, alanine and beta-alanine) for 19 pairs of ER-negative tumors (red; T1-T19) and adjacent non-cancerous tissues (blue; N1-19) that were analyzed using 3 independent platforms [discovery set (left), quantitation set (middle) and the validation set (right)]. Of note, one adjacent non-cancerous tissue, N18 (indicated with a green arrow), which was later found to contain cancerous components, grouped with tumor samples in every PCA analysis.



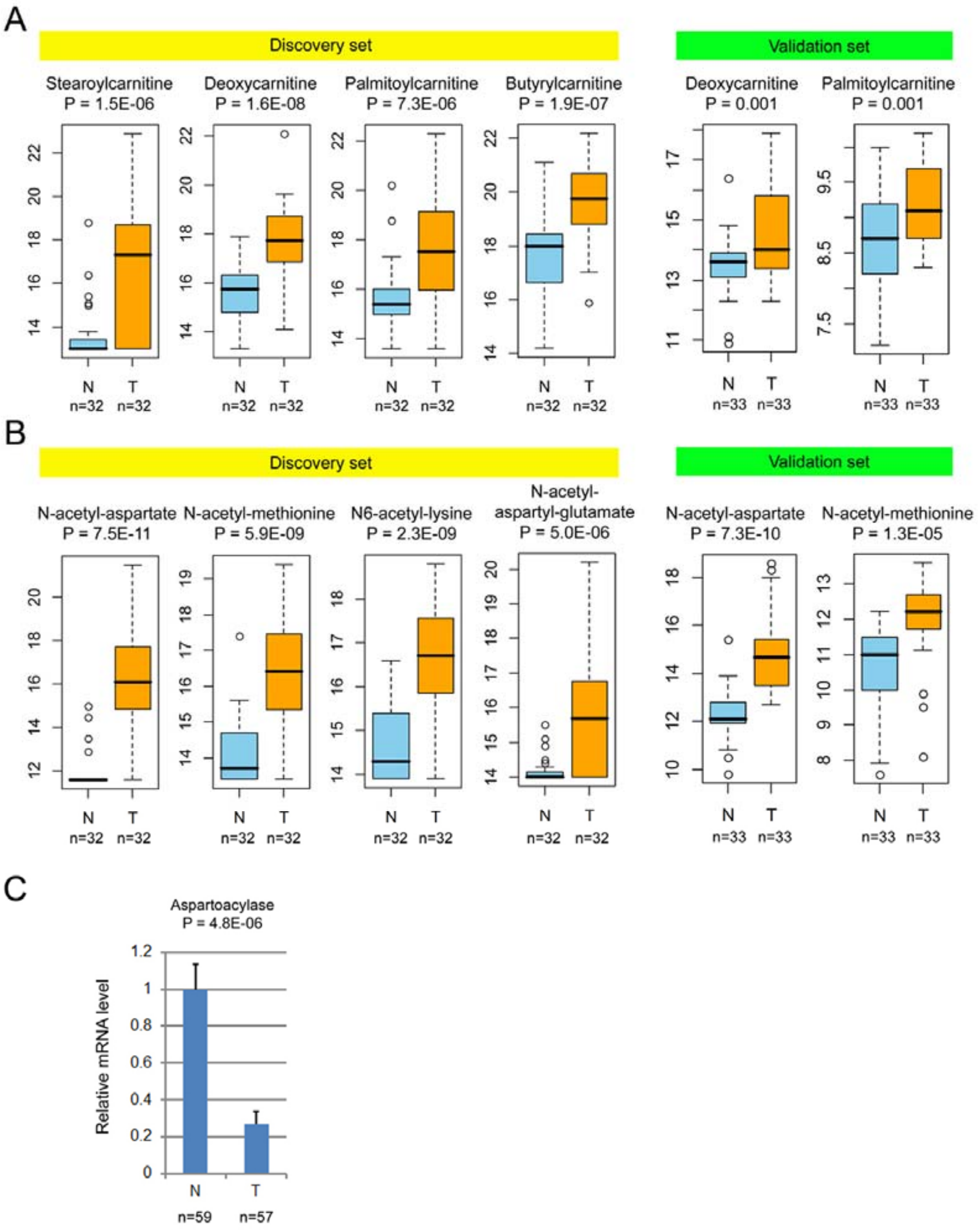
**Supplemental Figure 2. Metabolite patterns in breast cancer.** Heat maps representing variations for 296 metabolites in 67 tumors and 65 adjacent non-cancerous tissues. Columns represent individual tissue samples and rows refer to distinct metabolites. Shades of red and blue represent increased or decreased abundance of metabolites (compared to average). **A**: ER-positive tumors vs. paired non-cancerous tissue (n = 33). **B**: ER-negative tumors vs. paired non-cancerous tissue (n = 32). **C**: ER-negative (n = 34) vs. ER-positive tumors (n = 33). **D**: Age at diagnosis ≥ 50 (n = 34) vs. < 50 (n = 31). **E**: Patient’s body mass index (BMI) ≥ 30 (n = 29) vs. < 30 (n = 33). **F**: Post-menopausal (n = 33) vs. pre-menopausal patients (n = 26). **G**: Patient’s household income ≥ 60,000 USD (n = 11) vs. < 60,000 USD (n = 38). **H**: Patient’s education with school years > 12 (n = 22) vs. ≤ 12 (n = 38). **I**: Tumor grade 3 (n = 29) vs. 1 and 2 (n = 28). **J**: Neoadjuvant therapy (n = 9) vs. without (n = 53). **K**: TNM stage III (n = 15) vs. II and I (n = 52). **L**: Node metastasis positive (n = 29) vs. negative (n = 37). **M**: Numbers of differential metabolites in contrasts A-K (fold change > 2.5 or < 0.4; FDR < 5%). **N**: Differential metabolites overlapping in contrasts A,B,G. Only metabolites that were measurable in at least 40% of the tissues were included in the analysis.



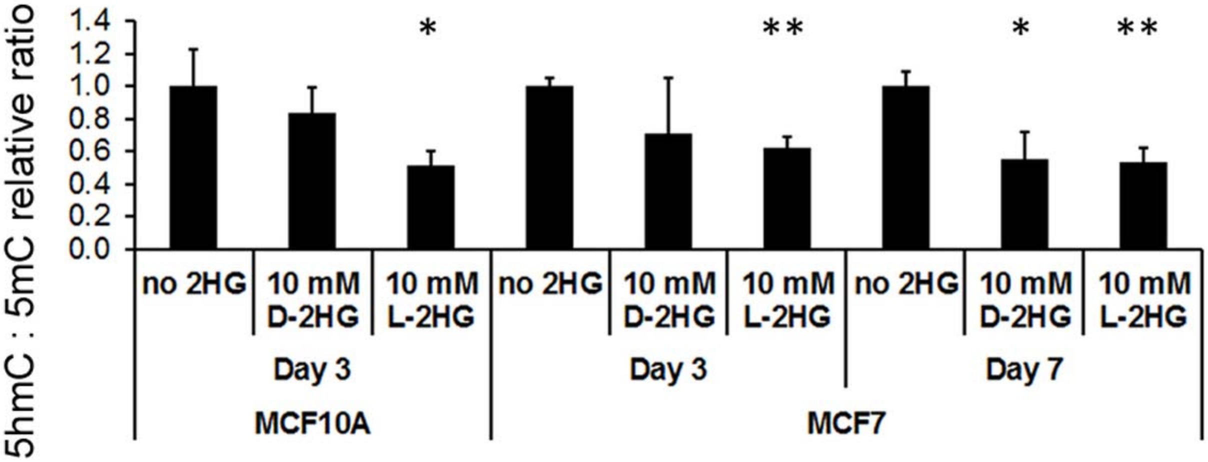
**Supplemental Figure 3. Metabolite-based contrasts for tumor grade and estrogen receptor status. A**: Z-score plots representing the deviation of the top 50 metabolites (as referred to in Figure 1) in grade 1 (blue; n = 8), grade 2 (green; n = 20) and grade 3 (red; n = 29) tumors from the average in grade 1 tumors. **B**: Z-score plots representing the deviation of the 50 metabolites in ER-negative (blue; n= 32) and ER-positive tumors (violet; n = 33) from the average in ER-positive tumors.



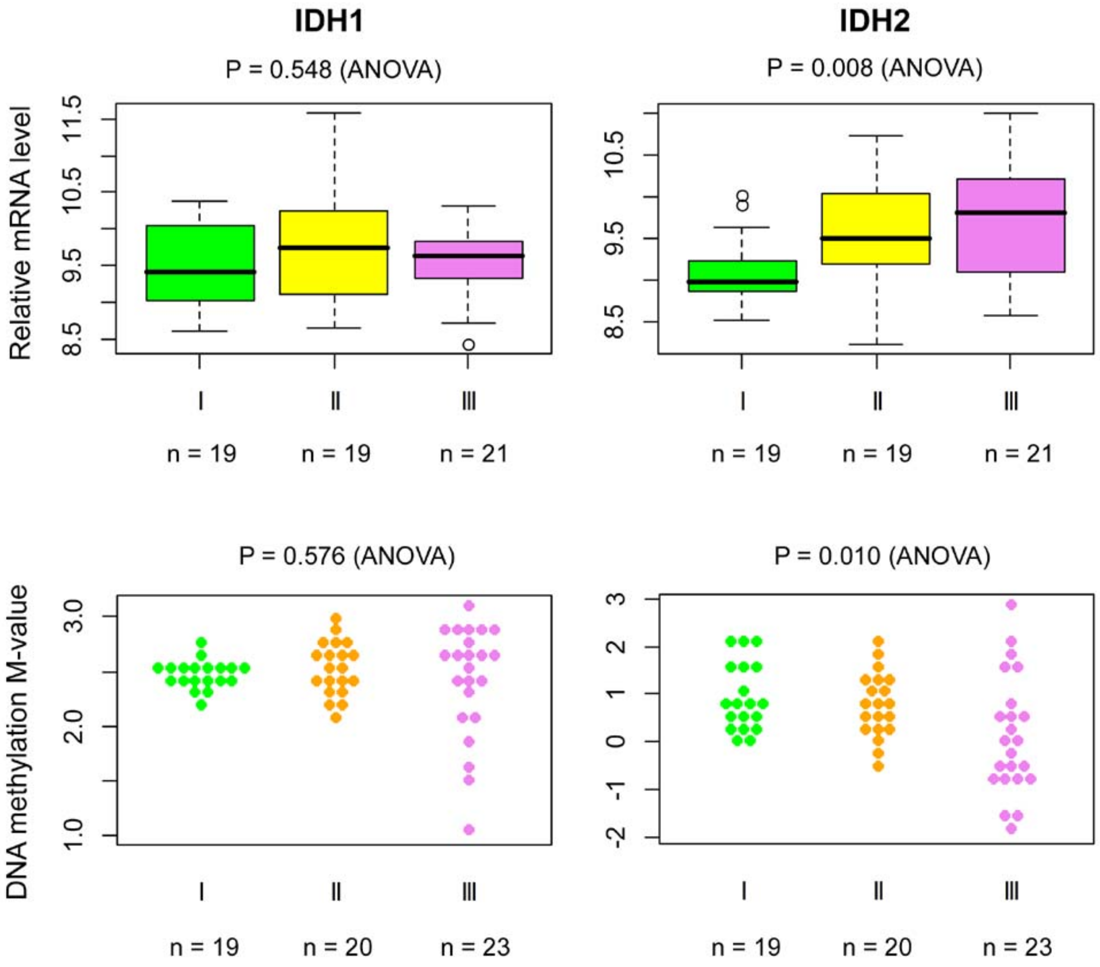
**Supplemental Figure 4**. **Metabolite-based classification of breast cancer subtypes.** Principal component analysis was performed with the 296 named metabolites using Partek Genomics Suite. Note that luminal A tumors formed a cluster (green ellipsoid) while most triple-negative and HER2-positive tumors deviated from it. Green: luminal A tumors (n = 23); red: triple-negative tumors (n = 17); orange: ER-negative/HER2-positive tumors (n = 10); blue: ER-positive/HER2-positive tumors (n = 7).



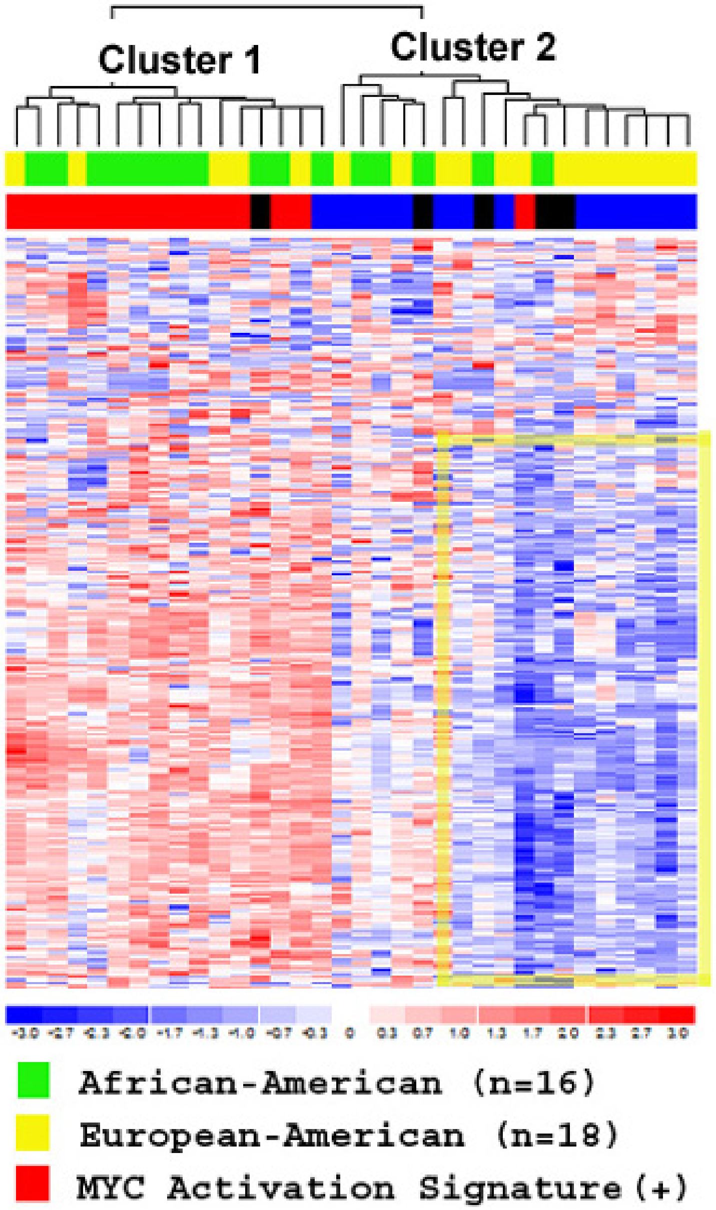
**Supplemental Figure 5. Two metabolite categories with increased abundance in ER-negative tumors.** Relative abundance of carnitines (**A**) and N-acetylamino acids (**B**) in ER-negative breast tumors (T) and adjacent non-cancerous tissue (N) for the discovery (left) and validation sets (right). **C:** Down-regulation of aspartoacylase in breast tumors (T). (**A**) and (**B**): Box plots depict the median value and interquartile range (IQR) shown by the box, the maximum and minimum values within the range from the first quartile minus IQR x 1.5 to the third quartile plus IQR x 1.5 by whiskers, and extreme values outside the range by open circles. Statistical test: Welch test.



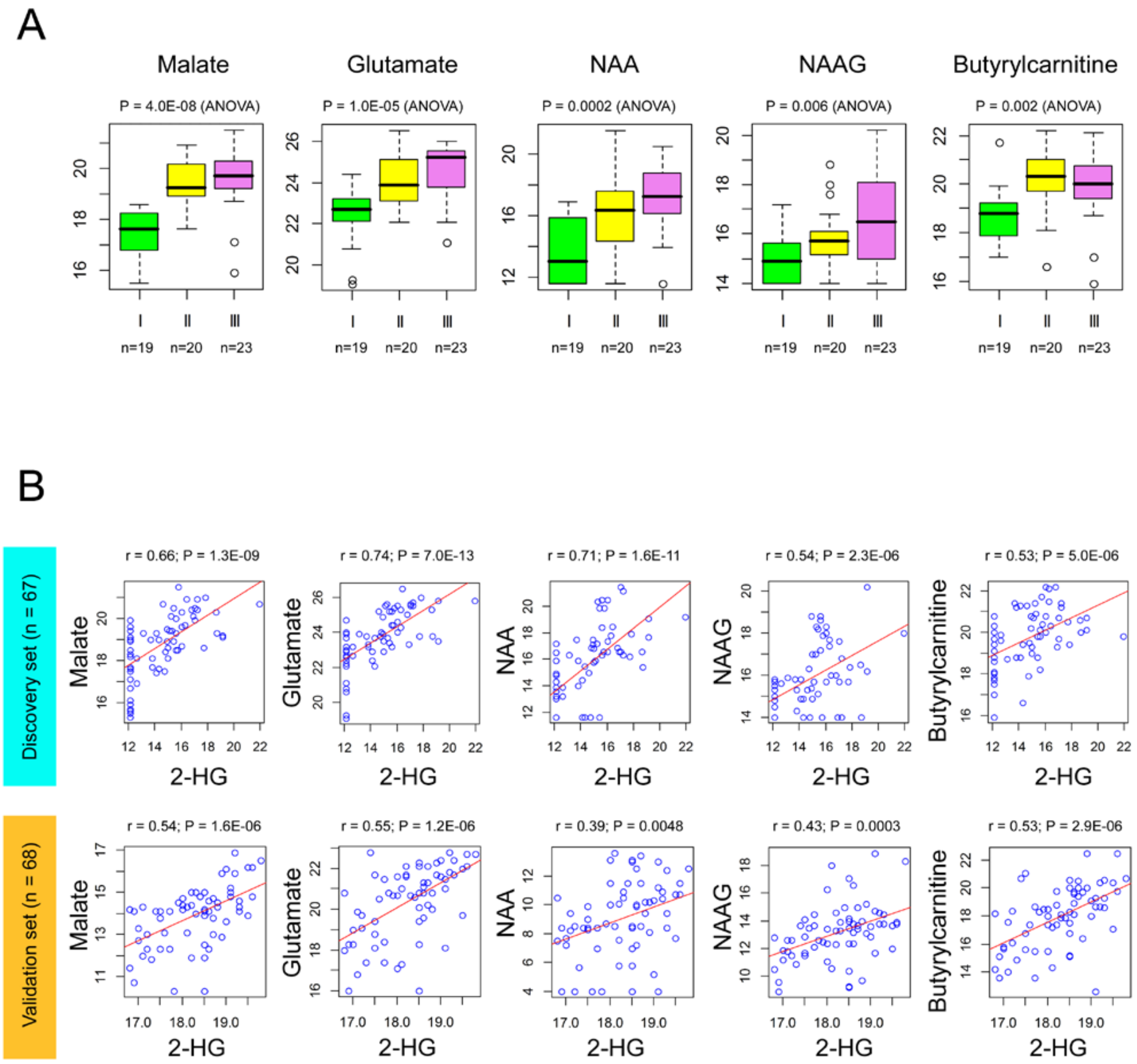
**Supplemental Figure 6. Decreased abundance of 5-hydroxymethylcytosine (5hmC) after treatment with 2HG.** Shown is the ratio between 5hmC and 5-methylcytosine (5mC) in genomic DNA after treating MCF7 cells with 10 mM of D-2HG or L-2HG for either 3 days or 7 days (* $P < 0.05$; ** $P < 0.01$; t-test). A similar shift in this ratio occurred in the nontumorigenic MCF10A cells after a 3 day exposure to 2HG. Graph shows mean ± SD.



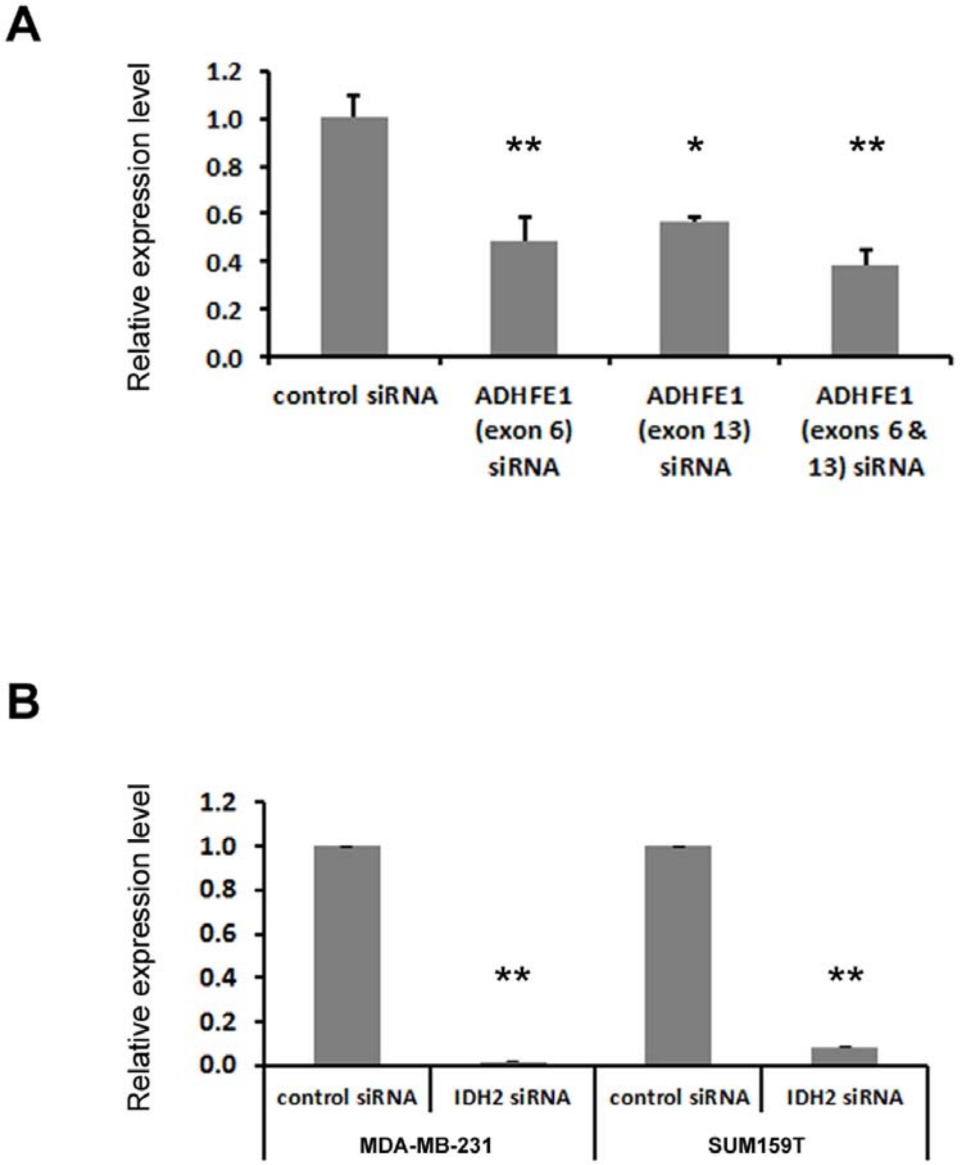
**Supplemental Figure 7. Isocitrate dehydrogenase expression in breast cancer subgroups I – III.** Mitochondrial isocitrate dehydrogenase-2 (IDH2) is up-regulated in subgroup III tumors. Gene expression levels (upper panels, n = 59, GeneChip Human Gene 1.0 ST Array) and DNA methylation M-values (lower panels, n = 62, Human Methylation 450 BeadChips) for cytoplasmic isocitrate dehydrogenase (IDH1) and IDH2 in breast cancer subgroups I, II and III.



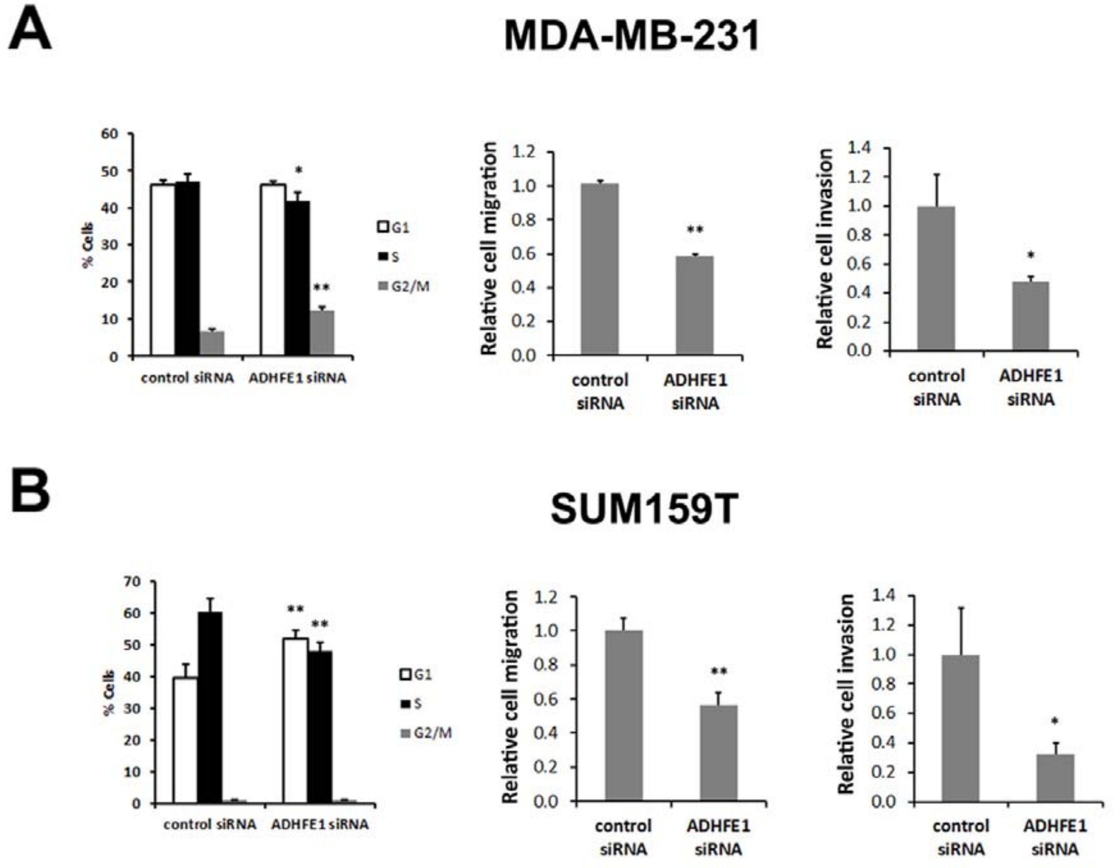
**Supplemental Figure 8. Metabolite-based classification of ER-negative breast tumors is a MYC signature-driven classification.** Heatmap for metabolite abundance with color-coded bars above it that indicate sample classes (green: African-American and yellow: European-American; red: MYC signature-positive tumors and blue: MYC signature-negative tumors; black: gene expression data not available). Unsupervised hierarchical clustering based on abundance of 296 named tissue metabolites separates ER-negative breast tumors (n = 34) into two main clusters. The heatmap shows that most tumors in cluster 1 express a MYC activation signature (second bar, red) defined by a Core MYC Gene Expression signature [Chandriani et al. (24)]. Yellow frame highlights metabolites decreased in a subset of breast tumors, representing mostly European-American patients. Cluster 1 is defined by MYC signature-positive tumors.



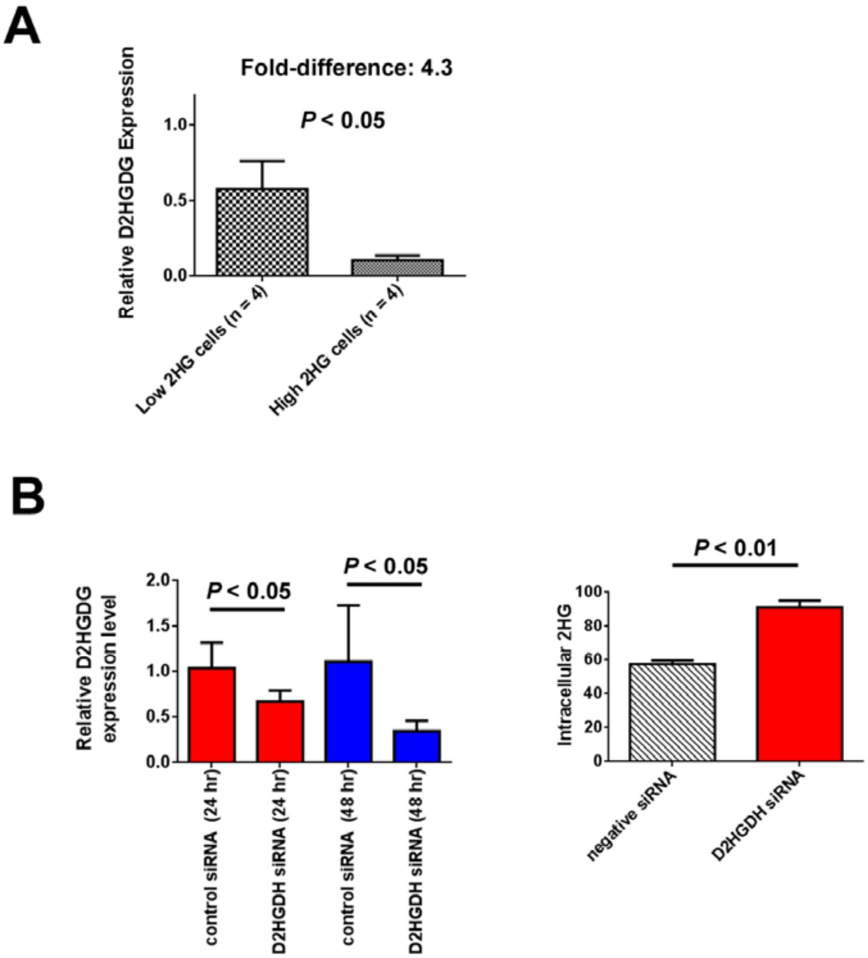
**Supplemental Figure 9. Correlation of five metabolites with 2-hydroxyglutarate (2-HG) levels in breast tumors. A**: Tissue abundance of malate, glutamate, N-acetyl-L-aspartate (NAA), N-acetyl-aspartyl-glutamate (NAAG), and butyrylcarnitine in breast cancer subgroups I - III of the discovery set (n = 62). **B**: Correlation between the above-mentioned metabolites and 2-HG in discovery and validation sets (n = 67 and n = 68, respectively).



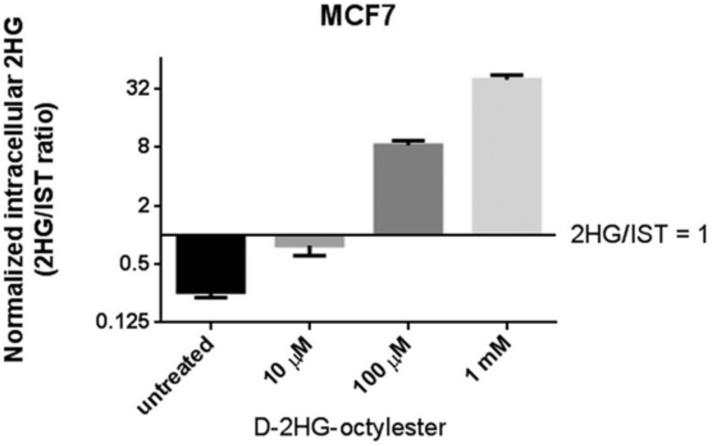
**Supplemental Figure 10. Reduced ADHFE1 and IDH2 mRNA in siRNA-transfected cells.** Human breast cancer cells were transfected with negative control siRNA (Silencer® Select Negative Control), IDH2 or ADHFE1 Select siRNAs and the reduced expression of the genes was confirmed after 24 hrs by qRT-PCR analysis. **A**: Expression of ADHFE1 mRNA in MDA-MB-231 cells after transfection with either negative control siRNA or siRNAs targeting the gene. Shown is the reduced mRNA expression after transfection with siRNAs targeting either ADHFE1 exon 6, exon 13, or both (cotransfection of both siRNAs). * $P < 0.05$; ** $P < 0.01$ (t-test). **B**: Expression of IDH2 mRNA in MDA-MB-231 and SUM159T cells after transfection with either negative control siRNA or siRNA targeting the gene. ** $P < 0.01$ (t-test). Graphs show mean ± SD (n = 4).

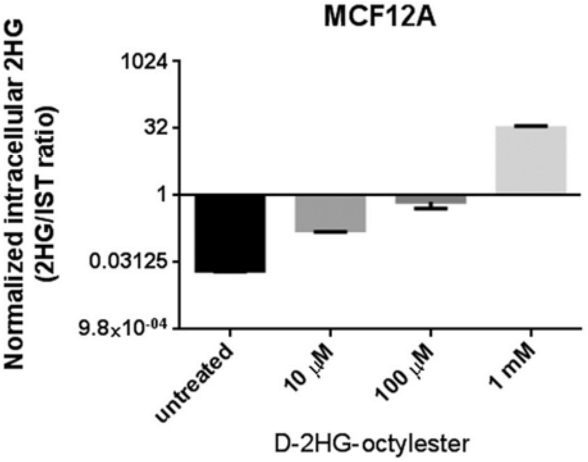


**Supplemental Figure 11. Effect of ADHFE1 knockdown on cell cycle, migration and invasion.** MDA-MB-231 (**A**) and SUM159T cells (**B**) were transfected with ADHFE1 siRNA. Cell cycle distribution, migration and invasion were all assessed at 48 hrs after transfection. Sample numbers were n = 4 for cell cycle analysis, n = 3 for migration & invasion of MDA-MB-231 cells, and n = 4 for migration & invasion of SUM159T cells. $^{**}P < 0.01$; $^{*}P < 0.05$ (t-test). All graphs show mean ± SD.

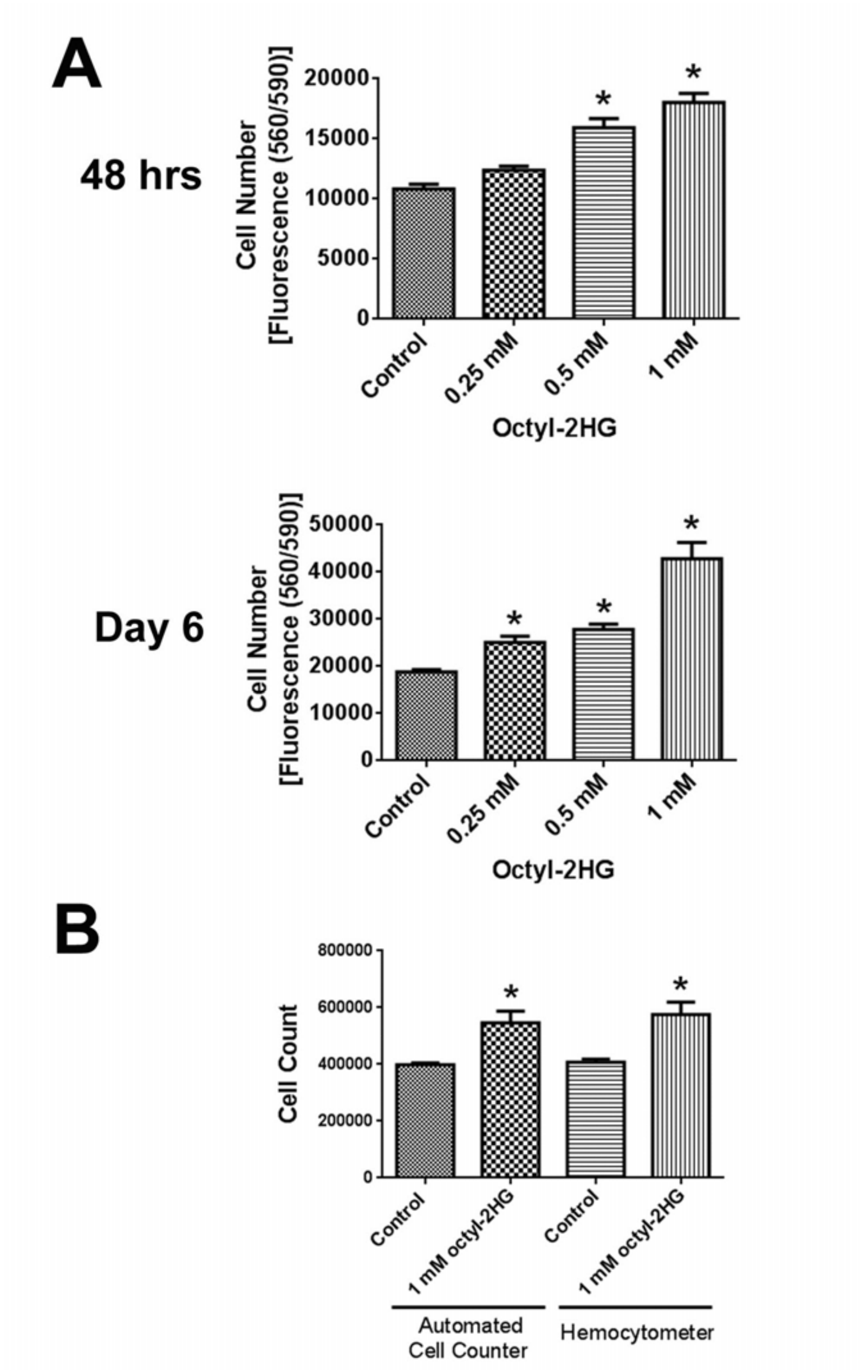


**Supplemental Figure 12. Relationship between D-2-hydoxyglutarate dehydrogenase (D2HGDG) expression and intracellular 2-hydroxyglutarate (2HG) in breast cancer cells. A**: Cell lines (MCF-7, MCF10A, MCF12A, T47D) with basal 2HG levels (low 2HG cells) have an average 4.3-fold higher mRNA expression of D2HGDG than cell lines with aberrant accumulation of 2HG (MDA-MB-231, MDA-MB-468, SUM159T, Hs578T). Graph shows qRT-PCR results with mean ± SEM and *P* value for t-test. **B**: Knockdown of D2HGDG mRNA expression with siRNA in MCF7 cells induces a moderate 1.6-fold increase of 2HG in these cells. 2HG levels were measured 48 hrs after transfection with siRNA. Graphs show mean ± SD (n = 4) and *P* values for t-test.





**Supplemental Figure 13. Accumulation of intracellular 2-hydroxyglutarate (2HG) in two human mammary epithelial cell lines after adding octyl-2HG to the culture medium.** MCF7 and MCF12A cells (with low endogenous 2HG) were cultured for 48 hrs in presence of various concentrations of cell-permeable octyl-2HG and the intracellular accumulation of 2HG was measured in cell pellets using mass spectrometry. 2HG accumulated > 100 fold in these cells when 1 mM octyl-2HG was added to the culture medium. N = 3 for all experiments. Intracellular 2HG was significantly different between untreated and treated cells for all octyl-2HG concentrations ($P < 0.05$). Cells were plated in 100 mm dishes and treated with octyl-2HG [(2R)-2-hydroxyglutaric acid octyl ester (SLR Biosciences, Burlington, MA)] for 48 hrs. To prepare cell pellets, cells were washed 3x with PBS, trypsinized and counted using the TC10 Automated Cell Counter (Bio-Rad, Hercules, CA). 2HG is normalized to cell counts and internal standard (IST).



**Supplemental Figure 14. Octyl-2HG induces increased proliferation of MCF7 cells. A**: MCF-7 cells ($5\times10^3$ cells/well) were seeded in 96-well plates in RPMI medium with 5% FBS and treated with different concentrations of octyl-2HG (0 mM, 0.25 mM, 0.5 mM and 1 mM). Proliferation of cells was measured after 2 and 6 days with the Cell Titer-Blue® Cell Viability Assay. Fluorescence of each well was recorded at 560/590($_{ex/em}$) using a FLUOstar Omega Fluorometer (BMG Labtech, Ortenberg, Germany). **B**: Cell numbers were also determined by directly counting cells after 6 days with and without 1 mM octyl-2HG using two different counting methods. Graphs show mean ± SD (n = 4). * $P < 0.05$ (compared to control; t-test).

**Supplemental Table 1 – Patient characteristics in the discovery cohort**

| | | All N = 67 | AA[1] N = 32 | EA[2] N = 35 | *P* value[3] |
|---|---|---|---|---|---|
| | | | **Mean ± SD** | | **t test** |
| Age (years) (n=65) | | 54±15.8 | 52.9±16.1 | 54.9±15.7 | 0.62 |
| Tumor size (cm across) (n = 60) | | 4.2±2.7 | 3.8±2.3 | 4.5±3.0 | 0.25 |
| West-African ancestry (%) (range among AA: 67% to 95%) | | | 83.2±8.7 | 2.2±0.5 | <0.001 |
| European ancestry (%) | | | 16.0±8.8 | 97.6±4.7 | <0.001 |
| | | **N** | **N** | **N** | **Fisher's exact test** |
| ER Status | Negative | 34 | 16 | 18 | |
| | Triple-negative[4] (basal-like[5]) | 17(16) | 8(8) | 9(8) | |
| | HER2-positive[6] (unknown) | 10(3) | 7(1) | 3(2) | |
| | Positive | 33 | 16 | 17 | 1.0[7] |
| | HER-negative | 23 | 10 | 13 | |
| | HER2-positive (unknown) | 7(3) | 3(3) | 4 | |
| TNM Stage | I | 6 | 2 | 4 | |
| | II | 46 | 24 | 22 | |
| | III | 15 | 6 | 9 | 0.63 |
| Grade | 1 | 8 | 1 | 7 | |
| | 2 | 20 | 8 | 12 | |
| | 3 | 29 | 18 | 11 | 0.029[8] |
| | Unknown | 10 | 5 | 5 | |
| p53 mutation | Negative | 51 | 25 | 26 | |
| | Positive | 16 | 7 | 9 | 0.78 |
| Menopause | No | 26 | 14 | 12 | |
| | Yes | 33 | 14 | 19 | 0.44[8,9] |
| | Unknown | 8 | 4 | 4 | |
| Income | Less than $15,000 | 12 | 10 | 2 | |
| | $15,000 to $60,000 | 26 | 13 | 13 | |
| | More than $60,000 | 11 | 1 | 10 | <0.01[8] |
| | Unknown | 18 | 8 | 10 | |
| Body mass index | ≤24.9 | 18 | 7 | 11 | |
| | 25.0 to 29.9 | 15 | 5 | 10 | |
| | ≥ 30.0 | 29 | 17 | 12 | 0.25[8] |
| | Unknown | 5 | 3 | 2 | |
| Neoadjuvant therapy | No | 53 | 30 | 23 | |
| | Yes | 9 | 1 | 8 | 0.026[8] |
| | Unknown | 5 | 1 | 4 | |

SD = standard deviation. [1] AA=African-American, [2]EA=European-American. Race/ethnicity by self-identification; [3]comparing AA versus EA; [4]Negative for estrogen, progesterone, and HER2 receptor expression; [5]Basal-like expression signature (PAM50-defined) and/or immunohistochemistry (IHC)-based (ER-negative, HER2-negative, cytokeratin 5/6-positive or EGFR-positive); [6]DAKO HercepTest™ IHC-positive (score 3) or IHC score 2 and HER2-enriched by gene expression; [7]ER-negative versus ER-positive; [8]Unknown not included; [9]Premenopausal was defined as still having menstrual periods. Postmenopausal is defined as being 55 years or older, stated as not having menstrual periods, or hysterectomy with both ovaries removed.

**Supplemental Table 3 – Patient characteristics in the validation cohort**

| | | All<br>N = 70 | AA[1]<br>N = 49 | EA[2]<br>N = 21 | *P* value[3] |
|---|---|---|---|---|---|
| | | | **Mean ± SD** | | **t test** |
| Age (years) (n = 70) | | 50.6±11 | 51±10.7 | 49.6±12 | 0.6 |
| Tumor size (cm across) (n = 63) | | 4.4±2.9 | 4.2±2.7 | 4.9±3.5 | 0.45 |
| West-African ancestry (%) among AA (range: 52% to 94%; n = 39) | | | 81.8±8.2 | | |
| | | **N** | **N** | **N** | **Fisher's exact test** |
| ER-negative | | 70 | 49 | 21 | |
| | Triple-negative | 39 | 29 | 10 | |
| TNM Stage | I | 10 | 7 | 3 | |
| | II | 45 | 30 | 15 | |
| | III/IV | 15 | 12 | 3 | 0.69 |
| Grade | 1 | 1 | 1 | 0 | |
| | 2 | 10 | 6 | 4 | |
| | 3 | 49 | 39 | 10 | 0.41[4] |
| | Unknown | 10 | 3 | 7 | |
| p53 mutation | Negative | 53 | 35 | 18 | |
| | Positive | 17 | 14 | 3 | 0.24 |
| Menopause | No | 20 | 14 | 6 | |
| | Yes | 33 | 24 | 9 | 1.0[4] |
| | Unknown | 17 | 11 | 6 | |
| Income | Less than $15,000 | 10 | 10 | 0 | |
| | $15,000 to $60,000 | 32 | 22 | 10 | |
| | More than $60,000 | 6 | 1 | 5 | <0.01[4] |
| | Unknown | 22 | 16 | 6 | |
| Body mass index | ≤24.9 | 20 | 11 | 9 | |
| | 25.0 to 29.9 | 20 | 14 | 6 | |
| | ≥ 30.0 | 27 | 21 | 6 | 0.25[4] |
| | Unknown | 3 | 3 | 0 | |
| Neoadjuvant therapy | No | 37 | 29 | 8 | |
| | Yes | 29 | 18 | 11 | 0.18[4] |
| | Unknown | 14 | 12 | 2 | |

SD = standard deviation. [1] AA=African-American, [2] EA=European-American. Race/ethnicity is determined by self-identification; [3]comparing AA with EA; [4]Unknown not included.

**Supplemental Table 4:** Top-ranked metabolites with abundance differences between African-American (n = 8) and European-American (n = 9) triple-negative tumors

| More abundant in AA tumor | | |
|---|---|---|
| **Metabolite ID** | **Fold Change** | **FDR (%)** |
| 2-arachidonoylglycerophosphocholine | 11.27 | 0.00 |
| 2-palmitoylglycerophosphocholine | 10.34 | 0.00 |
| stearoylcarnitine | 9.19 | 0.00 |
| sphingosine | 8.99 | 0.00 |
| stachydrine | 8.15 | 0.00 |
| 2-oleoylglycerophosphocholine | 8.04 | 0.00 |
| 2-oleoylglycerophosphoethanolamine | 6.96 | 0.00 |
| 2-docosahexaenoylglycerophosphoethanolamine | 6.71 | 0.00 |
| sphinganine | 6.68 | 0.00 |
| 2-linoleoylglycerophosphocholine | 6.26 | 0.00 |
| 2-docosapentaenoylglycerophosphoethanolamine | 6.18 | 0.00 |
| docosatrienoate (223n3) | 5.75 | 0.00 |
| 1-palmitoylglycerophosphocholine | 5.69 | 0.00 |
| uracil | 5.64 | 0.00 |
| glucose-6-phosphate (G6P) | 5.47 | 0.00 |
| putrescine | 5.43 | 0.00 |
| palmitoylcarnitine | 5.31 | 0.00 |
| 1-stearoylglycerophosphocholine | 5.29 | 0.00 |
| 1-palmitoylglycerol (1-monopalmitin) | 5.23 | 1.18 |
| cytidine 5'-diphosphocholine | 5.13 | 1.63 |
| 2-linoleoylglycerophosphoethanolamine | 4.97 | 0.00 |
| docosadienoate (222n6) | 4.92 | 0.00 |
| fructose-6-phosphate | 4.74 | 0.00 |
| erythronate | 4.73 | 0.00 |
| X - 13421 | 4.69 | 1.18 |
| cytidine 5'-monophosphate (5'-CMP) | 4.48 | 0.00 |
| X - 13391 | 4.38 | 1.18 |
| phosphoethanolamine | 4.35 | 1.18 |
| X - 12660 | 4.25 | 0.00 |
| 2-arachidonoylglycerophosphoethanolamine | 4.18 | 0.00 |
| isovalerylcarnitine | 4.11 | 1.18 |
| mannose-6-phosphate | 4.08 | 1.18 |
| X - 8994 | 3.79 | 2.15 |
| 1-oleoylglycerophosphocholine | 3.76 | 1.18 |
| X - 13134 | 3.72 | 0.00 |
| X - 5229 | 3.70 | 0.00 |
| 2-hydroxyglutarate | 3.69 | 3.12 |
| N-acetylaspartate (NAA) | 3.68 | 1.18 |
| X - 12627 | 3.68 | 0.00 |
| inositol 1-phosphate (I1P) | 3.54 | 1.63 |
| X - 15161 | 3.29 | 2.15 |
| dihomo-linoleate (202n6) | 3.23 | 1.18 |
| cysteine-glutathione disulfide | 3.17 | 4.90 |
| X - 13418 | 3.15 | 2.66 |
| glutathione, oxidized (GSSG) | 3.14 | 4.90 |
| X - 13396 | 3.11 | 1.63 |
| X - 3094 | 3.10 | 1.18 |
| beta-alanine | 3.07 | 1.18 |
| cysteine | 3.06 | 1.63 |
| oleoylcarnitine | 3.02 | 2.66 |
| deoxycarnitine | 3.00 | 1.18 |
| dihomo-linolenate (203n3 or n6) | 2.97 | 1.18 |
| glutamate | 2.95 | 1.18 |
| X - 12051 | 2.95 | 1.18 |
| 1-stearoylglycerophosphoinositol | 2.83 | 2.15 |
| malate | 2.82 | 1.18 |
| 1-stearoylglycerol (1-monostearin) | 2.80 | 1.63 |
| inosine | 2.78 | 1.63 |
| X - 12801 | 2.77 | 3.12 |
| succinylcarnitine | 2.76 | 3.12 |
| hydroxyisovaleroyl carnitine | 2.75 | 2.15 |
| 3-(4-hydroxyphenyl)lactate | 2.75 | 1.18 |
| eicosenoate (201n9 or 11) | 2.71 | 1.63 |
| fumarate | 2.63 | 1.63 |
| N-acetyl-aspartyl-glutamate (NAAG) | 2.61 | 3.12 |
| N-acetylserine | 2.59 | 1.63 |
| arachidonate (204n6) | 2.58 | 1.63 |
| X - 13413 | 2.57 | 3.12 |
| 3-phosphoglycerate | 2.55 | 1.63 |
| docosahexaenoate (DHA; 226n3) | 2.54 | 1.63 |
| cystine | 2.53 | 3.12 |
| X - 15117 | 2.52 | 2.66 |
| N-acetylglucosamine | 2.51 | 3.12 |

| Less abundant in AA tumor | | |
|---|---|---|
| **Metabolite ID** | **Fold Change** | **FDR (%)** |
| X - 12786 | 0.17 | 6.26 |
| histamine | 0.18 | 6.26 |
| X - 13066 | 0.39 | 6.26 |

**Supplemental Table 5:** Top-ranked metabolites with abundance differences between ER-negative (n = 34) and ER-positive (n = 33) tumors

| More abundant in ER(-) tumor; FDR < 5% | | |
|---|---|---|
| **Metabolite ID** | **Fold Change** | **FDR (%)** |
| docosatrienoate | 5.97 | 0.00 |
| 1-palmitoylglycerol | 3.46 | 0.00 |
| 1-palmitoylglycerophosphocholine | 3.32 | 0.00 |
| dihomo-linolenate | 3.26 | 0.00 |
| 1-stearoylglycerophosphoinositol | 3.17 | 0.00 |
| docosadienoate | 3.12 | 0.00 |
| 1-stearoylglycerophosphocholine | 3.03 | 0.00 |
| 1-oleoylglycerophosphocholine | 2.96 | 0.00 |
| sphingosine | 2.83 | 0.00 |
| 2-hydroxyglutarate | 2.79 | 0.00 |
| uracil | 2.76 | 0.00 |
| 2-arachidonoylglycerophosphocholine | 2.76 | 0.00 |
| cystathionine | 2.76 | 0.00 |
| 5,6-dihydrouracil | 2.63 | 0.00 |
| docosapentaenoate | 2.62 | 0.00 |
| 2-palmitoylglycerophosphocholine | 2.60 | 0.00 |
| adrenate | 2.60 | 0.00 |
| 2-oleoylglycerophosphocholine | 2.54 | 0.00 |
| cytidine 5'-diphosphocholine | 2.52 | 0.00 |
| arachidonate | 2.50 | 0.00 |
| 1-stearoylglycerol | 2.48 | 0.00 |
| docosapentaenoate | 2.44 | 0.00 |
| kynurenine | 2.44 | 0.00 |
| 2-docosapentaenoylglycerophosphoethanolamine | 2.39 | 0.00 |
| N-acetylglucosamine | 2.33 | 0.00 |
| dihomo-linoleate | 2.30 | 0.00 |
| aspartylleucine | 2.26 | 0.00 |
| fructose-6-phosphate | 2.18 | 0.72 |
| 1-palmitoylplasmenylethanolamine | 2.17 | 0.00 |
| stearidonate | 2.15 | 0.00 |
| 1-arachidonoylglycerophosphoinositol | 2.15 | 0.00 |
| 2-linoleoylglycerophosphocholine | 2.14 | 0.00 |
| sphinganine | 2.13 | 0.72 |
| eicosapentaenoate | 2.12 | 0.00 |
| inositol 1-phosphate | 2.09 | 0.72 |
| beta-alanine | 2.08 | 0.00 |
| docosahexaenoate | 2.06 | 0.00 |
| N-acetylaspartate (NAA) | 1.97 | 2.00 |
| 2-arachidonoylglycerophosphoethanolamine | 1.95 | 0.00 |
| 1-stearoylglycerophosphoethanolamine | 1.95 | 0.00 |
| stearoylcarnitine | 1.94 | 1.55 |
| 2-methylbutyroylcarnitine | 1.94 | 0.00 |
| ribose | 1.93 | 0.72 |
| 2-docosahexaenoylglycerophosphoethanolamine | 1.90 | 1.28 |
| isovalerylcarnitine | 1.89 | 0.72 |
| eicosenoate | 1.84 | 0.72 |
| palmitoleate | 1.83 | 0.00 |
| 2-linoleoylglycerophosphoethanolamine | 1.81 | 1.28 |
| cystine | 1.80 | 1.55 |
| 10-nonadecenoate | 1.80 | 0.00 |
| glycerol | 1.79 | 0.72 |
| 1-arachidonoylglycerophosphoethanolamine | 1.78 | 0.00 |
| guanine | 1.78 | 0.72 |
| N-acetylneuraminate | 1.78 | 0.00 |
| glucose-6-phosphate | 1.78 | 2.40 |
| 2-oleoylglycerophosphoethanolamine | 1.76 | 2.40 |
| linolenate [alpha or gamma] | 1.74 | 0.72 |
| glutamate | 1.70 | 1.28 |
| mannose-6-phosphate | 1.70 | 3.04 |
| glycerol 2-phosphate | 1.69 | 1.55 |
| tryptophan betaine | 1.66 | 1.55 |
| succinylcarnitine | 1.66 | 2.40 |
| fumarate | 1.65 | 1.28 |
| oleate | 1.65 | 1.55 |
| hydroxyisovaleroyl carnitine | 1.63 | 1.28 |
| 10-heptadecenoate | 1.62 | 0.72 |
| glycylleucine | 1.62 | 1.28 |
| margarate | 1.60 | 0.72 |
| malate | 1.59 | 1.55 |
| palmitoylcarnitine | 1.59 | 3.45 |
| myristoleate | 1.58 | 0.72 |
| phosphoethanolamine | 1.58 | 4.37 |
| gamma-glutamylleucine | 1.58 | 0.72 |
| N1-methyladenosine | 1.57 | 1.28 |
| 1,3-dihydroxyacetone | 1.57 | 1.55 |
| cis-vaccenate | 1.56 | 2.40 |
| phosphate | 1.56 | 1.28 |
| palmitoyl sphingomyelin | 1.56 | 1.28 |
| lactate | 1.55 | 1.28 |
| 2-hydroxypalmitate | 1.53 | 1.28 |
| mannose | 1.53 | 1.28 |
| myristate | 1.52 | 0.72 |
| 13-HODE | 1.52 | 2.40 |
| N-acetylglucosamine 6-phosphate | 1.51 | 2.00 |
| deoxycarnitine | 1.51 | 3.04 |

| Less abundant in ER(-) tumor; FDR < 20% | | |
|---|---|---|
| **Metabolite ID** | **Fold Change** | **FDR (%)** |
| histamine | 0.40 | 7.93 |
| glutathione, oxidized (GSSG) | 0.55 | 14.43 |
| ophthalmate | 0.57 | 10.90 |
| 3-phosphoglycerate | 0.58 | 10.90 |
| glutathione, reduced (GSH) | 0.61 | 15.50 |
| pyruvate | 0.62 | 14.43 |

**Supplemental Table 6:** Relative abundance of 2-hydroxyglutarate in human breast cancer cell lines

| | Diagnosis | ER | PR | HER2 | Cell lines | 2HG abundance per cell (femtomole) | 2HG cellular concentration† (mM) |
|---|---|---|---|---|---|---|---|
| 1 | Cancer | NEG | NEG | NEG | MDA-MB-231 | 7.65 | 5.10 |
| 2 | Cancer | NEG | NEG | NEG | MDA-MB-468 | 6.83 | 4.55 |
| 3 | Cancer | POS | POS | POS | MDA-MB-361 | 4.42 | 2.94 |
| 4 | Cancer | NEG | NEG | POS | HCC-1419 -Rep 1 | 4.10 | 2.73 |
| | Cancer | NEG | NEG | POS | HCC-1419 -Rep-2 | 3.00 | 2.00 |
| 5 | Cancer | NEG | NEG | NEG | SUM159T | 3.41 | 2.27 |
| 6 | Cancer | POS | POS | POS | BT-474 | 0.16 | 0.11 |
| 7 | Cancer | NEG | NEG | NEG | HCC-70 | 0.12 | 0.08 |
| 8 | Cancer | POS | POS | NEG | T47D -Rep-1 | 0.08 | 0.06 |
| | Cancer | POS | POS | NEG | T47D -Rep-2 | 0.05 | 0.03 |
| 9 | Cancer | NEG | NEG | POS | HCC-1569 | 0.05 | 0.03 |
| 10 | Cancer | NEG | NEG | NEG | HCC-1187 | 0.04 | 0.03 |
| 11 | Cancer | NEG | NEG | POS | HCC-1954 | 0.04 | 0.02 |
| 12 | Non-tumorigenic | NEG | NEG | NEG | HBL-100 | 0.03 | 0.02 |
| 13 | Non-tumorigenic | NEG | NEG | NEG | MCF-12A | 0.02 | 0.01 |
| 14 | Non-tumorigenic | NEG | NEG | NEG | MCF-10A | 0.02 | 0.01 |
| 15 | Cancer | POS | POS | NEG | MCF-7 -Rep-1 | 0.02 | 0.01 |
| | Cancer | POS | POS | NEG | MCF-7 -Rep-2 | 0.02 | 0.01 |
| 16 | Cancer | NEG | NEG | NEG | HCC-1143 | 0.01 | 0.01 |
| 17 | Cancer | NEG | NEG | POS | UACC-812 | 0.01 | 0.01 |

† Concentration was estimated assuming the average cellular volume to be 1.3 picoliter.

**Supplemental Table 7:** Association of tumor subgroups with breast cancer survival in the NCI study using multivariable Cox regression analysis

| | Univariable Cox regression | | | | Multivariable Cox regression[†] | | | |
|---|---|---|---|---|---|---|---|---|
| | **HR** | **95% CI** | ***P*** | ***n*** | **HR** | **95% CI** | ***P*** | ***n*** |
| **5-Year survival** | | | | | | | | |
| DNA methylation subgroup | | | | | | | | |
| Subgroup I | 1 | | | 19 | 1 | | | 19 |
| Subgroup II | 2.86 | 0.76 - 10.81 | 0.121 | 20 | 5.49 | 0.59 - 51.30 | 0.136 | 20 |
| Subgroup III | 4.52 | 1.28 - 15.88 | 0.019 [§] | 23 | 12.80 | 1.34 - 122.31 | 0.027 [§] | 23 |
| **10-Year survival** | | | | | | | | |
| DNA methylation subgroup | | | | | | | | |
| Subgroup I | 1 | | | 19 | 1 | | | 19 |
| Subgroup II | 2.52 | 0.77 - 8.19 | 0.126 | 20 | 2.98 | 0.53 - 16.67 | 0.213 | 20 |
| Subgroup III | 3.48 | 1.13 - 10.70 | 0.030 [§] | 23 | 7.02 | 1.20 - 41.01 | 0.031 [§] | 23 |

[†] Adjusted for age at diagnosis, TNM stage, ER status, race, p53 mutation, and chemotherapy.

[§] $P < 0.05$

**Supplemental Table 8:** DNA methylation signature for subgroup III predicts survival in the Dedeurwaerder et al. dataset

| | **Univariable Cox regression** | | | | **Multivariable Cox regression**[†] | | | |
|---|---|---|---|---|---|---|---|---|
| | **HR** | **95% CI** | ***P*** | ***n*** | **HR** | **95% CI** | ***P*** | ***n*** |
| **5-Year survival** | | | | | | | | |
| Subgroup III signature | | | | | | | | |
| Low | 1 | | | 117 | 1 | | | 117 |
| High | 2.10 | 0.98 - 4.49 | 0.055 | 117 | 2.53 | 1.17 - 5.44 | 0.018 [§] | 117 |
| **10-Year survival** | | | | | | | | |
| Subgroup III signature | | | | | | | | |
| Low | 1 | | | 117 | 1 | | | 117 |
| High | 2.10 | 1.08 - 4.09 | 0.029 [§] | 117 | 2.45 | 1.25 - 4.81 | 0.009 [§] | 117 |

[†] Adjusted for age and ER status.

[§] $P < 0.05$

**Supplemental Table 9:** Genes that were differentially expressed in subgroup III tumors compared to subgroup I tumors

| UP in subgroup III | | |
|---|---|---|
| **Gene Symbol** | **Fold Change** | **q-value (%)** |
| TPX2 | 3.960 | 0.000 |
| CENPF | 3.904 | 0.000 |
| ANLN | 3.878 | 0.000 |
| ASPM | 3.779 | 0.000 |
| NUF2 | 3.743 | 0.000 |
| TTK | 3.583 | 0.000 |
| BUB1 | 3.543 | 0.000 |
| CASP14 | 3.532 | 0.051 |
| TOP2A | 3.505 | 0.000 |
| FOXM1 | 3.383 | 0.000 |
| CDC20 | 3.377 | 0.000 |
| PLK1 | 3.366 | 0.000 |
| C15orf42 | 3.293 | 0.000 |
| CDK1 | 3.291 | 0.000 |
| GINS1 | 3.273 | 0.000 |
| DLGAP5 | 3.266 | 0.000 |
| UBE2T | 3.223 | 0.000 |
| KIF18A | 3.194 | 0.000 |
| CENPI | 3.193 | 0.000 |
| AURKA | 3.163 | 0.000 |
| CENPE | 3.161 | 0.000 |
| KIF20A | 3.112 | 0.000 |
| DTL | 3.099 | 0.000 |
| DEPDC1 | 3.074 | 0.000 |
| CXCL10 | 3.048 | 0.000 |
| KIF14 | 3.014 | 0.000 |
| CCNB2 | 3.002 | 0.000 |
| FAM72D | 2.993 | 0.000 |
| MYBL2 | 2.970 | 0.000 |
| CEP55 | 2.946 | 0.000 |
| CKS2 | 2.937 | 0.000 |
| PRC1 | 2.937 | 0.000 |
| STIL | 2.912 | 0.000 |
| SGOL1 | 2.897 | 0.000 |
| CKAP2L | 2.894 | 0.000 |
| ARHGAP11A | 2.872 | 0.000 |
| HIST2H4A | 2.795 | 0.000 |
| HIST2H4A | 2.795 | 0.000 |
| MELK | 2.773 | 0.000 |
| CDC6 | 2.765 | 0.000 |
| NCAPH | 2.757 | 0.000 |
| MKI67 | 2.749 | 0.000 |
| PBK | 2.741 | 0.000 |
| LRP8 | 2.732 | 0.000 |
| BRIP1 | 2.723 | 0.000 |
| PRR11 | 2.722 | 0.000 |
| KIF23 | 2.679 | 0.000 |
| NEK2 | 2.675 | 0.000 |
| CXCL11 | 2.673 | 0.000 |
| CDKN3 | 2.648 | 0.000 |
| KIF11 | 2.644 | 0.000 |
| SKA3 | 2.639 | 0.000 |
| ATAD2 | 2.614 | 0.000 |
| IQGAP3 | 2.611 | 0.000 |
| HIST2H3D | 2.607 | 0.000 |
| KIF4A | 2.603 | 0.000 |
| EZH2 | 2.590 | 0.000 |
| XRCC2 | 2.557 | 0.000 |
| KPNA2 | 2.552 | 0.000 |
| SLC7A5 | 2.547 | 0.000 |

| DOWN in subgroup III | | |
|---|---|---|
| **Gene Symbol** | **Fold Change** | **q-value (%)** |
| MUCL1 | 0.104 | 0.000 |
| AK5 | 0.126 | 0.000 |
| FABP4 | 0.149 | 0.000 |
| SCGB2A2 | 0.154 | 0.000 |
| ANKRD30A | 0.160 | 0.000 |
| PROL1 | 0.164 | 0.000 |
| MIR145 | 0.176 | 0.000 |
| KRT15 | 0.192 | 0.000 |
| WIF1 | 0.194 | 0.000 |
| PIP | 0.197 | 0.000 |
| ANKRD30B | 0.205 | 0.000 |
| CYP4Z1 | 0.205 | 0.000 |
| OGN | 0.208 | 0.000 |
| COL14A1 | 0.229 | 0.000 |
| FIGF | 0.236 | 0.000 |
| LRP2 | 0.241 | 0.000 |
| TSHZ2 | 0.241 | 0.000 |
| ADH1B | 0.247 | 0.000 |
| TSHZ2 | 0.249 | 0.000 |
| NEK10 | 0.252 | 0.000 |
| ABCA9 | 0.258 | 0.000 |
| PIGR | 0.259 | 0.000 |
| MYH11 | 0.263 | 0.000 |
| ABCA6 | 0.266 | 0.000 |
| ABCA8 | 0.266 | 0.000 |
| CD36 | 0.267 | 0.000 |
| CCL28 | 0.281 | 0.000 |
| GPC3 | 0.283 | 0.000 |
| IGSF10 | 0.285 | 0.000 |
| CYP4X1 | 0.286 | 0.000 |
| CHRDL1 | 0.288 | 0.000 |
| CNTNAP3 | 0.289 | 0.000 |
| CNTNAP3 | 0.289 | 0.000 |
| AREG | 0.292 | 0.000 |
| SVEP1 | 0.292 | 0.000 |
| PTN | 0.293 | 0.000 |
| MFAP4 | 0.293 | 0.000 |
| SCUBE2 | 0.295 | 0.000 |
| CNTNAP3 | 0.295 | 0.000 |
| ACTG2 | 0.296 | 0.000 |
| NTRK2 | 0.296 | 0.000 |
| STEAP4 | 0.298 | 0.000 |
| LTF | 0.298 | 0.000 |
| PGR | 0.299 | 0.000 |
| TNXB | 0.305 | 0.000 |
| TNXB | 0.305 | 0.000 |
| TNXA | 0.309 | 0.000 |
| APOD | 0.311 | 0.000 |
| C7 | 0.314 | 0.000 |
| PTHLH | 0.315 | 0.000 |
| ADIPOQ | 0.315 | 0.051 |
| SFRP1 | 0.317 | 0.000 |
| FAM189A2 | 0.319 | 0.000 |
| CYP4Z2P | 0.323 | 0.171 |
| CNTNAP3 | 0.323 | 0.000 |
| GSTM5 | 0.324 | 0.000 |
| MAOB | 0.327 | 0.000 |
| RBMS3 | 0.327 | 0.000 |
| KRT14 | 0.328 | 0.000 |
| KIT | 0.328 | 0.000 |

| | | |
|---|---|---|
| NCAPG | 2.544 | 0.000 |
| BUB1B | 2.537 | 0.000 |
| CCNA2 | 2.532 | 0.000 |
| NUSAP1 | 2.522 | 0.000 |
| KIF15 | 2.513 | 0.000 |
| KPNA2 | 2.511 | 0.000 |
| RACGAP1 | 2.496 | 0.000 |
| GBP5 | 2.488 | 0.051 |
| KIFC1 | 2.481 | 0.000 |
| ECT2 | 2.481 | 0.000 |
| EXO1 | 2.478 | 0.000 |
| HIST1H3H | 2.476 | 0.000 |
| TDO2 | 2.473 | 0.000 |
| NDC80 | 2.468 | 0.000 |
| BLM | 2.466 | 0.000 |
| IDO1 | 2.426 | 0.085 |
| RAD51AP1 | 2.412 | 0.000 |
| C12orf48 | 2.407 | 0.000 |
| GGH | 2.399 | 0.000 |
| CASC5 | 2.393 | 0.000 |
| HIST1H3B | 2.384 | 0.000 |
| KIAA0101 | 2.365 | 0.000 |
| MYBL1 | 2.362 | 0.000 |
| HIST1H2AI | 2.348 | 0.000 |
| TRIP13 | 2.344 | 0.000 |
| KIF2C | 2.343 | 0.000 |
| FAM111B | 2.342 | 0.000 |
| SNORD78 | 2.340 | 0.000 |
| MCM10 | 2.338 | 0.000 |
| CCNE2 | 2.323 | 0.000 |
| FAM72D | 2.321 | 0.000 |
| CCNE1 | 2.313 | 0.000 |
| SNORD75 | 2.305 | 0.000 |
| FAM72D | 2.302 | 0.000 |
| FAM72D | 2.300 | 0.000 |
| CCNB1 | 2.297 | 0.000 |
| KIAA1524 | 2.291 | 0.000 |
| CHEK1 | 2.291 | 0.000 |
| SMC4 | 2.283 | 0.000 |
| ASF1B | 2.271 | 0.000 |
| GPSM2 | 2.268 | 0.000 |
| CDKN2A | 2.265 | 0.000 |
| FANCI | 2.253 | 0.000 |
| NEIL3 | 2.244 | 0.000 |
| SPC25 | 2.199 | 0.000 |
| KIFC1 | 2.194 | 0.000 |
| GSDMC | 2.189 | 0.114 |
| CDCA2 | 2.185 | 0.000 |
| RRM2 | 2.181 | 0.000 |
| SQLE | 2.177 | 0.000 |
| HORMAD1 | 2.175 | 0.842 |
| ARHGAP11B | 2.167 | 0.000 |
| SLC7A11 | 2.161 | 0.000 |
| MTBP | 2.155 | 0.000 |
| UHRF1 | 2.155 | 0.000 |
| FAM83D | 2.152 | 0.000 |
| DNA2 | 2.136 | 0.000 |
| SPAG5 | 2.134 | 0.000 |
| CIT | 2.134 | 0.000 |
| POLQ | 2.132 | 0.000 |
| CDCA8 | 2.124 | 0.000 |
| CDC45 | 2.123 | 0.000 |
| HIST2H2AB | 2.121 | 0.000 |
| TK1 | 2.114 | 0.000 |

| | | |
|---|---|---|
| SCGB1D2 | 0.330 | 0.000 |
| IGF1 | 0.332 | 0.000 |
| OR5P2 | 0.333 | 0.000 |
| IGJ | 0.334 | 0.114 |
| ABCA10 | 0.335 | 0.000 |
| CNTNAP3 | 0.335 | 0.000 |
| FGF10 | 0.336 | 0.000 |
| HMGCS2 | 0.337 | 0.000 |
| C5orf4 | 0.339 | 0.000 |
| SRPX | 0.339 | 0.000 |
| C2orf40 | 0.341 | 0.000 |
| SNORD114-3 | 0.341 | 0.000 |
| TNS4 | 0.345 | 0.000 |
| SLC28A3 | 0.346 | 0.000 |
| NRG1 | 0.347 | 0.000 |
| TFF1 | 0.348 | 0.000 |
| EGR3 | 0.349 | 0.000 |
| OR5P3 | 0.350 | 0.000 |
| HSPB6 | 0.350 | 0.000 |
| TP63 | 0.352 | 0.000 |
| ST8SIA6 | 0.354 | 0.000 |
| ERBB4 | 0.354 | 0.051 |
| CNN1 | 0.355 | 0.000 |
| PLEKHH2 | 0.355 | 0.000 |
| PIK3C2G | 0.358 | 0.000 |
| C21orf34 | 0.358 | 0.000 |
| IGF2 | 0.359 | 0.000 |
| HLF | 0.359 | 0.000 |
| KLHL13 | 0.360 | 0.000 |
| LHFP | 0.360 | 0.000 |
| SYNPO2 | 0.361 | 0.000 |
| FHL1 | 0.362 | 0.000 |
| C21orf15 | 0.363 | 0.000 |
| RUNX1T1 | 0.363 | 0.000 |
| DCLK1 | 0.364 | 0.000 |
| ADH1C | 0.366 | 0.000 |
| FMOD | 0.368 | 0.000 |
| TSPAN7 | 0.371 | 0.000 |
| LRRN1 | 0.371 | 0.051 |
| FOSB | 0.372 | 0.000 |
| MAMDC2 | 0.372 | 0.000 |
| AREG | 0.373 | 0.000 |
| MME | 0.373 | 0.000 |
| LPL | 0.377 | 0.000 |
| CHL1 | 0.378 | 0.000 |
| F2RL2 | 0.382 | 0.000 |
| DMD | 0.383 | 0.000 |
| SEMA5A | 0.383 | 0.000 |
| STC2 | 0.383 | 0.000 |
| TCN1 | 0.384 | 0.114 |
| MMP16 | 0.385 | 0.000 |
| ALDH1A1 | 0.390 | 0.000 |
| LPHN3 | 0.390 | 0.000 |
| NTN4 | 0.391 | 0.000 |
| MEG3 | 0.391 | 0.000 |
| TGFBR3 | 0.392 | 0.000 |
| RELN | 0.393 | 0.000 |
| COL6A6 | 0.393 | 0.000 |
| ABI3BP | 0.395 | 0.000 |
| VSNL1 | 0.396 | 0.000 |
| SLIT3 | 0.396 | 0.000 |
| KRT23 | 0.396 | 0.171 |
| ANGPTL1 | 0.396 | 0.000 |
| AQP1 | 0.399 | 0.000 |

| | | |
|---|---|---|
| A2ML1 | 2.113 | 0.171 |
| ADAMDEC1 | 2.111 | 0.114 |
| FANCB | 2.110 | 0.000 |
| KIF20B | 2.097 | 0.000 |
| SNRPG | 2.083 | 0.000 |
| PLK4 | 2.082 | 0.000 |
| MAD2L1 | 2.074 | 0.000 |
| CHML | 2.068 | 0.000 |
| MMP1 | 2.066 | 0.215 |
| HMMR | 2.063 | 0.000 |
| HIST1H2BF | 2.062 | 0.000 |
| SNORA71C | 2.062 | 0.000 |
| DIAPH3 | 2.062 | 0.000 |
| C15orf23 | 2.060 | 0.000 |
| HIST1H1B | 2.059 | 0.000 |
| CXCL9 | 2.057 | 0.628 |
| HIST2H3A | 2.053 | 0.000 |
| HIST2H3A | 2.053 | 0.000 |
| GEN1 | 2.049 | 0.000 |
| HIST1H3I | 2.047 | 0.000 |
| SNORD27 | 2.045 | 0.000 |
| GAS2L3 | 2.036 | 0.000 |
| E2F5 | 2.035 | 0.000 |
| CENPN | 2.034 | 0.000 |
| CENPO | 2.034 | 0.000 |
| CENPA | 2.025 | 0.000 |
| FBXO45 | 2.024 | 0.000 |
| RSAD2 | 2.023 | 0.114 |
| S100P | 2.018 | 0.000 |
| GINS2 | 2.016 | 0.000 |
| HJURP | 2.013 | 0.000 |
| LAPTM4B | 2.013 | 0.000 |
| HIST1H2BM | 2.004 | 0.000 |
| DNAH14 | 2.003 | 0.000 |
| LIN9 | 2.002 | 0.000 |

| | | |
|---|---|---|
| BOC | 0.399 | 0.000 |
| C8orf84 | 0.401 | 0.000 |
| PLAT | 0.401 | 0.000 |
| SFRP4 | 0.401 | 0.000 |
| HBB | 0.403 | 0.171 |
| NOSTRIN | 0.405 | 0.000 |
| DCN | 0.405 | 0.000 |
| EMCN | 0.407 | 0.000 |
| FREM1 | 0.408 | 0.000 |
| ZNF385D | 0.408 | 0.000 |
| DARC | 0.409 | 0.000 |
| LAMA2 | 0.412 | 0.000 |
| CMYA5 | 0.412 | 0.000 |
| RBP4 | 0.413 | 0.842 |
| EGR1 | 0.413 | 0.000 |
| MEOX2 | 0.414 | 0.000 |
| PLIN1 | 0.416 | 0.051 |
| SEMA6D | 0.417 | 0.000 |
| TFAP2B | 0.417 | 0.467 |
| OXTR | 0.417 | 0.000 |
| OVCH2 | 0.418 | 0.000 |
| NDRG2 | 0.419 | 0.000 |
| EBF1 | 0.420 | 0.000 |
| CACHD1 | 0.420 | 0.000 |
| TAT | 0.420 | 0.628 |
| CXCL14 | 0.421 | 0.000 |
| PLIN4 | 0.423 | 0.093 |
| PHYHD1 | 0.424 | 0.000 |
| SPRY2 | 0.424 | 0.000 |
| FMO1 | 0.425 | 0.000 |
| MMP7 | 0.425 | 0.842 |
| PAMR1 | 0.425 | 0.000 |
| LMOD1 | 0.425 | 0.000 |
| GLI3 | 0.426 | 0.000 |
| FMO2 | 0.426 | 0.114 |
| SPATA18 | 0.427 | 0.000 |
| CNTN1 | 0.427 | 0.000 |
| SPARCL1 | 0.429 | 0.000 |
| LPAR1 | 0.429 | 0.000 |
| ANK2 | 0.429 | 0.000 |
| LIFR | 0.430 | 0.000 |
| LYVE1 | 0.430 | 0.467 |
| TTC6 | 0.431 | 0.628 |
| SLC7A2 | 0.431 | 0.215 |
| SORBS1 | 0.431 | 0.000 |
| GEM | 0.431 | 0.000 |
| TNXB | 0.433 | 0.000 |
| DPT | 0.433 | 0.000 |
| GPX2 | 0.434 | 0.000 |
| NPY1R | 0.434 | 1.352 |
| PI15 | 0.436 | 2.578 |
| JAM2 | 0.436 | 0.000 |
| GABRP | 0.437 | 2.090 |
| CXCL12 | 0.439 | 0.000 |
| FRZB | 0.439 | 0.000 |
| PDLIM3 | 0.440 | 0.000 |
| MMRN1 | 0.440 | 0.000 |
| SAA1 | 0.442 | 0.000 |
| HPSE2 | 0.442 | 0.000 |
| CAV1 | 0.443 | 0.000 |
| NRG1 | 0.443 | 0.000 |
| PROS1 | 0.443 | 0.000 |
| WLS | 0.443 | 0.000 |
| SEMA3C | 0.444 | 0.000 |

| | | |
|---|---|---|
| RHOJ | 0.444 | 0.000 |
| CLDN8 | 0.445 | 0.093 |
| ZFHX4 | 0.448 | 0.000 |
| TMEM47 | 0.448 | 0.000 |
| PDGFD | 0.449 | 0.000 |
| SORBS2 | 0.450 | 0.000 |
| MATN2 | 0.450 | 0.000 |
| AKAP12 | 0.451 | 0.000 |
| AMIGO2 | 0.452 | 0.000 |
| PCDH18 | 0.452 | 0.000 |
| SESN3 | 0.452 | 0.000 |
| EGR2 | 0.453 | 0.000 |
| MPPED2 | 0.453 | 0.000 |
| SLIT2 | 0.454 | 0.000 |
| ETV1 | 0.454 | 0.000 |
| ADAMTS5 | 0.455 | 0.000 |
| PALMD | 0.455 | 0.000 |
| NOVA1 | 0.457 | 0.093 |
| SMOC2 | 0.458 | 0.000 |
| CPA3 | 0.458 | 0.000 |
| MYLK | 0.458 | 0.000 |
| SEMA3D | 0.460 | 0.000 |
| HMCN1 | 0.460 | 0.000 |
| C10orf81 | 0.461 | 0.093 |
| SPRY1 | 0.461 | 0.000 |
| FOS | 0.462 | 0.000 |
| ARHGEF38 | 0.462 | 0.093 |
| ITM2A | 0.463 | 0.000 |
| PROS1 | 0.463 | 0.000 |
| BICC1 | 0.465 | 0.000 |
| GAGE12J | 0.465 | 0.000 |
| GPD1 | 0.465 | 0.114 |
| ADH1A | 0.466 | 0.051 |
| CCDC80 | 0.466 | 0.093 |
| FAT4 | 0.467 | 0.000 |
| FHL2 | 0.468 | 0.000 |
| ITIH5 | 0.469 | 0.000 |
| TNC | 0.469 | 0.000 |
| AGR2 | 0.470 | 2.578 |
| GHR | 0.470 | 0.051 |
| PDZRN3 | 0.471 | 0.000 |
| DUSP6 | 0.471 | 0.000 |
| SYNM | 0.471 | 0.051 |
| TAGLN | 0.472 | 0.000 |
| SNORD115-11 | 0.474 | 0.366 |
| ABLIM3 | 0.474 | 0.000 |
| PODN | 0.474 | 0.000 |
| C7orf58 | 0.474 | 0.000 |
| HBA1 | 0.475 | 0.215 |
| HBA2 | 0.475 | 0.215 |
| GALNTL1 | 0.475 | 0.000 |
| AOC3 | 0.475 | 0.000 |
| NR2F1 | 0.476 | 0.000 |
| FXYD1 | 0.476 | 0.000 |
| CLDN11 | 0.477 | 0.000 |
| ISM1 | 0.477 | 0.000 |
| PLSCR4 | 0.478 | 0.000 |
| GPAM | 0.479 | 0.114 |
| DLC1 | 0.479 | 0.000 |
| ADAMTS9 | 0.480 | 0.000 |
| THRSP | 0.481 | 0.093 |
| IL33 | 0.482 | 0.051 |
| RIC3 | 0.482 | 0.000 |
| CCL14-CCL15 | 0.482 | 0.000 |

| | | |
|---|---|---|
| CCL21 | 0.482 | 0.114 |
| GRAMD3 | 0.483 | 0.000 |
| THRB | 0.483 | 0.000 |
| LDB2 | 0.483 | 0.000 |
| CA12 | 0.483 | 0.842 |
| STAC2 | 0.483 | 0.051 |
| FBLN5 | 0.483 | 0.000 |
| COL17A1 | 0.484 | 0.000 |
| SNORD113-4 | 0.484 | 0.000 |
| PDK4 | 0.484 | 0.171 |
| RERG | 0.485 | 0.269 |
| C7orf63 | 0.485 | 0.000 |
| MEIS2 | 0.485 | 0.000 |
| OMD | 0.485 | 0.114 |
| SERPINA3 | 0.485 | 0.467 |
| C13orf33 | 0.486 | 0.171 |
| PDE1C | 0.488 | 0.000 |
| CX3CL1 | 0.488 | 0.000 |
| FAT2 | 0.489 | 0.000 |
| NFIA | 0.489 | 0.000 |
| FLRT2 | 0.489 | 0.000 |
| C15orf51 | 0.490 | 0.000 |
| C15orf51 | 0.490 | 0.000 |
| SLC12A2 | 0.491 | 0.000 |
| PDE5A | 0.492 | 0.000 |
| ZFP36 | 0.492 | 0.000 |
| KRT5 | 0.492 | 0.842 |
| SNORD114-2 | 0.492 | 0.000 |
| MAOA | 0.493 | 0.171 |
| FRMD6 | 0.493 | 0.000 |
| ADAMTS15 | 0.494 | 0.000 |
| CAV2 | 0.494 | 0.093 |
| C5orf36 | 0.494 | 0.000 |
| MGP | 0.497 | 0.051 |
| ACACB | 0.497 | 0.051 |
| LOC349196 | 0.497 | 0.000 |
| LOC349196 | 0.497 | 0.000 |
| CX3CR1 | 0.497 | 0.000 |
| LUZP2 | 0.497 | 0.000 |
| CREB5 | 0.498 | 0.000 |
| PER2 | 0.498 | 0.000 |
| GRP | 0.499 | 0.000 |
| NAV3 | 0.499 | 0.000 |
| SAMD5 | 0.499 | 0.000 |

**Supplemental Table 10:** Genes that were differentially expressed between 2-hydroxyglutarate (2HG)-high and 2HG-low tumors (top 33% versus lowest 33%)

| UP in 2HG-high tumor | | |
|---|---|---|
| **Gene Symbol** | **Fold Change** | **q-value (%)** |
| CXCL10 | 2.573 | 0.219 |
| ANLN | 2.538 | 0.000 |
| NUF2 | 2.508 | 0.000 |
| BUB1 | 2.508 | 0.000 |
| TOP2A | 2.477 | 0.000 |
| CDC6 | 2.469 | 0.000 |
| TTK | 2.429 | 0.000 |
| CENPF | 2.421 | 0.000 |
| ERBB2 | 2.406 | 0.000 |
| TPX2 | 2.367 | 0.000 |
| FOXM1 | 2.365 | 0.000 |
| ASPM | 2.348 | 0.000 |
| MMP1 | 2.348 | 0.272 |
| SGOL1 | 2.288 | 0.000 |
| CDK1 | 2.280 | 0.000 |
| KIF20A | 2.246 | 0.000 |
| CENPE | 2.241 | 0.000 |
| AURKA | 2.210 | 0.000 |
| CCNB2 | 2.195 | 0.000 |
| GGH | 2.172 | 0.000 |
| ESRP1 | 2.167 | 0.000 |
| GRB7 | 2.164 | 0.000 |
| CDC20 | 2.160 | 0.000 |
| MELK | 2.148 | 0.000 |
| SKA3 | 2.134 | 0.000 |
| UBE2T | 2.132 | 0.000 |
| CKAP2L | 2.124 | 0.000 |
| HORMAD1 | 2.123 | 2.201 |
| XRCC2 | 2.123 | 0.000 |
| TDO2 | 2.115 | 0.272 |
| CXCL11 | 2.114 | 1.078 |
| C15orf42 | 2.098 | 0.000 |
| CEP55 | 2.092 | 0.000 |
| DEPDC1 | 2.062 | 0.000 |
| RAD21 | 2.058 | 0.000 |
| KIF14 | 2.057 | 0.000 |
| SLC7A11 | 2.057 | 0.000 |
| KIF23 | 2.053 | 0.000 |
| GSDMB | 2.044 | 0.121 |
| CCNE2 | 2.041 | 0.000 |
| PLK1 | 2.039 | 0.121 |
| MYBL1 | 2.033 | 0.272 |
| FAM83B | 2.033 | 0.272 |
| CKS2 | 2.029 | 0.121 |
| NCAPH | 2.025 | 0.000 |
| DKK1 | 2.014 | 0.889 |
| STIL | 2.012 | 0.000 |
| ATAD2 | 2.011 | 0.000 |
| CD24 | 2.006 | 0.219 |
| PRAME | 2.001 | 0.121 |

| DOWN in 2HG-high tumor | | |
|---|---|---|
| **Gene Symbol** | **Fold Change** | **q-value (%)** |
| FABP4 | 0.098 | 0.000 |
| CD36 | 0.206 | 0.000 |
| ADH1B | 0.231 | 0.000 |
| ADIPOQ | 0.241 | 0.000 |
| RBP4 | 0.279 | 0.000 |
| OGN | 0.282 | 0.000 |
| PROL1 | 0.293 | 0.000 |
| FIGF | 0.294 | 0.000 |
| GPC3 | 0.296 | 0.000 |
| AK5 | 0.303 | 0.000 |
| PIGR | 0.304 | 0.000 |
| MIR145 | 0.311 | 0.000 |
| CHRDL1 | 0.315 | 0.000 |
| LPL | 0.323 | 0.000 |
| ABCA9 | 0.330 | 0.000 |
| SCGB2A2 | 0.331 | 0.272 |
| ADH1C | 0.335 | 0.000 |
| HBB | 0.336 | 0.000 |
| SNORD114-3 | 0.338 | 0.000 |
| ABCA8 | 0.339 | 0.000 |
| STEAP4 | 0.340 | 0.000 |
| IGSF10 | 0.341 | 0.000 |
| ABCA6 | 0.341 | 0.000 |
| HSPB6 | 0.342 | 0.000 |
| LYVE1 | 0.343 | 0.000 |
| TGFBR3 | 0.343 | 0.000 |
| C7 | 0.351 | 0.000 |
| MFAP4 | 0.359 | 0.000 |
| PLIN1 | 0.360 | 0.000 |
| COL14A1 | 0.361 | 0.000 |
| TNXB | 0.363 | 0.000 |
| TNXA | 0.365 | 0.000 |
| FOSB | 0.366 | 0.000 |
| PIP | 0.368 | 0.889 |
| CYP4X1 | 0.368 | 0.167 |
| SVEP1 | 0.372 | 0.000 |
| CCL28 | 0.373 | 0.000 |
| PLIN4 | 0.374 | 0.000 |
| APOD | 0.376 | 0.000 |
| FHL1 | 0.378 | 0.000 |
| SRPX | 0.378 | 0.000 |
| HBA1 | 0.386 | 0.000 |
| HBA2 | 0.386 | 0.000 |
| MYH11 | 0.387 | 0.000 |
| GPD1 | 0.391 | 0.000 |
| DPT | 0.395 | 0.000 |
| CHL1 | 0.402 | 0.000 |
| IGF1 | 0.402 | 0.000 |
| KRT15 | 0.403 | 0.000 |
| FGF10 | 0.407 | 0.000 |
| GSTM5 | 0.408 | 0.000 |
| C5orf4 | 0.409 | 0.000 |
| TSPAN7 | 0.410 | 0.000 |

|  |  |  |
|---|---|---|
| SLC28A3 | 0.413 | 0.000 |
| LTF | 0.414 | 0.219 |
| SAA1 | 0.417 | 0.000 |
| CNTNAP3 | 0.417 | 0.000 |
| PLEKHH2 | 0.417 | 0.000 |
| ERBB4 | 0.418 | 0.219 |
| C13orf33 | 0.418 | 0.000 |
| TSHZ2 | 0.419 | 0.000 |
| PDK4 | 0.420 | 0.000 |
| TCN1 | 0.423 | 0.521 |
| ADH1A | 0.423 | 0.000 |
| PDLIM3 | 0.427 | 0.000 |
| LHFP | 0.427 | 0.000 |
| ANGPTL1 | 0.428 | 0.000 |
| TIMP4 | 0.433 | 0.000 |
| AOC3 | 0.436 | 0.000 |
| RBMS3 | 0.439 | 0.000 |
| ALDH1A1 | 0.440 | 0.000 |
| DARC | 0.440 | 0.000 |
| ACACB | 0.441 | 0.000 |
| COL6A6 | 0.442 | 0.000 |
| AQP7P1 | 0.442 | 0.000 |
| WIF1 | 0.444 | 0.521 |
| ANPEP | 0.445 | 0.000 |
| SLIT3 | 0.445 | 0.000 |
| EBF1 | 0.448 | 0.000 |
| CIDEC | 0.450 | 0.000 |
| ABCA10 | 0.451 | 0.000 |
| CYP4Z1 | 0.453 | 3.706 |
| AKR1C2 | 0.454 | 0.219 |
| ZFP36 | 0.454 | 0.000 |
| CNN1 | 0.454 | 0.000 |
| CXCL12 | 0.456 | 0.000 |
| MMRN1 | 0.456 | 0.000 |
| CPA3 | 0.457 | 0.000 |
| GPAM | 0.458 | 0.000 |
| LIFR | 0.461 | 0.000 |
| DCN | 0.463 | 0.000 |
| ADAMTS1 | 0.464 | 0.000 |
| SYNPO2 | 0.465 | 0.000 |
| MAMDC2 | 0.465 | 0.000 |
| PPARG | 0.465 | 0.000 |
| EGR3 | 0.468 | 0.000 |
| SEMA3D | 0.470 | 0.000 |
| PALMD | 0.470 | 0.000 |
| ZNF385D | 0.473 | 0.000 |
| ADAMTS9 | 0.473 | 0.000 |
| SORBS1 | 0.473 | 0.000 |
| THRSP | 0.474 | 0.000 |
| GHR | 0.476 | 0.000 |
| FOS | 0.476 | 0.000 |
| CFD | 0.476 | 0.000 |
| NTRK2 | 0.479 | 0.272 |
| KIT | 0.481 | 0.272 |
| ABI3BP | 0.481 | 0.000 |
| FMOD | 0.482 | 0.000 |
| DMD | 0.482 | 0.000 |
| GLYAT | 0.485 | 0.000 |

| | | |
|---|---|---|
| AQP1 | 0.485 | 0.000 |
| IL33 | 0.487 | 0.000 |
| CAV2 | 0.489 | 0.000 |
| WLS | 0.489 | 0.000 |
| CDON | 0.489 | 0.000 |
| RUNX1T1 | 0.490 | 0.000 |
| EMCN | 0.491 | 0.000 |
| LAMA2 | 0.492 | 0.000 |
| SNORD114-2 | 0.492 | 0.000 |
| CAV1 | 0.493 | 0.000 |
| IGJ | 0.493 | 3.706 |
| LALBA | 0.495 | 1.512 |
| STC2 | 0.496 | 0.521 |
| GPX3 | 0.497 | 0.000 |
| LIPE | 0.497 | 0.000 |
| ANK2 | 0.499 | 0.000 |

**Supplemental Table 11:** Subgroup III gene expression signature is associated with survival in independent datasets

| | Univariable Cox regression | | | | Multivariable Cox regression[†] | | | |
|---|---|---|---|---|---|---|---|---|
| | **HR** | **95% CI** | ***P*** | ***n*** | **HR** | **95% CI** | ***P*** | ***n*** |
| **van de Vijver dataset** | | | | | | | | |
| **5-Year survival** | | | | | | | | |
| Subgroup III signature | | | | | | | | |
| Low | 1 | | | 148 | 1 | | | 148 |
| High | 4.26 | 2.13 - 8.56 | $4.6 \times 10^{-5}$ [§] | 147 | 2.73 | 1.29 - 5.79 | 0.009 [§] | 147 |
| **10-Year survival** | | | | | | | | |
| Subgroup III signature | | | | | | | | |
| Low | 1 | | | 148 | 1 | | | 148 |
| High | 3.59 | 2.13 - 6.06 | $1.7 \times 10^{-6}$ [§] | 147 | 2.74 | 1.56 - 4.80 | 0.0004 [§] | 147 |
| **Kao dataset** | | | | | | | | |
| **5-Year survival** | | | | | | | | |
| Subgroup III signature | | | | | | | | |
| Low | 1 | | | 164 | 1 | | | 164 |
| High | 2.27 | 1.25 - 4.11 | 0.007 [§] | 163 | 2.15 | 1.18 - 3.93 | 0.013 [§] | 163 |
| **10-Year survival** | | | | | | | | |
| Subgroup III signature | | | | | | | | |
| Low | 1 | | | 164 | 1 | | | 164 |
| High | 1.92 | 1.21 - 3.05 | 0.006 [§] | 163 | 1.80 | 1.13 - 2.89 | 0.014 [§] | 163 |

[†] Adjusted for ER status in *van de Vijver* dataset, where all patients are younger than 53 and at TNM stage I or II.
Adjusted for age at diagnosis, TNM stage and ER status in Kao dataset.

[§] $P < 0.05$

**Supplemental Table 12:** Gene expression signature of 2-hydroxyglutarate (2HG)-high tumors predicts survival in independent datasets

| | Univariable Cox regression | | | | Multivariable Cox regression[†] | | | |
|---|---|---|---|---|---|---|---|---|
| | **HR** | **95% CI** | ***P*** | ***n*** | **HR** | **95% CI** | ***P*** | ***n*** |
| **van de Vijver dataset** | | | | | | | | |
| **5-Year survival** | | | | | | | | |
| 2HG-high tumor signature | | | | | | | | |
| Low | 1 | | | 148 | 1 | | | 148 |
| High | 3.77 | 1.92 - 7.40 | 0.0001 [§] | 147 | 2.48 | 1.21 - 5.08 | 0.013 [§] | 147 |
| **10-Year survival** | | | | | | | | |
| 2HG-high tumor signature | | | | | | | | |
| Low | 1 | | | 148 | 1 | | | 148 |
| High | 3.08 | 1.86 - 5.11 | $1.3 \times 10^{-5}$ [§] | 147 | 2.36 | 1.38 - 4.03 | 0.002 [§] | 147 |
| **Kao dataset** | | | | | | | | |
| **5-Year survival** | | | | | | | | |
| 2HG-high tumor signature | | | | | | | | |
| Low | 1 | | | 164 | 1 | | | 164 |
| High | 2.26 | 1.25 - 4.10 | 0.007 [§] | 163 | 2.06 | 1.13 - 3.74 | 0.018 [§] | 163 |
| **10-Year survival** | | | | | | | | |
| 2HG-high tumor signature | | | | | | | | |
| Low | 1 | | | 164 | 1 | | | 164 |
| High | 1.80 | 1.14 - 2.86 | 0.012 [§] | 163 | 1.63 | 1.02 - 2.61 | 0.040 [§] | 163 |

[†] Adjusted for ER status in van de Vijver dataset, where all patients are younger than 53 and at TNM stage I or II.
Adjusted for age at diagnosis, TNM stage and ER status in Kao dataset.

[§] $P < 0.05$

**Supplemental Table 13:** Signaling pathways that are potentially altered in subgroup III tumors when compared with subgroup I tumors because of DNA methylation.

| **KEGG Pathway Term** | **Fold Enrichment** | ***P* value** | **FDR (%)** |
|---|---|---|---|
| Focal Adhesion | 3.41 | 0.01 | 13.5 |
| WNT Signaling | 3.89 | 0.02 | 15.7 |
| Calcium Signaling | 3.33 | 0.03 | 26.6 |
| Gap Junction | 4.39 | 0.06 | 45.6 |
| Glutathione Metabolism | 5.87 | 0.09 | 60.6 |

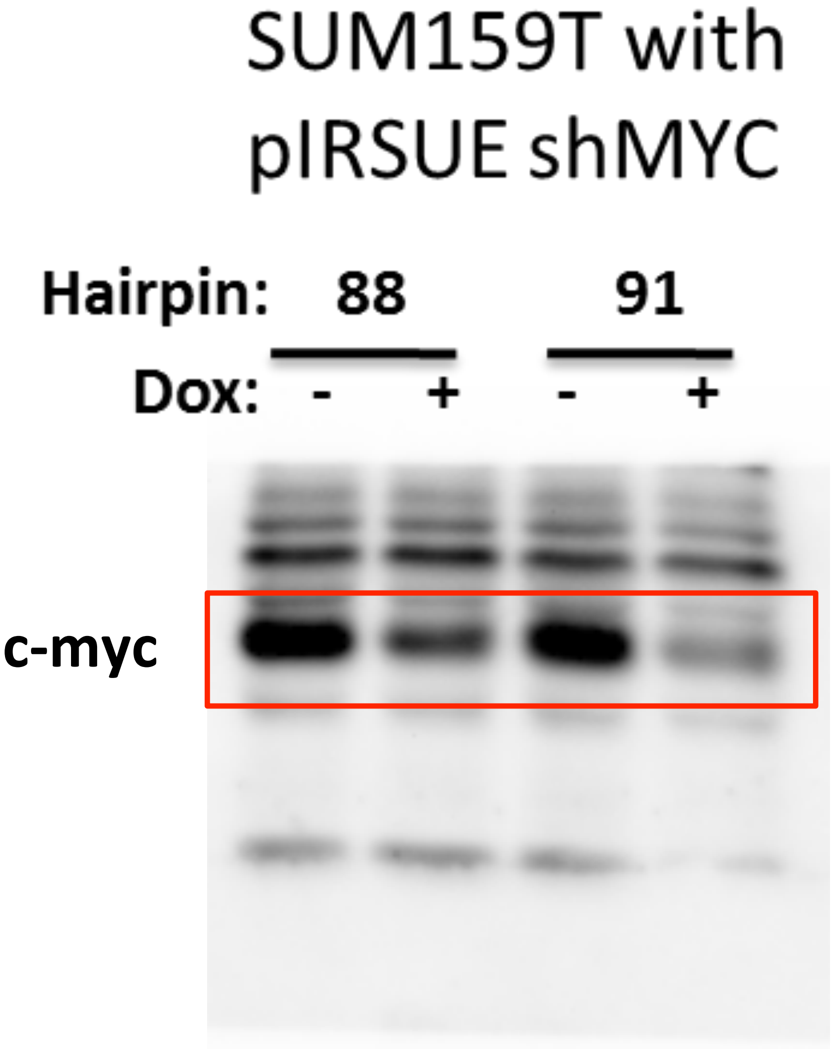
The shown KEGG pathway terms were enriched for genes that showed higher DNA methylation [log2(fold change) > 1, FDR < 5%] and lower expression levels [log2(fold change) < -0.5, FDR < 5%] in subgroup III compared with subgroup I (= 159 genes). FDR = false discovery rate.

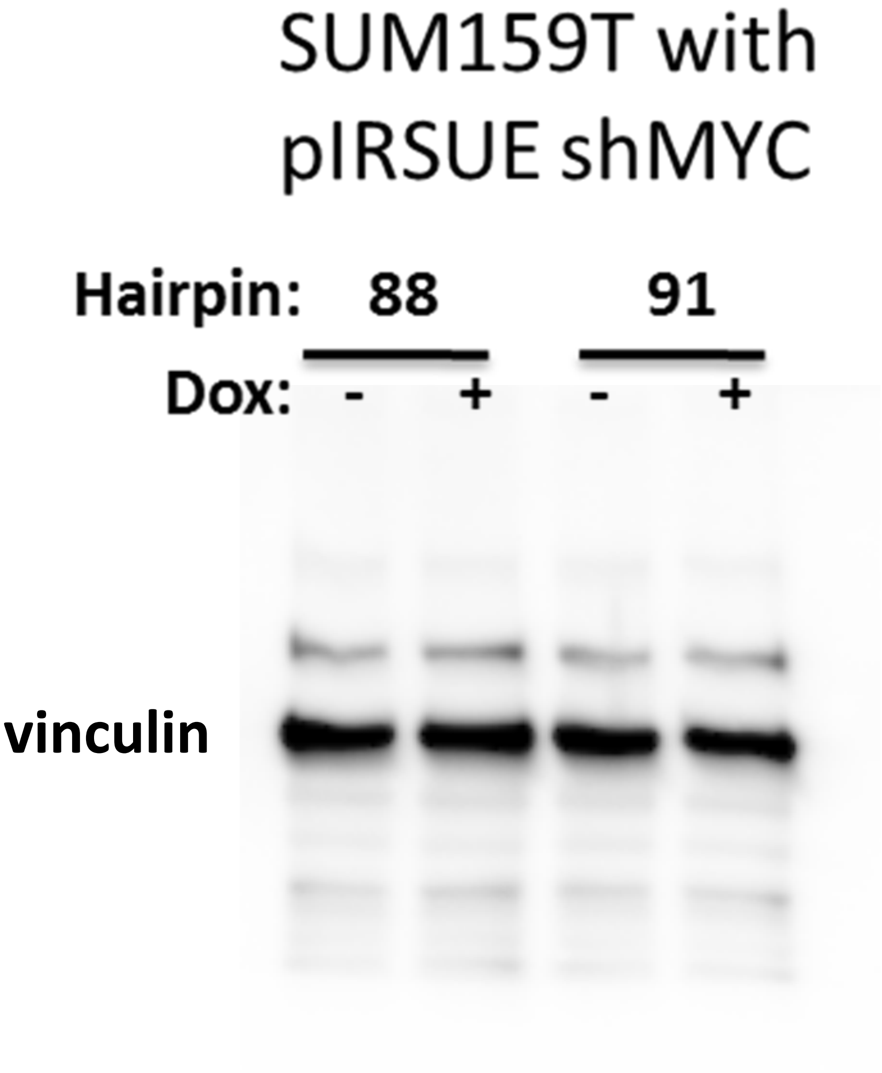
**Supplemental Table 14**: Genes that were differentially expressed between glutamine-low (n = 15) and glutamine-high (n = 15) tumors (median as cutoff)

| UP in Glutamine-low tumor | | |
|---|---|---|
| **Gene Symbol** | **Fold Change** | **q-value (%)** |
| C8orf85 | 2.524 | 9.244 |
| SNORA38B | 2.425 | 9.244 |
| OVOS | 2.182 | 9.244 |
| FAM72D | 2.082 | 9.244 |
| CENPF | 2.040 | 9.244 |
| NUF2 | 2.039 | 9.244 |
| EPCAM | 2.034 | 9.244 |
| PSAT1 | 2.016 | 9.244 |

| DOWN in Glutamine-low tumor | | |
|---|---|---|
| **Gene Symbol** | **Fold Change** | **q-value (%)** |
| CYP4Z1 | 0.175 | 0.000 |
| CYP4Z2P | 0.183 | 0.000 |
| CYP4X1 | 0.339 | 5.293 |
| C4A | 0.341 | 0.000 |
| CPA3 | 0.360 | 0.000 |
| SFRP2 | 0.361 | 0.000 |
| ADIPOQ | 0.387 | 9.244 |
| F2RL2 | 0.395 | 3.939 |
| CILP | 0.415 | 3.736 |
| GPC3 | 0.418 | 6.843 |
| HLA-DQB1 | 0.424 | 3.736 |
| SIDT1 | 0.427 | 3.736 |
| SLC44A4 | 0.438 | 7.841 |
| SFRP4 | 0.439 | 3.736 |
| SLC7A8 | 0.443 | 0.000 |
| AR | 0.445 | 3.939 |
| NAT1 | 0.450 | 6.843 |
| MFAP4 | 0.458 | 6.952 |
| FOXA1 | 0.485 | 7.841 |
| CCDC80 | 0.486 | 6.843 |
| OMD | 0.490 | 5.293 |
| CHIT1 | 0.490 | 5.293 |
| GATA3 | 0.490 | 9.244 |
| ARHGEF38 | 0.500 | 5.293 |
| VGLL3 | 0.500 | 5.293 |

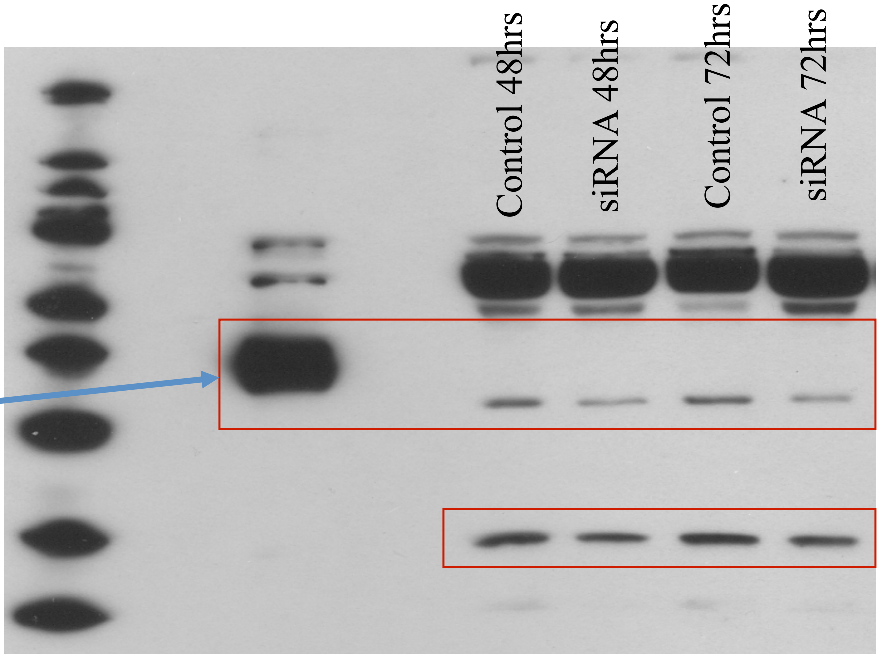
# Uncut blots for Figure 8E





# Uncut blots for Figure 9A: Knockdown of ADHFE1 in MDA-MB-231 cells transfected with the ADHFE1-targeting siRNA #1 (after 48 and 72 hrs)

## Western for ADHFE1

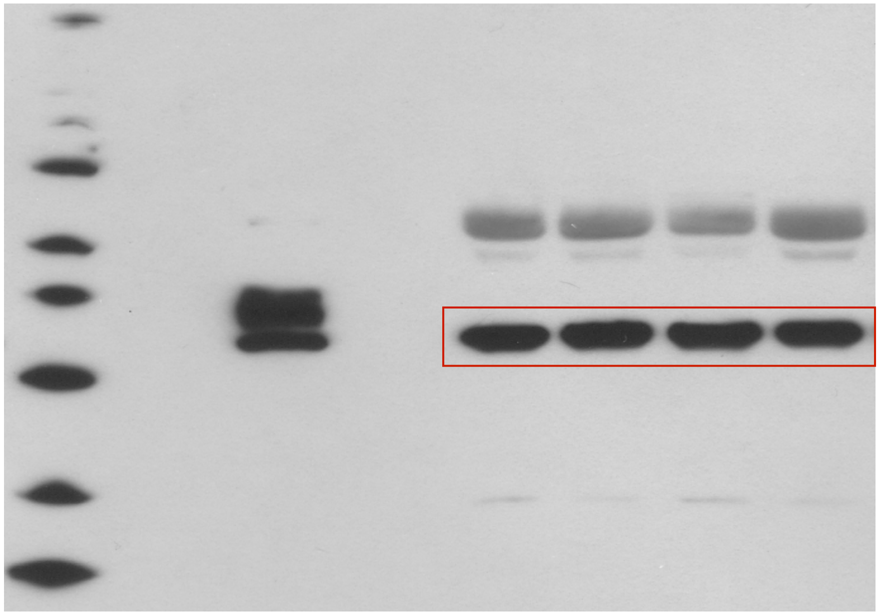


**Positive Control:**
**Protein encoded by higher molecular weight Flag-tagged ADHFE1-transgene overexpressed in MCF7 cells**

**ADHFE1 (45-50 KDa)**
**at predicted molecular size**

**ADHFE1v (27-30 KDa)**
**predicted spice variant**

## Western for beta-actin

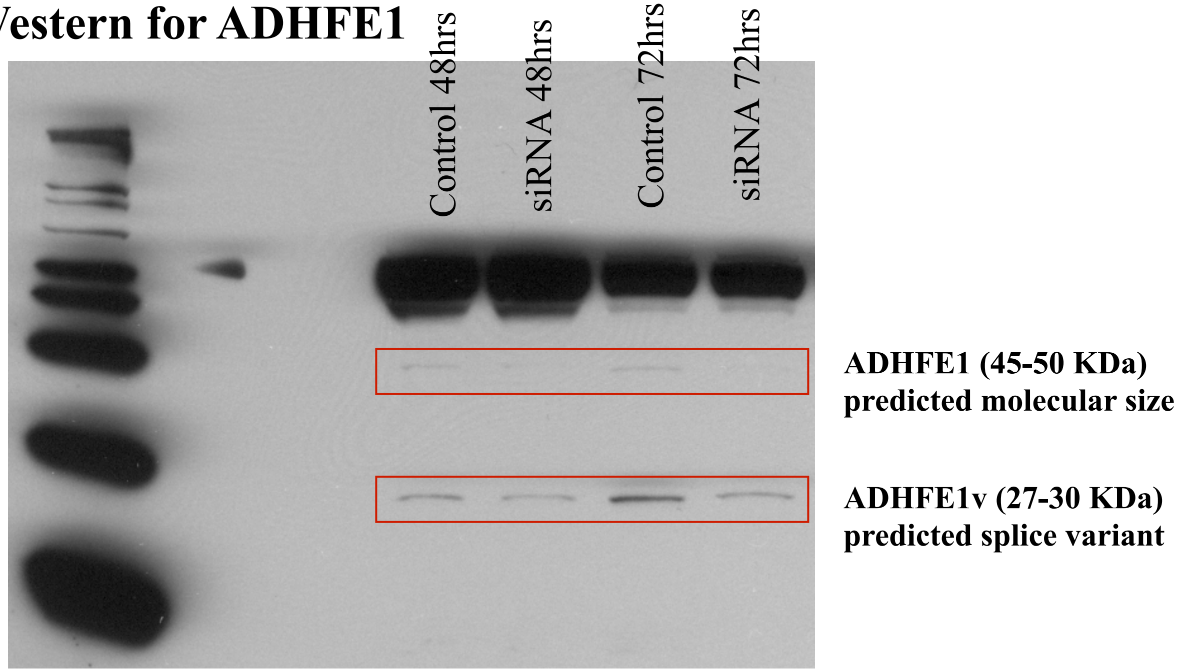


**β-Actin**

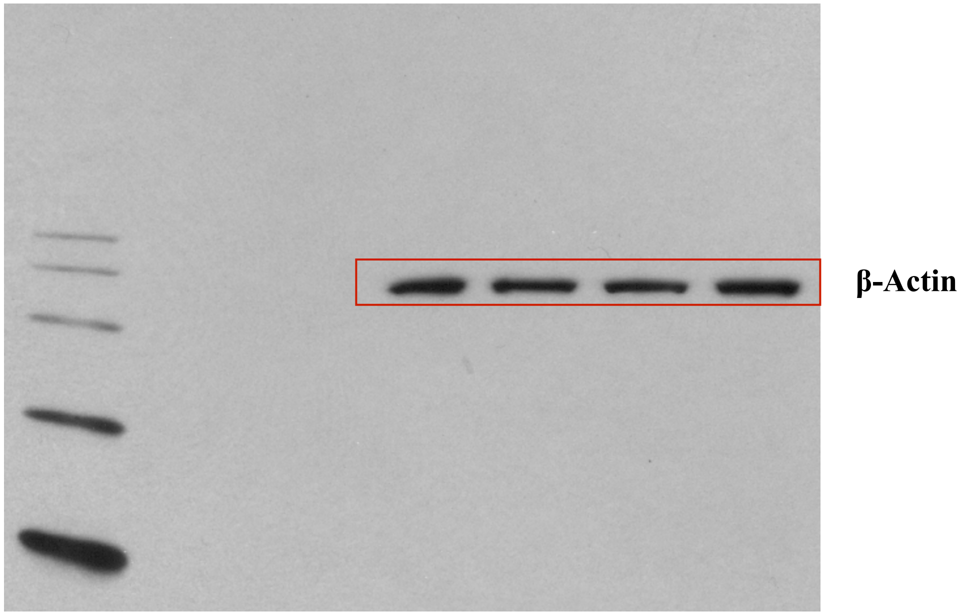
## Uncut blots for Figure 9A:

**Knockdown of ADHFE1 in MDA-MB-231 cells transfected with the ADHFE1-targeting siRNA #2 (after 48 and 72 hrs)**

**Western for ADHFE1**

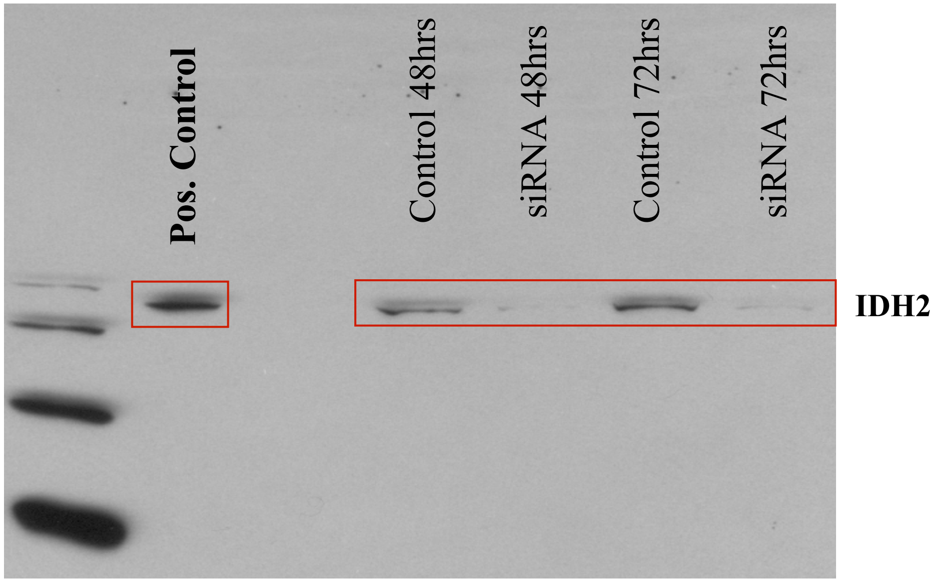


**Western for beta-actin**

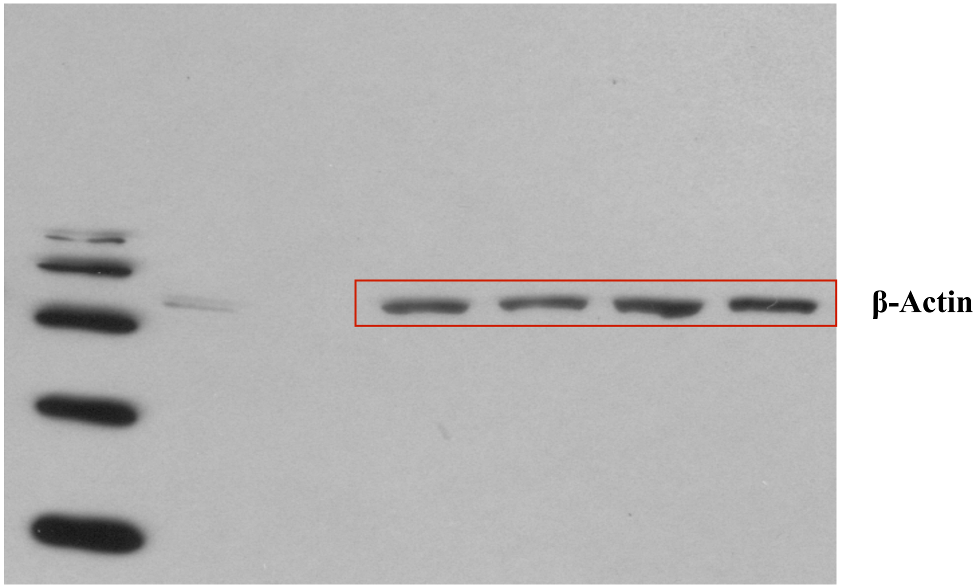


**Uncut blots for Figure 9B:** **Western for IDH2 protein using extracts from MDA-MB-231 cells transfected with the siRNA targeting IDH2 (after 48 and 72 hrs)**

**Western for IDH2**



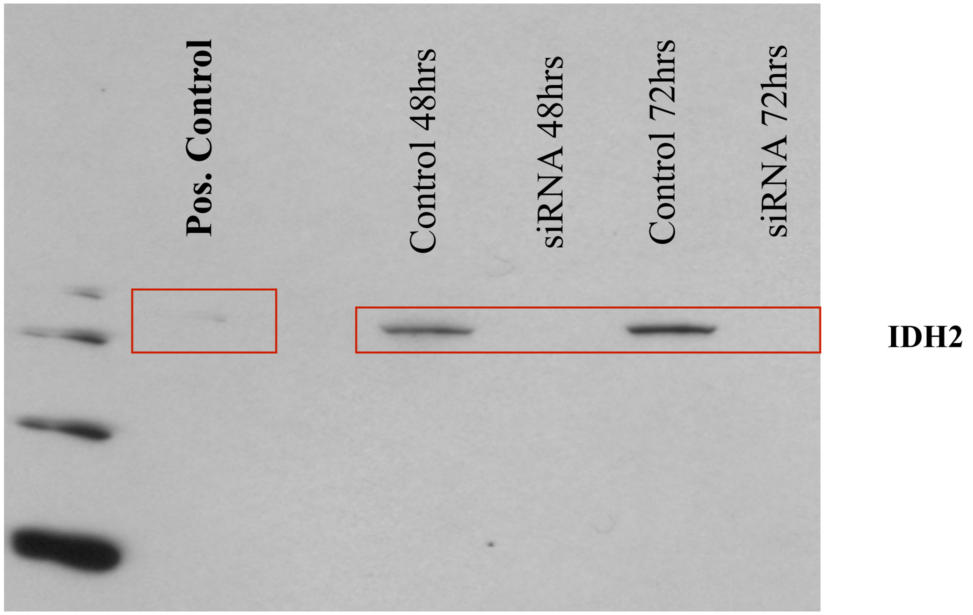
**Western for beta-actin**



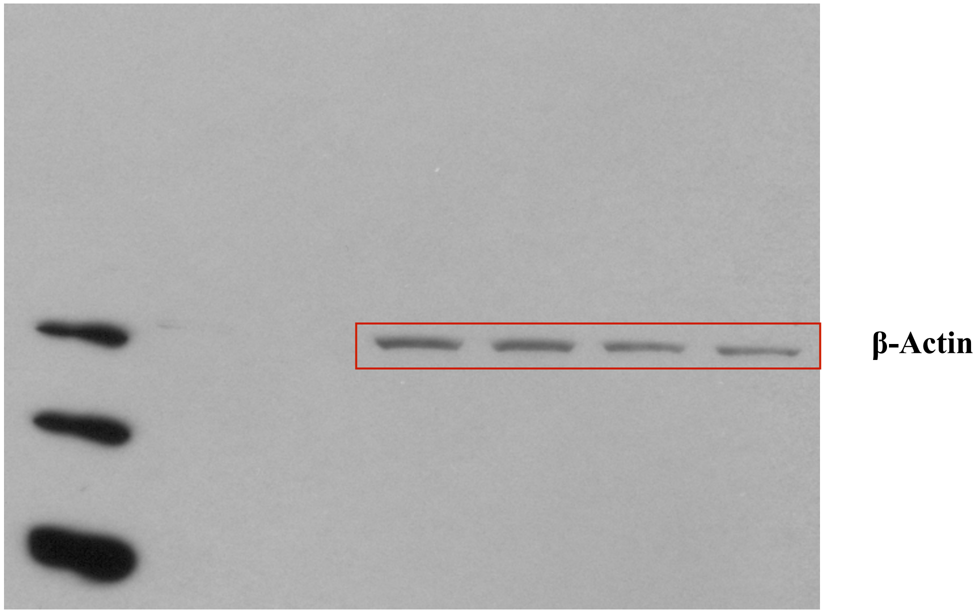
**Uncut blots for Figure 9B:**

**Western for IDH2 protein using extracts from SUM159T cells transfected with the siRNA targeting IDH2 (after 48 and 72 hrs)**

**Western for IDH2**

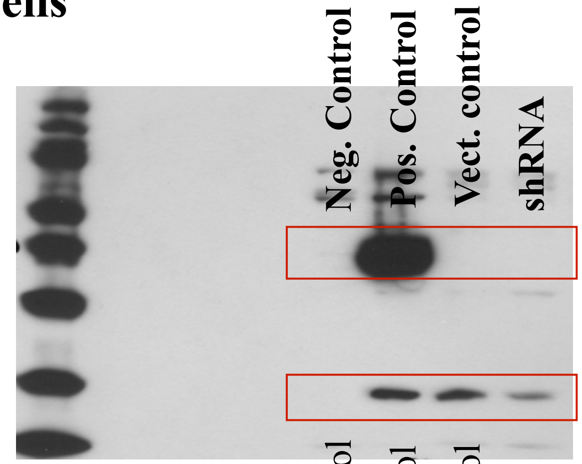


**Western for beta-actin**



# Uncut blots for Figure 9D: Knockdown of ADHFE1 using shRNA in mitochondrial lysates from MDA-MB-231 cells

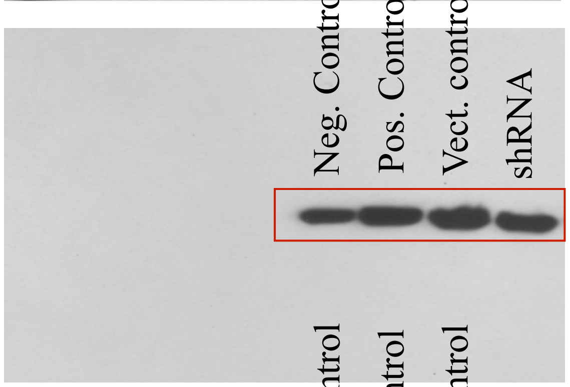
## Western for ADHFE1



**Flag-tagged ADHFE1 transgene expressed in MCF7 cells**

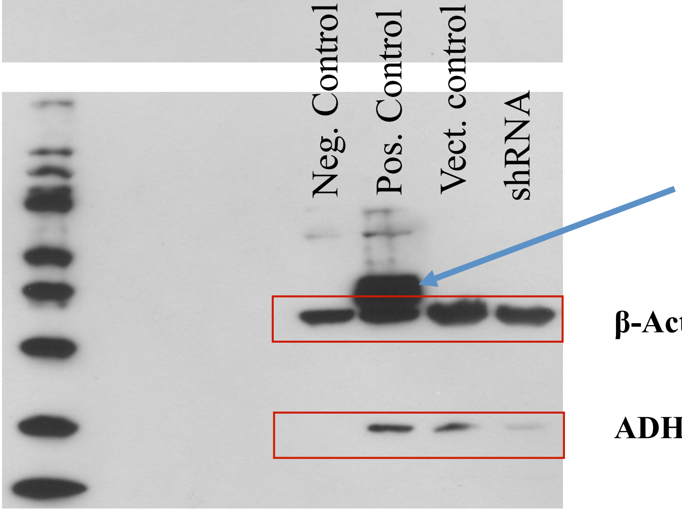
**ADHFE1v (27-30 KDa)**

## Western for IDH2 (loading control)



**IDH2 (mitochondrial enzyme)**

## Western for beta actin (2nd loading control)



Flag-tagged ADHFE1 transgene

**β-Actin**

**ADHFE1v (27-30 KDa)**

| Tumor LHC | Tumor Cel file GSE37751 | DNA methylation subgroup | Myc signature | Metabolon Tumor ID | Normal LHC | Normal Cel file GSE37751 | Metabolon Normal ID |
|---|---|---|---|---|---|---|---|
| 3264 | 3264 | 1 | 0 | 499657 | 3265 | 3265 | 499527 |
| 3631 | 3631 | 3 | 1 | 499659 | 3630 | 3630 | 499529 |
| 3661 | 3661 | 1 | 0 | 499661 | 3660 | no | 499531 |
| 4333 | 4333 | 1 | 0 | 499663 | 5838 | 5838 | 499535 |
| 4663 | 4663 | 3 | 1 | 499665 | 4662 | 4662 | 499533 |
| 6066 | 6066 | 3 | 1 | 499667 | 6065 | 6065 | 499537 |
| 6076 | 6076 | 3 | 0 | 499669 | 6075 | 6075 | 499539 |
| 6335 | 6335 | 1 | 0 | 499671 | 6334 | 6334 | 499541 |
| 6501 | 6501 | 2 | 1 | 499673 | 6500 | 6500 | 499543 |
| 7549 | 7549 | 3 | 1 | 499675 | 7548 | 7548 | 499545 |
| 7748 | 7748 | 1 | 0 | 499677 | 7747 | 7747 | 499547 |
| 8010 | 8010 | 3 | 0 | 499679 | 8009 | no | 499549 |
| 8115 | no | no | no | 499681 | 8114 | 8114 | 499551 |
| 8126 | 8126 | 3 | 1 | 499683 | 12434 | 12434 | 499603 |
| 8180 | 8180 | 2 | 0 | 499685 | 8179 | 8179 | 499553 |
| 8233 | no | 3 | no | 499687 | 8232 | no | 499555 |
| 8254 | 8254 | 2 | 0 | 499689 | no | no | no |
| 8332 | 8332 | 2 | 1 | 499691 | 8331 | 8331 | 499557 |
| 8406 | 8406 | 3 | 0 | 499693 | 8405 | 8405 | 499559 |
| 8480 | 8480 | 2 | 0 | 499695 | 8479 | 8479 | 499561 |
| 8549 | 8549 | 1 | 0 | 499697 | 8548 | 8548 | 499563 |
| 9513 | 9513 | 1 | 0 | 499699 | 9512 | 9512 | 499565 |
| 9520 | 9520 | 3 | 1 | 499701 | 9519 | 9519 | 499567 |
| 9525 | 9525 | 2 | 0 | 499703 | 9524 | no | 499569 |
| 9769 | 9769 | 2 | 1 | 499705 | 9768 | no | 499571 |
| 9791 | 9791 | 3 | 1 | 499707 | 9790 | 9790 | 499573 |
| 10249 | 10249 | 3 | 1 | 499709 | 10248 | 10248 | 499575 |
| 10669 | 10669 | 2 | 0 | 499711 | 10668 | 10668 | 499577 |
| 10693 | 10693 | 2 | 1 | 499713 | 10692 | no | 499579 |
| 10741 | 10741 | 3 | 1 | 499715 | 10740 | 10740 | 499581 |
| 10745 | 10745 | 3 | 1 | 499717 | 10744 | 10744 | 499583 |
| 10863 | 10863 | no | 0 | 499719 | 10864 | 10864 | 499585 |
| 11262 | 11262 | 3 | 0 | 499721 | 11246 | 11246 | 499587 |
| 11292 | 11292 | 3 | 1 | 499723 | 11291 | 11291 | 499589 |
| 11473 | 11473 | 1 | 0 | 499725 | 11472 | 11472 | 499591 |
| 11855 | no | 2 | no | 499727 | 11854 | no | 499595 |
| 12025 | 12025 | 1 | 0 | 499729 | 11541 | 11541 | 499593 |
| 12163 | 12163 | 2 | 0 | 499731 | 12162 | no | 499597 |
| 12286 | 12286 | 1 | 0 | 499733 | 12285 | 12285 | 499599 |
| 12380 | 12380 | 3 | 1 | 499735 | 12379 | 12379 | 499601 |
| 12645 | 12645 | 2 | 0 | 499737 | 12644 | 12644 | 499605 |
| 13164 | no | 3 | no | 499739 | 13163 | 13163 | 499607 |
| 13216 | 13216 | no | 0 | 499741 | NONE | no | no |
| 13528 | 13528 | 2 | 1 | 499743 | 13527 | no | 499609 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 13708 | 13708 | no | 1 | 499745 | 13707 | 13707 | 499611 |
| 14458 | no | no | 0 | 499747 | 14470 | 14470 | 499613 |
| 15121 | no | 2 | no | 499749 | 15120 | no | 499615 |
| 16880 | 16880 | 3 | 0 | 499751 | 16879 | 16879 | 499617 |
| 18382 | 18382 | 2 | 0 | 499753 | 19885 | 19885 | 499623 |
| 18477 | no | 1 | no | 499755 | 18476 | no | 499619 |
| 19270 | 19270 | 1 | 0 | 499757 | 19271 | no | 499621 |
| 20018 | 20018 | 1 | 1 | 499759 | 20017 | 20017 | 499625 |
| 20020 | 20020 | 2 | 0 | 499761 | 26470 | 26470 | 499653 |
| 20644 | 20644 | 2 | 0 | 499763 | 20645 | 20645 | 499627 |
| 21061 | 21061 | 1 | 0 | 499765 | 21062 | 21062 | 499629 |
| 21091 | 21091 | 3 | 1 | 499767 | 21090 | 21090 | 499631 |
| 21167 | 21167 | 2 | 0 | 499769 | 21166 | no | 499633 |
| 21274 | 21274 | 3 | 1 | 499771 | 21273 | 21273 | 499635 |
| 21275 | 21275 | 1 | 0 | 499773 | 21276 | 21276 | 499637 |
| 21319 | 21319 | 1 | 0 | 499775 | 21318 | 21318 | 499639 |
| 21882 | 21882 | 3 | 1 | 499777 | 21881 | 21881 | 499641 |
| 21966 | 21966 | 1 | 0 | 499779 | 21965 | 21965 | 499643 |
| 22530 | 22530 | 2 | 0 | 499781 | 22529 | 22529 | 499645 |
| 24028 | 24028 | 1 | 0 | 499783 | 24027 | no | 499647 |
| 24818 | 24818 | 2 | 0 | 499785 | 24817 | no | 499649 |
| 25689 | 25689 | 3 | 1 | 499787 | 25688 | no | 499651 |
| 26666 | 26666 | 1 | 1 | 499789 | 26665 | no | 499655 |
| | | | | | | | |
| | | | 1=positive | | | | |