NEOPLASIA
www.neoplasia.com

Volume 16 Number 5 May 2014 pp. 390–402 390

Pathway-Centric Integrative Analysis Identifies RRM2 as a Prognostic Marker in Breast Cancer Associated with Poor Survival and Tamoxifen Resistance[1,2,3]

Nagireddy Putluri*,†,‡,4, Suman Maity*,†,‡,4, Ramakrishna Kommangani*,4, Chad J. Creighton§,¶, Vasanta Putluri*,†,‡, Fengju Chen§,¶, Sarmishta Nanda#, Salil Kumar Bhowmik*,†,‡, Atsushi Terunuma, Tiffany Dorsey**, Agostina Nardone#, Xiaoyong Fu#, Chad Shaw§,¶, Tapasree Roy Sarkar††, Rachel Schiff*,¶,#,§, John P. Lydon*, Bert W. O'Malley*,‡,¶, Stefan Ambs**, Gokul M. Das‡‡, George Michailidis§§ and Arun Sreekumar*,†,‡,¶**

*Department of Molecular and Cell Biology, Baylor College of Medicine, Houston, TX, USA; †Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, TX, USA; ‡Alkek Center for Molecular Discovery, Baylor College of Medicine, Houston, TX, USA; §Department of Medicine, Baylor College of Medicine, Houston, TX, USA; ¶Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA; #Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX, USA; **Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; ††Department of Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ‡‡Department of Pharmacology and Therapeutics, Roswell Park Cancer Institute, Buffalo, NY, USA; §§Department of Statistics, University of Michigan, Ann Arbor, MI, USA

Abstract

Breast cancer (BCa) molecular subtypes include luminal A, luminal B, normal-like, HER-2–enriched, and basal-like tumors, among which luminal B and basal-like cancers are highly aggressive. Biochemical pathways associated with patient survival or treatment response in these more aggressive subtypes are not well understood. With the limited availability of pathologically verified clinical specimens, cell line models are routinely used for pathway-centric studies. We measured the metabolome of luminal and basal-like BCa cell lines using mass spectrometry, linked metabolites to biochemical pathways using Gene Set Analysis, and developed a novel rank-based method

Address all correspondence to: Arun Sreekumar, PhD, Department of Molecular and Cell Biology, Verna and Marrs McLean Department of Biochemistry, and Alkek Center for Molecular Discovery, Baylor College of Medicine, Houston, TX 77030 or George Michailidis, PhD, Department of Statistics, University of Michigan, Ann Arbor, MI-48109. E-mails: gmichail@umich.edu, arun.sreekumar@bcm.edu

[1] This research was supported by the following grant support: Susan Komen Foundation (to A.S., N.P., R.S., and B.W.O.), PG 1221410 (to R.S.), National Institute of Health (NIH) U01 CA167234 (to A.S., N.P., and G.M.), NSF DMS-1161759 (to A.S. and G.M.), National Science Foundation (NSF) DMS-12-28164 (to G.M.) and National Science Foundation (NSF) DMS-11617838 (to G.M.), NCI P30 CA125123 (to C.J.C. and F.C.), RP120092 (to A.S. and N.P.), National Institute of Health (NIH) HD-07857 (to B.W.O.), RP120092 from Cancer Prevention Research Institute of Texas (to A.S. and N.P.), funds from the Alkek Center for Molecular Discovery (to A.S.), Breast Cancer Research Foundation (to R.S.), Dan L Duncan Cancer Center, NCI-P30CA016056-32 for Pathology Resource Network (PRN), and PG 1221410 (to R.S and B.W.O), Roswell Park Cancer Institute.
[2] Conflict of interest: The authors declare no conflict of interest.
[3] This article refers to supplementary materials, which are designated by Tables W1 to W9 and Figures W1 to W7 and are available online at www.neoplasia.com.
[4] Authors contributed equally to this work.
Received 10 February 2014; Revised 15 May 2014; Accepted 19 May 2014

1476-5586/14
http://dx.doi.org/10.1016/j.neo.2014.05.007

to select pathways on the basis of their enrichment in patient-derived omics data sets and prognostic relevance. Key mediators of the pathway were then characterized for their role in disease progression. Pyrimidine metabolism was altered in luminal *versus* basal BCa, whereas the combined expression of its associated genes or expression of one key gene, *ribonucleotide reductase subunit M2* (*RRM2*) alone, associated significantly with decreased survival across all BCa subtypes, as well as in luminal patients resistant to tamoxifen. Increased RRM2 expression in tamoxifen-resistant patients was verified using tissue microarrays, whereas the metabolic products of RRM2 were higher in tamoxifen-resistant cells and in xenograft tumors. Both genetic and pharmacological inhibition of this key enzyme in tamoxifen-resistant cells significantly decreased proliferation, reduced expression of cell cycle genes, and sensitized the cells to tamoxifen treatment. Our study suggests for evaluating RRM2-associated metabolites as noninvasive markers for tamoxifen resistance and its pharmacological inhibition as a novel approach to overcome tamoxifen resistance in BCa.

Neoplasia (2014) 16, 390–402

Introduction

In the United States, breast cancer (BCa) is the most common cancer diagnosed in women and the second highest cause of cancer-related deaths among them [1]. Once diagnosed, many important characteristics of BCa are used to determine optimal treatment and prognosis. These characteristics include tumor size, estrogen and progesterone receptor status (ER and PR), HER-2/*neu* status, histologic subtype, nuclear grade, lymph node status, and margin status [2,3], all of which provide limited insight into the molecular pathways driving disease progression. Breast tumors are clinically stratified into subgroups on the basis of ER and HER-2 expression and the so-called triple-negative tumors (TN: ER, PR, and HER-2 negative) for which currently there is no targeted therapy. Hence, TN subtype tumors are often treated using conventional chemotherapeutics [4,5].

To obtain a better understanding of the pathways associated with estrogen-induced molecular alterations, numerous studies have examined gene and protein expression profiles using high-throughput omics-based technologies [6–18]. However, the application of metabolomics to define pathways associated with BCa has been limited. Unlike the genome and the proteome, the metabolome defines the actual physiological state of the tumor, is computationally tractable, less complex (than the other -omics), and more importantly, reveals potential metabolites that can be measured in noninvasive body fluids in a clinical context. Some researchers have used mass spectrometry to examine the metabolome associated with BCa [19–21] as well as to determine altered metabolites and biochemical pathways associated with the various subtypes of tumors [22–28].

In the current study, using a robust mass spectrometry platform [29–32], we measured metabolic alterations in luminal and basal BCa cell lines [33] and ranked pathways using a Gene Set Analysis (GSA)–based enrichment approach [34]. The enriched pathways were then selected on the basis of their relevance in patient-derived luminal and basal-like BCa tissues, by examining pre-existing gene and metabolic expression data sets. Following this, the selected pathways were further stratified on the basis of their association with survival of patients with BCa using publicly available gene expression data sets containing information on patient outcome. A novel rank-based method was developed that took into account the degree of enrichment of the pathways in each of the molecular data sets, as well as its prognostic potential, to generate a cumulative rank score. This was then finally used to stratify the pathways for subsequent downstream validation studies. This systematic stepwise selection enabled us to identify pyrimidine metabolism as a key biochemical pathway associated with aggressive BCa in general and with tamoxifen resistance in patients with luminal BCa. Importantly, using *in vitro* and *in vivo* BCa models, the translational and clinical relevance of pyrimidine metabolism and the gene associated with one of its key enzymes, ribonucleotide reductase subunit M2 (RRM2) was established.

Methods

Cell Lines

Breast cell lines (basal-like or mesenchymal breast cancer—BT549, HS578, MDA MB 231, MDA MB 436, and MDA MB 468; luminal breast cancer—BT474, MCF-7, MDA MB 453, and T47D) were purchased from American Type Culture Collection (Manassas, VA; see Supplementary Table 1 for description of the cell lines). SUM159PT basal BCa cells were kindly gifted by Dr Ethier (Medical University of South Carolina (MUSC) Hollings Cancer Center, Charleston, SC). MDA MB 231, MDA MB 453, HS578T, and MCF-7 L [35,36] were grown in Dulbecco's modified Eagle's solutin (DMEM)–GlutaMAX media (Invitrogen Corp, Carlsbad, CA) supplemented with 10% FBS (Hyclone Laboratories/Thermo Scientific, Rockford, IL) and 1% penicillin-streptomycin (Hyclone Laboratories). MDA MB 436, and MDA MB 468 were grown in L15 media (Life Technologies, Grand Island, NY) supplemented with 10% FBS (Hyclone Laboratories). T47D, BT 474, and BT549 cells were grown in RPMI (Invitrogen Corp) media supplemented with 10% FBS (Hyclone Laboratories) and 1% penicillin-streptomycin (Hyclone Laboratories). SUM 159 PT was grown in Ham F12, 5% insulin hydrocortisone (Life Technologies). All cells were maintained at 37°C and 5% CO_2. Before their analyses, cells were trypsinized, and the pellet was washed thrice with ice-cold phosphate-buffered saline (PBS), counted into 25 million aliquots, and stored at −140°C.

For studies to characterize the role of RRM2 in tamoxifen resistance, MCF-7 parental without treatment (MCF-7 L, parental) or either treated with tamoxifen for 48 hours [when cells are still sensitive, TAM sensitive (TAM-S)] or for long term until cells became resistant and resumed growth [TAM resistant, TAM-R)] as described earlier by Morrison et al. [35,36] were used. The parental

cells were grown in RPMI medium as described above. Tamoxifen-treated cells (TAM-S or TAM-R) were grown in phenol-red–free RPMI medium (Media Tech, Manassas, VA) containing charcoal stripped Fetal Bovine Serum, with 10 uM 4-hydroxytamoxifen (Sigma-Aldrich, St Louis, MO). In addition, xenografts generated using MCF parental cells (MCF-7 L) that were either untreated (termed parental) or treated with tamoxifen for ~2 weeks (short-term treatment and TAM-S) or ~3 months (long-term treatment and TAM-R), as described by Massarweh et al. were used.

Mass Spectrometry

Unbiased mass spectrometry–based methods were described earlier [30]. Sample preparation for mass spectrometry–based examination of metabolome is described in Supplementary Methods. The mass spectrometry portion of the unbiased profiling platform is based on a 1200 SL Rapid resolution LC and a 6520 Quadrupole Time of Flight mass spectrometer (Agilent Technologies, Santa Clara, CA). For unbiased profiling studies, real-time mass correction during mass spectrometry was achieved by infusion of a standard mixture of reference ions using an independent 1200 SL Rapid resolution LC isocratic pump equipped with 100:1 splitter to output a flow rate of 5 ml/min. The samples were independently examined in both positive and negative ionization modes using a dual electrospray ionization source. Detailed description of chromatographic methods used for separation of metabolites is given in Supplementary Methods.

The mass spectrometry portion of the targeted profiling platform was based on a 1200 SL Rapid resolution LC and a 6430 triple Quadrupole mass spectrometer (Agilent Technologies). The multiple reaction monitoring (MRM)–based measurement of levels of 76 metabolites was done using four different methods, which used either a reverse-phase or normal aqueous-phase chromatographic separation, before mass spectrometry. Details of the methods, associated chromatographic conditions, metabolites measured, and their corresponding MRM transitions are given in Supplementary Table 2 and Supplementary Methods.

Metabolomic Library, METLIN

METLIN (Agilent Technologies) was used to search the unbiased mass spectral data. METLIN was created using approximately 1800 commercially available compounds whose retention time was defined using the Reverse Phase (RP) chromatographic method described above.

Metabolomics Data Analysis

All the downstream processing and data analyses were performed using R statistical programming software (R Foundation for Statistical Computing, Vienna, Austria) [37]. After removing the compounds with more than 40% of missing values, the missing values in the remaining metabolites in the BCa cell lines were imputed using the K-Nearest Neighbour (KNN) algorithm ("imputation" package [38], $K = 5$). Following the preprocessing, of 673 compounds, the BCa cell line has 76 were unique metabolites, whereas our recently published BCa tissue metabolome data set [39] had a total of 219 compounds of which 168 metabolites were named.

Imputed data were median centered and Inter Quartile Range (IQR) scaled following $\log_2$ transformation. Two-sided t tests were performed to identify differential metabolites by comparing luminal and basal subtypes coupled with False discovery rate (FDR) adjustment (adjusted P values < .2) using the Benjamini Hochberg (BH) method [40] along with estimated fold change using the Differential Expression via Distance Summary (DEDS) package [41].

Analysis of Microarray Gene Expression Data

In this study, we used this gene expression data set (termed National Cancer Institute data set, GSE37751) for 46 matched BCa tissues derived from (Affymetrix, Santa Clara, CA, USA) GeneChip Human Gene 1.0 ST arrays, followed by Robust Mult-chip Average (RMA) normalization [39,42] found in the R bioconductor [43] package and differential expressions analysis using FDR-adjusted (Benjamini Hochberg) two-sided t tests (limma (linear models for microarray data) package). Heat maps with average linkage based hierarchical clustering of z-score transformed differentially expressed compounds and genes, were generated using the gplots [44] package.

Identification of Key Pathways Based on Enrichment Scores Describing Differential Expression and a Prognostic Value in BCa Subtypes

Selection of key pathways was accomplished using a two-step process that considered both the relative enrichment scores on the basis of a modified enrichment analysis, GSA (proposed by Tibshirani et al. {(http://www-stat.stanford.edu/~tibs/GSA/), [45,46]} and their prognostic potential on both the set of metabolites and their Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) (KEGG)–derived associated gene sets across multiple publically available BCa gene expression data sets (refer to Supplementary Table 3 for the list of metabolites and their associated genes and Supplementary Table 4 for the description of the data sets).

In the first step, the selection of top significant (enrichment P value) pathways included a group of 11 pathways that were altered between basal and luminal subtypes. Next, the gene sets associated with these 11 enriched pathways were obtained from KEGG to derive their pathway-specific score (averaged gene expression) and examined for their prognostic value in the publically available gene expression data sets (refer to Supplementary Table 4 for the list of data sets used) [47,48]. This was done by examining the association between pathway scores and years of metastasis-free patient survival using a Cox proportional hazards model. For selected pathway-associated gene sets or individual genes, correlation with metastasis-free survival was visualized using Kaplan-Meier (KM) plots. As a comparative reference for the prognostic analysis, samples in the Kessler compendium and van de Vijver data sets were also stratified using (GEO data set GEICAM9906) PAM50 and other known prognostic marker panels [49].

Once the enrichment scores for both expression levels and prognostic values were obtained for each pathway, they were combined to generate a combined rank (CR) score, details of which are described in Supplementary Methods.

Analysis of RRM2 Expression by Immunohistochemistry on Tissue Microarray of Patients with BCa

Tissue microarray construction. RRM2 protein expression was measured using a previously developed tissue microarray (TMA) (Pathology Resource Network, Roswell Park, NY) containing 192 biopsy specimens (each having three replicates). Tissues from patients who had surgeries performed between 1995 and 2008 at Roswell Park Cancer Institute (Buffalo, NY) were included in the TMA. Specimens for controls within the TMA consisted of multiple cores of normal tissue from 10 different organs including heart, colon, kidney,

adrenal, ovary, myometrium, brain, thyroid, lung, and prostate, thereby representing more than 20% of all the cores in a TMA. Appropriate Institutional Review Board approval consistent with federal, state, and local requirements was obtained for this project, and clinical and outcome data were deidentified.

Of the 192 patients whose tissues were arrayed on the TMA, 185 were ER positives, 5 were ER negatives, and the remaining 2 were not defined for their subtype. Further among the 192 tissues, 132 were PR positive, 58 were PR negative, and the remaining two patients were not defined for their PR status. Patients were administered with tamoxifen (156), anastrozole (21), or letrozole (2) or not treated (5). The treatment was given either in an adjuvant systemic setting (165) or neoadjuvant systemic setting (16). Two patients received systemic treatment for metastasis, one received for local recurrence, four did not receive any systemic treatment, one patient refused treatment, and three patients were not documented for the treatment modality. In addition, a subset of patients were also subjected to chemotherapy with the following distribution: AC (Adriamycin+ Cyclophosphamide (AC) (Sigma, St Louis, MO, USA)) (29), CMF IV (Cyclophosphamide+ Methotrexate+ Fluorouracil /5FU (CMF) (Sigma, St Louis, MO, USA)) cytoxan (6), AC/Taxol (48), Adriamycin/Taxotere (4), A-CMF (Adriamycin+Cyclophosphamide+ Methotrexate+ Fluorouracil /5FU(ACMF) (Sigma, St Louis, MO, USA)) (3), Taxotere (1), CEF (Cyclophosphamide + Epirubicin, + 5-Fluorouracil (CEF) (Sigma, St Louis, MO, USA)) (Epirubicin, 1). Seventy-eight patients did not receive any chemotherapy.

Importantly, in this study, the inclusion criteria for evaluating RRM2 expression in the TMA were that the patient should have ER+ tumors, be treated with tamoxifen in an adjuvant setting and without chemotherapy. Using these criteria a total of 45 patients were selected, of which recurrence was seen in 19 patients whereas the rest of the 26 patients were followed for recurrence-free survival for a median time of 8 years. Furthermore, among these 45 patients, 25 patients were reported to have died by the end of 8 years of median follow-up, of which 13 died of the disease, 6 died due to other complications, and 6 died of unknown causes.

Immunohistochemistry

Paraffin sections were cut at 4 μm, placed on charged slides, and dried at 60°C for 1 hour. Slides were cooled to room temperature, deparaffinized in three changes of xylene, and rehydrated using graded alcohols. For antigen retrieval, slides were heated in the steamer for either 40 or 60 minutes in citrate buffer (pH 6.0) (Biocare Medical, (Concord, CA) No. CB910), followed by a 20-minute cooldown. Endogenous peroxidase was quenched with aqueous 3% H_2O_2 for 10 minutes and washed with PBS/T (Tween-20). Slides were loaded on a Dako (X0909, Dako, CA) autostainer, serum-free protein block (Dako No. X0909) was applied for 5 minutes and blown off, and the RRM2 antibody (SC-81850, Santa Cruz Biotechnology (Santa Cruz (Dallas, TX)) was applied at 1:500 dilution for 1 hour. PBS/T was used to wash slides between each reagent application. Dako Mouse Envision (K4007) was applied for 30 minutes, followed by DAB (Dako No. K3468) for 10 minutes. Finally, the slides were removed from the autostainer, counterstained with hematoxylin, dehydrated, cleared, and coverslipped. Isotype-specific nonimmune IgG was used as a control to examine the specificity of RRM2 staining.

Aperio Slide Scanning and Image Analysis

TMA slides were digitally scanned using Aperio ScanScope (Aperio Technologies, Inc, Vista, CA) with ×20 bright-field microscopy. These images were then accessible using Spectrum (Aperio Technologies, Inc), a web-based digital pathology information management system. Slide images were automatically associated to a digital slide created in the Digital Slide table in Spectrum.

Once slides were scanned, Aperio ImageScope version 11.2.0.780 (Aperio Technologies, Inc) was used to view images for analysis. Slide image data fields were populated and images were examined for quality and were amended as necessary. Care was taken to avoid including areas of carcinoma *in situ* and regions with staining artifacts. When possible, representative areas of tumor were selected for analysis with a minimum target of 30 tumor cells per TMA core.

A Cytoplasm Algorithm (Aperio, Leica Microsystems, IL) that was calibrated to analyze DAB staining intensity and calculate the percentage of cells containing the stain within their cytoplasmic compartment was used. Staining thresholds were set for calling out positive stains, and the scores for average cytoplasm intensity for the selected regions were calculated on the basis of these thresholds. The staining intensity was stratified into the following four score values: 0, none; 1 +, weak; 2 +, moderate; and 3 +, strong. Concomitantly, for each of these staining scores, the percentage of cells that stained was also calculated. The staining intensity and the percentage of cells in each of the staining intensity bins were together used to calculate the H score that reflects the proportion of cells in the cytoplasm of the tumor that express RRM2 at the various staining intensity thresholds described above. In other words, the H score was calculated as follows: 1*(%1 +) + 2*(%2 +) + 3*(%3 +), and ranged from 0 to 300 with a score of 300 reflecting 100% of the cells having a staining intensity of 3 +. All the slides were counterstained with hematoxylin to reveal the morphologic detail of the surrounding tissue and to help identify nuclear and cytoplasmic compartments of the cells for analysis.

TMA Data Analysis

In this study, we selected tissues from the TMA for RRM2 analysis that met the following inclusion criteria: ER+ treated with tamoxifen in the adjuvant setting and no chemotherapy. A total of 45 tissues passed these criteria among which 19 recurred and the rest did not show recurrence for a median follow-up time of 8 years. Furthermore, among the 45 patients, 25 were reported to be dead by the end of the follow-up period either due to cancer (13) or other causes (12).

For the analysis, the median H score value for each of the samples was calculated using data from replicate cores. The median H score was then compared using nonparametric rank sum test for RRM2 expression among patients who were reported to be dead *versus* those reported to be alive for the 8-year median follow-up time period. Furthermore, KM plots using H score were used to examine the association of RRM2 with recurrence-free survival.

Genetic and Pharmacological Inhibition of RRM2 in TAM-R Cells and Associated Characterization

TAM-R cells were transfected using either negative control siRNA (ctrl-siRNA) small interfering Ribo Nucleic Acid (siRNA) or Smart pools of siRNA against RRM2 (RRM2-siRNA, Thermo Scientific/Dharmacon RNAi Technologies, Chicago, IL) using Lipofectamine 2000 reagent as described earlier [50]. At 48 hours of posttransfection, whole-cell extracts were made, and knockdown of RRM2 was verified using both Quantitative Polymerase Chain Reaction (QPCR) and immunoblot analysis. Bromodeoxyuridine (BrdU) labeling assay (Calbiochem/Millipore, Billerica, MA, USA) was used to assess the proliferation rate per manufacturer's instructions. For cell-cycle gene

analysis studies, TAM-R cells were transfected with ctrl-siRNA or RRM2-siRNA as described above. After 48 hours, total RNA was isolated from TAM-R cells, and cDNA was prepared from total RNA using iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA). Transcript levels of cyclin genes were measured by performing qPCR using SsoAdvanced universal qPCR supermix and gene specific primers as per manufacturer's instructions (Bio-Rad Laboratories). For 5-azacytidine (aza)–based studies, TAM-R cells were plated onto 96-well and six-well plates for assessing cell survival and RRM2 protein levels, respectively. The next day, cells were treated with control vehicle or aza at 25, 50, and 100 nM concentrations. Following this, after 72 hours, cells were subjected to MTT (Non-Radioactive Cell Proliferation Assay (Promega Corporation, Madison, WI)) assay and immunoblot analysis.

Results

Figure 1*A* shows the overall approach used in this study. Mass spectrometry was used to measure the relative levels of 673 compounds (includes 60 metabolites measured using targeted assays) across 10 BCa cell lines (luminal $n = 4$ and basal $n = 6$; see Supplementary Table 1 for description of cell lines), each analyzed as biologic triplicates. Metabolites from predefined liver pools ($n = 26$) were extracted in parallel and examined as process controls. Internal standards were spiked in during the extraction

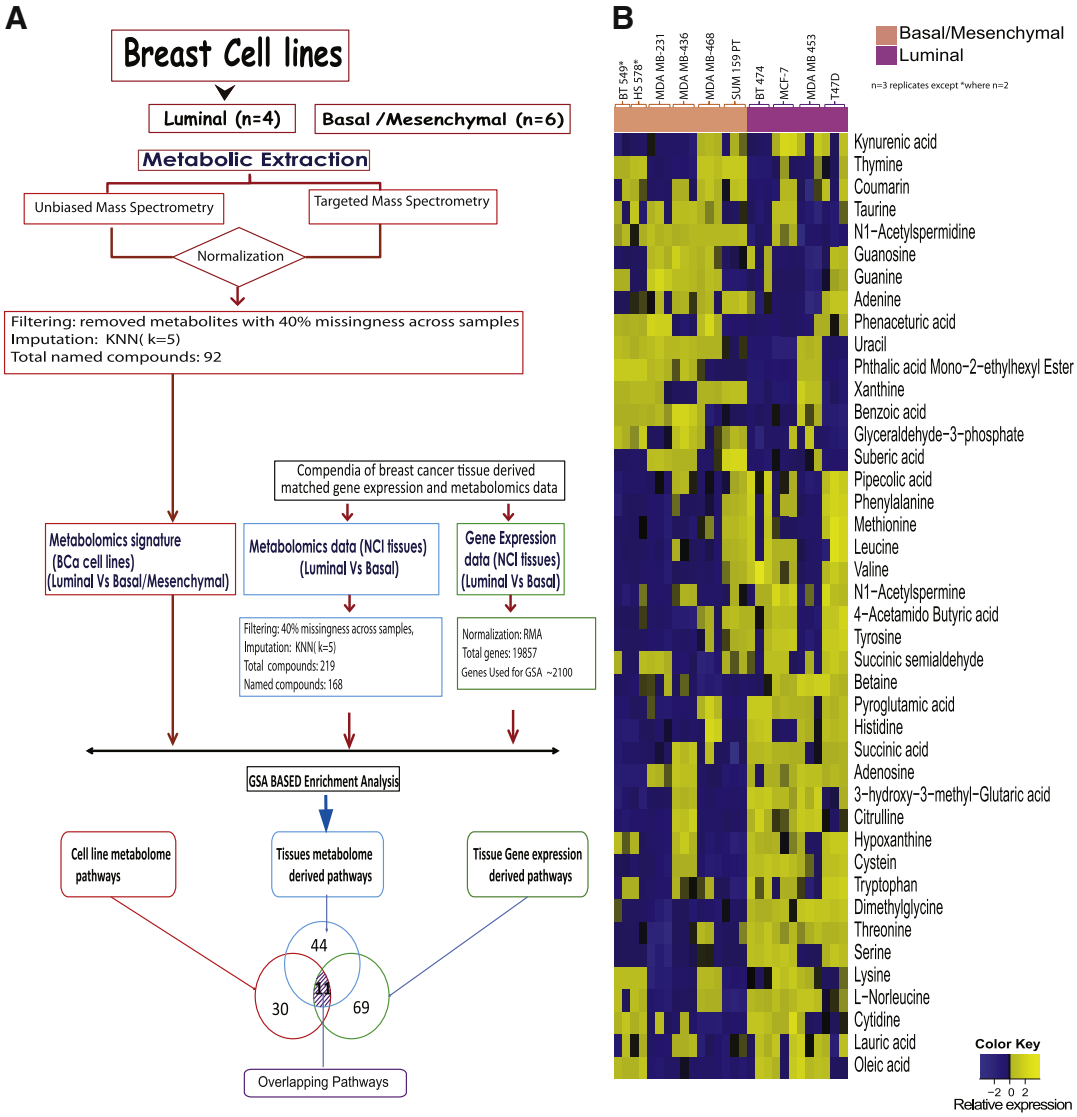


Figure 1. Metabolomic profiling of BCa cell lines. (A) Overview of metabolomic profiling and the integrative methodology. Mass spectrometry–based measurement of the cell line metabolome was followed by GSA-based identification of enriched biochemical pathways. These were compared with GSA-enriched metabolic pathways derived from an independent set that consisted of matched metabolome and transcriptome data derived from ~50 BCa tissues (46 specimens for transcriptome data). Pathway enrichment results from each of the three data sets were integrated using a novel rank-based method to nominate a set of 11 commonly enriched pathways. These were further curated using a similar rank-based approach for their clinical significance, i.e., association with time to metastasis-free survival. Overall, results of this stepwise enrichment were used to nominate pathways on the basis of both their biologic and clinical relevance. (B) Heat map overview of levels of named metabolites across luminal and basal/mesenchymal BCa cell lines. Shades of maize and blue represent elevated and reduced levels of metabolites, respectively (see color scale).

process to assess process variation. Mass spectral data were used to calculate differential metabolites that distinguish basal and luminal BCa cell lines. To obtain a pathway perspective, the metabolite data were enriched using GSA [34]. In parallel, the same strategy was used on matched pre-existing BCa tissue–derived gene expression and metabolic data to determine clinically relevant biochemical pathways. Subsequently, commonly enriched pathways in both the cell line and patient data sets were determined and evaluated for their association with BCa-specific survival. A novel rank-based scoring method incorporating information on both the biologic (enrichment in clinical tissue data sets) and clinical (prognostic) significance of the pathways was developed and employed to select key biochemical pathways for downstream validation studies. This method allows us to define key biochemical pathways associated with cancer progression, by integrating data from diverse omics data sets containing well- annotated clinical data.

Metabolic Profiles Associated with BCa Cell Lines

Before the analysis of the cell line data, the process controls, i.e., matrix-free internal standards and liver pools were evaluated for their variability. The range of coefficient of variation (% CV) for the eight internal standards in liver pool was 0.09% to 6%. A total of 76 named metabolites were measured across the luminal and basal BCa cell lines (Supplementary Figure 1 and Supplementary Table 1). Of these, 42 metabolites (FDR-adjusted P value <.2) were differentially altered between the luminal and basal subtypes (Figure 1*B*). Luminal BCa cell lines showed elevated levels of amino acids, like phenylalanine, tryptophan, and tyrosine, and branched chain amino acids like leucine, lysine, and valine, as well as higher levels of lauric acid and oleic acid. In contrast, levels of nucleotides like guanine, adenine, thymine, uracil, xanthine, and guanosine were elevated in basal BCa cells compared to the luminal counterparts.

Enrichment Analysis to Define Pathways Associated with Cell Line Metabolomics Data

To obtain better insight into the subtype-specific biochemical pathways distinguishing luminal from basal BCa cell lines, we applied GSA on the cell line–derived metabolomics data as described under the Methods section. As a first step in this process, all the named metabolites identified in cell lines (n = 53) were mapped to 200 biochemical pathways using the KEGG database (Supplementary Table 2) [51]. These pathways were then used for GSA-based enrichment analysis. Following this, we ranked the pathways on the basis of their FDR-corrected enrichment P value and selected 30 pathways (P value < .15; Supplementary Table 5) for subsequent integrative analysis with pathways obtained from tissue data sets. The enriched pathways in the cell line data include those that describe metabolism of purine, pyrimidine, glutamate, alanine-aspartate, glycine-serine-threonine, butanoate, β-alanine, taurine-hypotaurine, valine-leucine-isoleucine biosynthesis, pantothenate-CoA (Coenzyme A) biosynthesis and lysine degradation.

Enrichment of Biochemical Pathways in Luminal Basal-Like BCa Tissues

We used matched transcriptomic and metabolomic data sets for human breast tumors that were generated earlier by our group [39]. Metabolomics data were available for a total of 50 matched benign-tumor pairs that included 33 luminal and 17 basal-like/TN tumors. Here, a total of 219 compounds were measured, of which 98 metabolites were mapped to ~200 biochemical pathways in KEGG. Supplementary Figure 3*A* shows the overall heat map for the metabolites in luminal and basal-like BCa tissues. A total of 44 pathways were enriched (P value < .15) using GSA on the tissue metabolic data set that were subsequently ranked on the basis of their P value and used for integrative analysis with cell line–derived biochemical pathways (Supplementary Table 5 lists the enriched pathways).

The transcriptomic data, published by our group earlier [39], contained measurements for ~20,000 genes that were measured using Affymetrix ST arrays across 46 BCa tissues (ER+ n = 31 and basal like/TN, n = 15; see Supplementary Figure 3*B*, GEO Accession No. GSE39004/GSE37751). Among these, ~2100 genes were mapped to the same 200 KEGG biochemical pathways described above and used for GSA-based enrichment analysis. Supplementary Table 5 shows the list of the 69 pathways enriched using the gene expression data (P value < .15) that distinguish luminal *versus* basal BCa. These pathways were also ranked on the basis of their FDR-corrected enrichment P value and used for subsequent integrative analysis.

Selection of Robust Omics-Derived Pathways

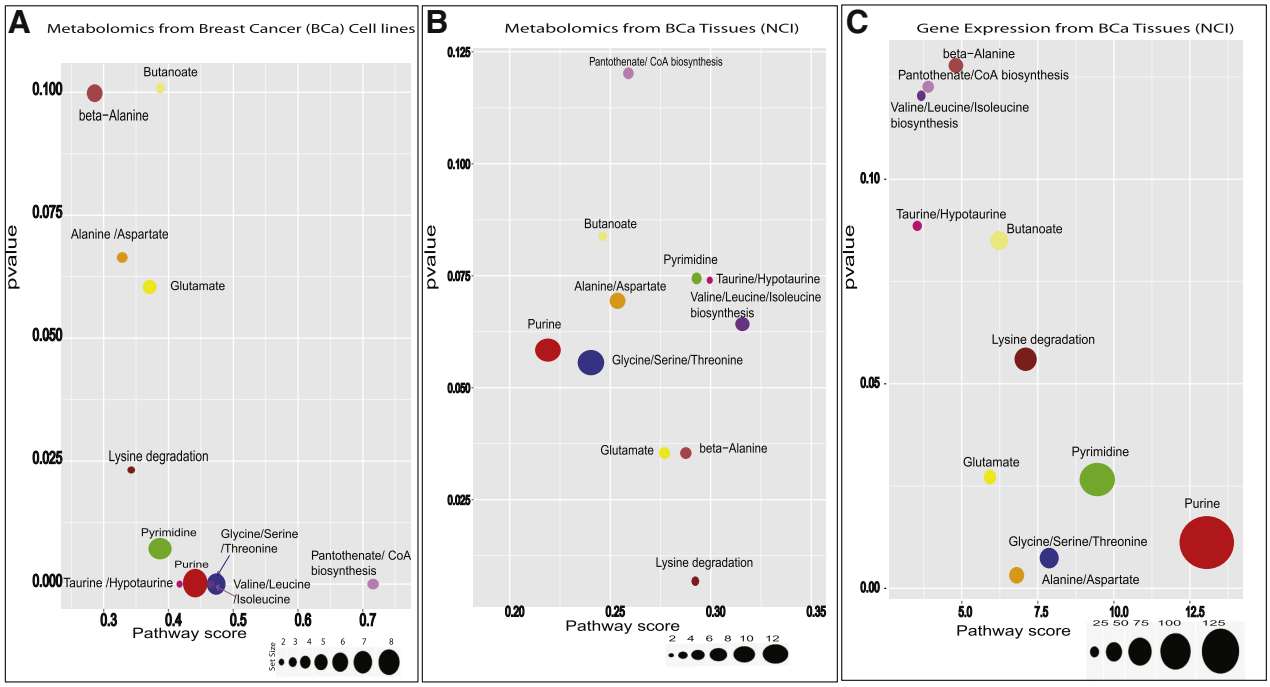
Having determined the enriched biochemical pathways and their respective ranks in the metabolome (cell line and tissue) and transcriptome (tissue) data sets, we next asked which of these metabolic processes were consistently enriched across all the three data sets. A total of 11 pathways distinguishing luminal *versus* basal BCa were uniformly enriched at a P value threshold <.15 in each of the three data sets (Figure 2, *A*–*C*). These 11 pathways included those that describe metabolism of purine, pyrimidine, glutamate, alanine-aspartate, glycine-serine-threonine, butanoate, β-alanine, taurine-hypotaurine, valine-leucine-isoleucine biosynthesis, pantothenate-CoA biosynthesis, and lysine degradation (Supplementary Figure 2). Supplementary Figure 3 shows the heat maps for the expression of the genes associated with the 11 biochemical pathways in luminal and basal breast tumors from our BCa tissue–derived transcriptomic data set [39].

Integration of Omics-Based Enrichment and Prognostic Ranks to Select Clinically Relevant Biochemical Pathways

The biochemical pathways that were consistently enriched across the different omics data sets were further filtered for their clinical relevance by examining their prognostic potential in publicly available BCa gene expression data sets having at least 10 years of clinical follow-up information. An *in silico* analysis was performed to investigate the association of each of the 11 metabolic pathways with patient survival using 10 independent gene expression data sets (refer to Supplementary Table 4). Among these, eight data sets (all on Affymetrix platform) are included in the Kessler compendium [48]. The additional two data sets include the van de Vijver data set (containing gene expression data on 295 patients with BCa [47] and the NCI transcriptome data set (GEO data set GSE39004) (used for integrative analysis above [39]. For the analysis, both the expression value and directionality of expression for each gene associated with the 11 commonly enriched pathways (described in Figure 2, *A*–*C*) were obtained from each of the above gene expression data sets. This information was then used to generate an average gene expression score for each pathway, which was then examined for association with years of metastasis-free patient survival using a Cox proportional

hazards model. The prognostic value of the top ranked pathway was visualized using KM plots across all samples in the Kessler compendium. Figure 2*D* shows the individual as well as integrated *P* values for each pathway obtained across all BCa tissues using all the 10 data sets described above.

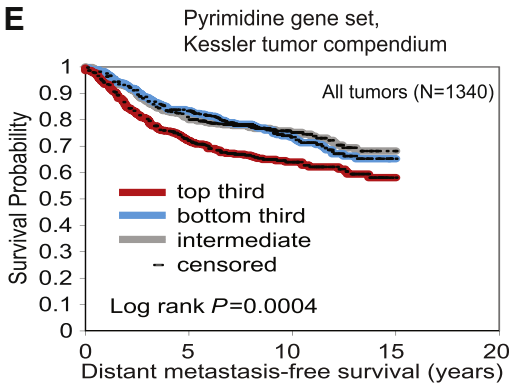
Notably, only pyrimidine metabolism had a *P* value <.1 for the CR Score, indicative of its biologic and clinical relevance in BCa (Table 1). In addition, pyrimidine metabolism was also significantly enriched in a subset of BCas with documented ER status (Supplementary Table 6). Consistent with all of this, KM plots also showed higher expression of pyrimidine metabolism pathway genes as being associated with shorter metastasis-free survival, across all BCa (Figure 2*E*, *N* = 1340, log-rank *P* value = .0004) as well as within the subset of ER+ tumors (Figure 2*F*, *N* = 686, log-rank *P* value = .003). In addition, higher expression of pyrimidine metabolism genes, correlated well with PAM50 [52], defined basal-like, Her-2–enriched, and luminal B tumor subtypes whereas being lower in luminal A tumors (Figure 2*G*). Higher cumulative expression of pyrimidine

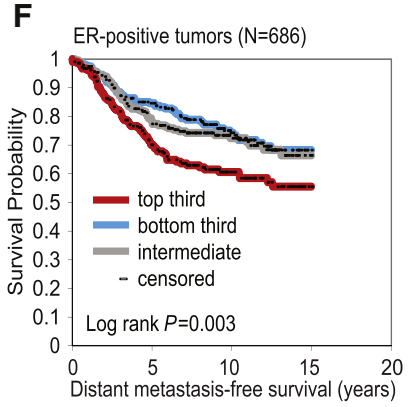


D Univariate Cox p-values, correlation with distant met-free survival

Pathway	Purine	Pyrimidine	Glutamate	Alanine and aspartate	Glycine, serine and threonine	Valine, leucine and isoleucine	Lysine degradation	beta-Alanine	Taurine and hypotaurine	Butanoate	Pantothenate and CoA
Chin	0.69	0.8867	0.352018	0.39	0.25636879	0.55	0.76493569	0.03179	0.51319	0.011	0.79977845
Desmedt	0.59	0.3912	0.337026	0.69	0.91571858	0.69	0.61169976	0.00043	0.60714	0.014	0.25917953
Loi	0.19	0.0032	0.256111	0.25	0.5335992	0.13	0.14728391	0.21924	0.13856	0.049	0.36948496
Minn (2005)	0.16	0.6214	0.395419	0.85	0.63029842	0.65	0.601937	0.69602	0.5168	0.698	0.52173749
Minn (2007)	0.25	0.0353	0.697554	0.02	0.27845509	0.37	0.3718909	0.48714	0.81765	0.641	0.94340073
Schmidt	0.05	0.0089	0.050765	0.72	0.75367514	0.06	0.59486639	0.21126	0.16844	0.815	0.22613129
Wang	0.28	0.0957	0.925174	0.47	0.01225123	0.07	0.00226724	0.03241	0.00709	0.153	0.82102774
Zhang	0.77	0.3518	0.352596	0.65	0.02941004	0.84	0.02911779	0.03179	0.51319	0.101	0.79977845
NCI.	0.461959	0.1726	0.043639	0.040559	0.8148938	0.019055	0.495968	0.26128	0.01985	0.277	0.6542126
van de Vijver	0.025749	0.001	0.516883	0.07	0.008	0.47	0.09	0.09851	0.43711	0.496	0.68546085

Red denote significant correlation with worse outcome.
Blue denote significant correlation with better outcome.





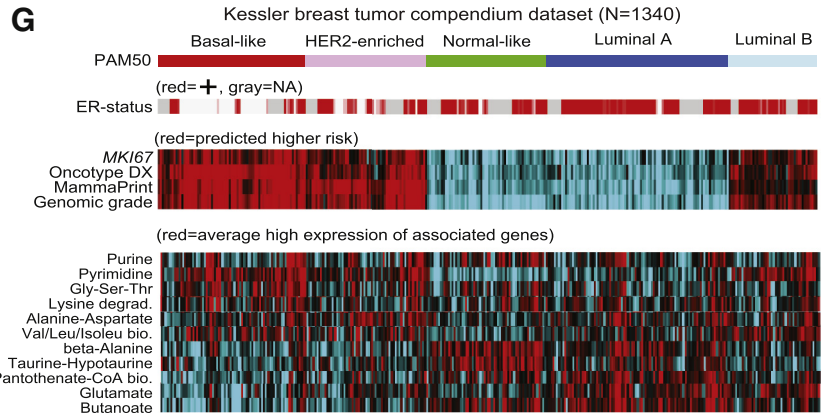


Table 1. Individual and Combined *P* Values for Biochemical Pathways Nominated by Integrative Analysis in BCa Data Sets

						Kessler Compendium								
p values → Pathway ↓	NCI. Tranc riptome	NCI. Meta bolome	BCL. Meta bolome	NCI. Survival ER+(31); ER-(15)	van de Vijver ER+(226); ER-(69)	Chin ER+(75); ER-(43)	Desmedt ER+(134); ER-(64)	Loi ER+(211); ER-(24)	Minn (2005) ER+(57); ER-(42)	Minn (2007) ER+(0); ER-(58)	Schmidt ER+(0); ER-(0)	Wang ER+(209); ER-(77)	Zhang ER+(0); ER-(0)	Combined Rank Score (CRS)
Purine	0.01	0.06	0.0002	0.46	0.03	0.69	0.59	0.19	0.16	0.25	0.05	0.28	0.77	0.27
Pyrimidine	0.03	0.07	0.007	0.17	0.001	0.89	0.39	0.003	0.62	0.04	0.009	0.10	0.35	0.07
Glutamate	0.03	0.04	0.06	0.04	0.52	0.35	0.34	0.26	0.40	0.70	0.05	0.93	0.35	0.59
Alanine and aspartate	0.003	0.07	0.07	0.04	0.07	0.39	0.69	0.25	0.85	0.02	0.72	0.47	0.65	0.59
Glycineserine and threonine	0.007	0.06	0	0.81	0.008	0.26	0.92	0.53	0.63	0.28	0.75	0.01	0.03	0.29
Valine leucine and isoleucine	0.12	0.06	0	0.02	0.47	0.55	0.69	0.13	0.65	0.37	0.06	0.07	0.84	0.58
Lysine degradation	0.06	0.007	0.02	0.50	0.09	0.76	0.61	0.15	0.60	0.37	0.59	0.002	0.03	0.23
beta-Alanine	0.13	0.04	0.10	0.26	0.10	0.03	0.0004	0.22	0.70	0.49	0.21	0.03	0.03	0.41
Taurine and hypotaurine	0.09	0.07	0	0.02	0.44	0.51	0.61	0.14	0.52	0.82	0.17	0.007	0.51	0.45
Butanoate	0.09	0.08	0.10	0.28	0.50	0.01	0.01	0.05	0.70	0.64	0.81	0.15	0.10	0.72
Pantothenate and CoA	0.12	0.12	0	0.65	0.69	0.80	0.26	0.37	0.52	0.94	0.23	0.82	0.80	0.99

Red denote significant correlation with worse outcome.
Blue denote significant correlation with better outcome.

metabolism genes also overlapped with expression patterns of MKi67 [53], Oncotype Dx [54], Mammaprint [55], and genome grade [56] (Figure 2*G*). In contrast to the above, genes belonging to the methionine and glutathione pathways showed opposing expression patterns in basal-like and Her-2–enriched tumors (Figure 2*G*).

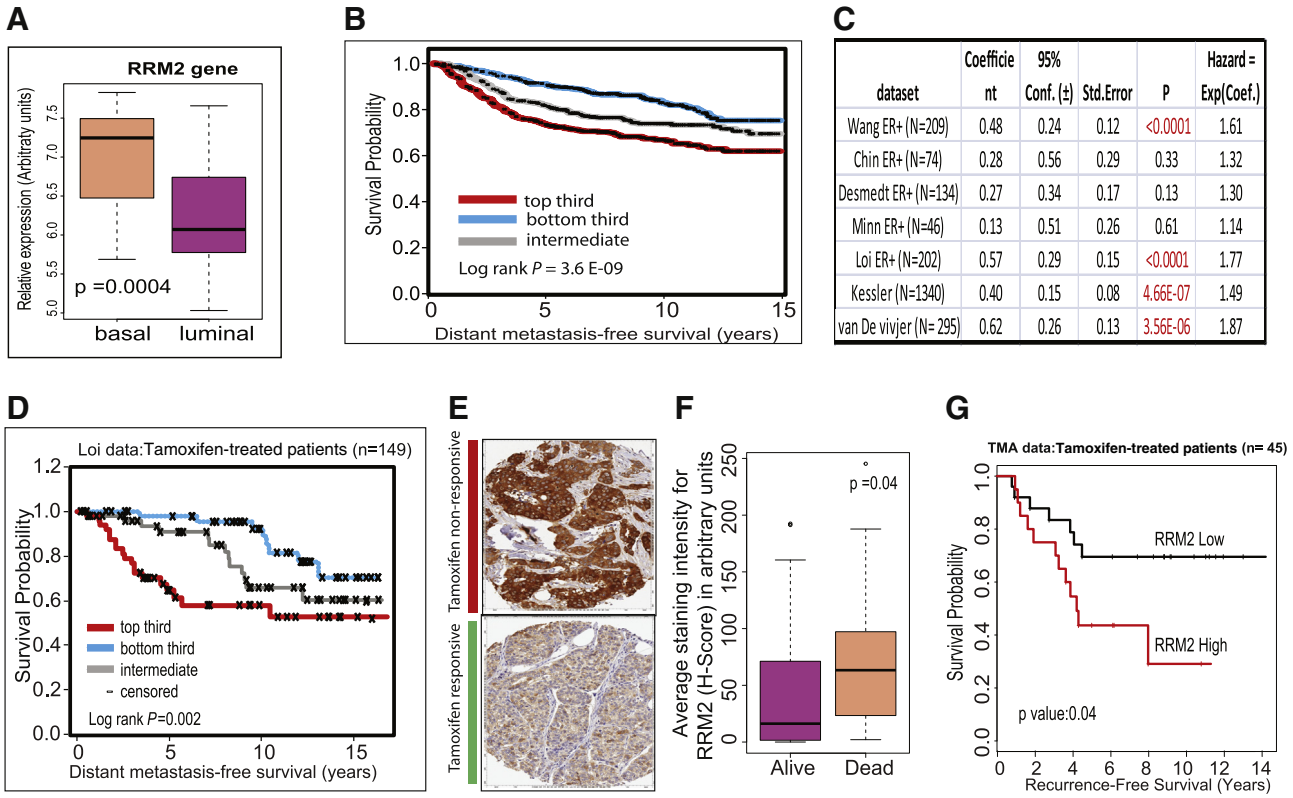
Higher Expression of RRM2 is Associated with Tamoxifen Resistance in BCa

We examined the genes within the pyrimidine metabolic pathway that were most highly altered in basal *versus* luminal BCa using our tissue-derived transcriptome data [39] Both uridine pyrophosphorylase 1 (UPP1, Supplementary Figure 4) and RRM2 (Figure 3*A*) were significantly ($P = .0004$) elevated in basal-like BCa compared to luminal tumors. Ribonucleotide reductase has three isoforms, M1 (RRM1), M2 (RRM2), and M2b (RRM2b), all of which are key enzymes in the process of DNA replication and catalyzes the conversion of UDP and cytidine diphosphate (CDP) to their respective deoxygenated versions (dUDP/dCDP), which are then used for DNA synthesis. Furthermore, higher levels of RRM2 but not RRM1 (refer to Supplementary Table 7) were associated with poor survival in the Kessler compendium (log-rank $P = 3.6E-09$; Figure 3*B*) as well as the Wong, Loi, and van de Vijver data sets, each of which contains a significant number of luminal A patients (Figure 3*C*) [47,57] who are conventionally treated with endocrine or hormone therapy. In light of the above finding, we asked whether RRM2 could distinguish luminal patients on the basis of their response to tamoxifen (TAM), a widely used endocrine agent for ER-positive BCa. Our analysis using the Loi data set (that contains 149 ER+ node-negative patients treated with TAM), showed that RRM2 by itself was able to distinguish patients who exhibit intrinsic resistance to TAM treatment and hence relapse within 2 to 5 years after starting the therapy (log-rank P value = .002; Figure 3*D*).

Consistent with this, a KM plot (Figure 3*D*) showed that about 70% of the patients who had low expression of RRM2 (bottom third) in this data set showed significantly longer metastasis-free survival, compared to about 50% who expressed higher levels of this pyrimidine metabolic gene (top third) and displayed worse outcome. Furthermore, RRM1 was not altered between tamoxifen responders and nonresponders in this data set, whereas RRM2b was not measured. Consistent with all of the above, in the Loi data set, pathway-centric enrichment of gene expression data stratifying patients into TAM responders *versus* nonresponders also significantly enriched pyrimidine metabolism (Supplementary Table 8).

We next verified the association between RRM2 overexpression and TAM resistance using TMA. Here, we restricted our analysis to ER+ BCa tissues from patients treated with only TAM in the adjuvant setting and followed up for tumor recurrence and survival for a median time period of 8 years. To begin with, the specificity of RRM2 antibody for the protein was established using an isotype nonimmune IgG control (Supplementary Figure 5). Following this, TMA staining using RRM2 antibody revealed significantly higher expression of the protein in patients who did not respond to TAM (Figure 3*E*). Consistent with this, patients who were reported to have died by the end of the follow-up

Figure 2. Gene expression signature analysis of metabolism-associated pathways in human breast tumors. (A) Graphical representation of enriched pathways obtained using GSA on cell line–derived metabolomics data. The enrichment *P* values and pathway ranking are given on the y- and x-axis, respectively. The circumference of each circle in the plot correlates to the number of metabolites in the pathway used for the GSA. (B) Same as in A, but for pathways obtained using GSA on tissue-derived metabolomics data. (C) Same as in A, but for pathways obtained using GSA on tissue-derived transcriptomics data. The circumference of each circle in the plot correlates to the number of genes in the pathway used for the GSA. (D) Associations with distant metastasis-free survival (by univariate Cox *P* value) involving combined sets of pathway-associated genes, for each of the indicated breast tumor gene expression data sets are presented. (E) KM plots show the association between combined expression of pyrimidine metabolism–associated genes and metastasis-free survival across 1340 BCa tissues in the Kessler compendium (tumors binned by top third, bottom third, and middle third of scores). (F) Same as in E, for pyrimidine metabolism–associated genes and metastasis-free survival across 686 ER + BCa tissues. (G) Heat maps show average expression of genes associated with the 11 enriched pathways in the Kessler compendium of breast tumor expression profiles ($N = 1340$). Relative expression for prognostic signatures MKI67, Oncotype Dx, Mammaprint, and Genomic grade are also shown.



dataset	Coefficient	95% Conf. (±)	Std.Error	P	Hazard = Exp(Coef.)
Wang ER+ (N=209)	0.48	0.24	0.12	<0.0001	1.61
Chin ER+ (N=74)	0.28	0.56	0.29	0.33	1.32
Desmedt ER+ (N=134)	0.27	0.34	0.17	0.13	1.30
Minn ER+ (N=46)	0.13	0.51	0.26	0.61	1.14
Loi ER+ (N=202)	0.57	0.29	0.15	<0.0001	1.77
Kessler (N=1340)	0.40	0.15	0.08	4.66E-07	1.49
van De vijver (N= 295)	0.62	0.26	0.13	3.56E-06	1.87

Figure 3. Pyrimidine metabolism–associated *RRM2* is differently expressed in breast cancer subtypes and is a predictor of outcome. (A) Box plot shows the relative expression of RRM2 in basal and luminal BCa tissues. (B) KM plot showing the association of RRM2 expression with time to metastasis-free survival in patients with BCa (N = 1340). Higher expression of RRM2 was significantly (log-rank P = 3.6E-09) associated with poor survival in BCa. (C) Table shows results of univariate Cox P values of RRM2 in each of the publically available data sets and its association with distant metastasis-free survival. (D) KM plot shows the association of RRM2 expression with time to metastasis-free survival in patients with tamoxifen-treated BCa (Loi data set, N = 149). Higher expression of RRM2 was significantly (log-rank P = .002) associated with tumors having intrinsic tamoxifen resistance and poor survival in this patient group. (E) RRM2 protein expression was generally higher in patients who did not respond to TAM compared to those who responded to the treatment. (F) RRM2 expression was significantly higher (Wilcoxon rank sum, P = .04) in patients who were reported to be dead *versus* those who were alive, post-TAM treatment for a median follow-up time of 8 years. G) KM plot confirms significant association (log-rank test, P = .04) of RRM2 expression with early tumor recurrence.

period expressed significantly higher RRM2 (Wilcoxon rank sum P = .04) compared to those who reported to be alive (Figure 3*F*). In addition, higher RRM2 expression associated significantly with earlier tumor recurrence (log-rank test, P = .04; Figure 3*G*), confirming the prognostic value of RRM2 in patients with luminal BCa treated with TAM.

Next, to determine the role of RRM2 in TAM resistance, RRM2-associated metabolites were measured in MCF-7 L cells that were treated with TAM either for a short-term (TAM-S) or longer time period (TAM-R) [58,59], as well as xenograft tumors [58] generated using MCF-7 L cells that were treated *in vivo* with TAM for 2 weeks (TAM-S) or 3 months (TAM-R). As expected from our earlier findings, in both cell lines and xenograft tumors, the ratio of the metabolite product:metabolite substrate for RRM2 (dUDP/UDP and dCDP/CDP) were significantly higher in TAM-R cells (P < .01; Figure 4, *A* and *B*, TAM-Cell lines and Figure 4, *C* and *D*, TAM-Xenograft),compared to TAM-S controls. A similar profile was also obtained for TAM-R cells when compared with parental untreated controls (MCF-7 L, Figure 4, *A* and *B*). Corroborating these findings, both protein (Figure 4*E*; refer to Supplementary Figure 6 C and D, *C* and *D*, for data using untreated parental cells) and mRNA levels (Figure 4*F*) of RRM2 were higher in TAM-R cells compared to TAM-S controls. Furthermore, transcript levels of RRM2 were also elevated in TAM-R xenograft tumors compared to TAM-S counterparts (Supplementary Figure 7). To further substantiate the association of RRM2 with TAM resistance, siRNA-based knockdown of the gene was carried out (immunoblot for Knock Down (KD) in Figure 4*G*) in TAM-R cells. As expected, siRNA-mediated KD of RRM2 expression in the resistant cells resulted in a significant decrease in cell proliferation compared to nontarget siRNA controls (Figure 4*H*). Consistent with this, the transcript levels for cyclins regulating G_1-S transition Cyclin D2 (CCND2) and Cyclin E1 (CCNE1) were significantly (P < .05) reduced in the KD cells with no change in the mRNA and levels of S-phase– or G2-phase–associated cyclins (CCNA1 and CCNB1) (Figure 4*I*).

Having garnered substantial evidence on the association of RRM2 with TAM resistance, we next evaluated the ability of the DNA methyltransferase inhibitor aza to sensitize TAM-R cells to TAM treatment. This was motivated by a recent report that showed down-regulation of RRM2 transcript and protein levels by aza in leukemic cell lines [60]. Consistent with the reported findings, treatment with 25 to 100 uM aza completely abolished RRM2 protein expression in

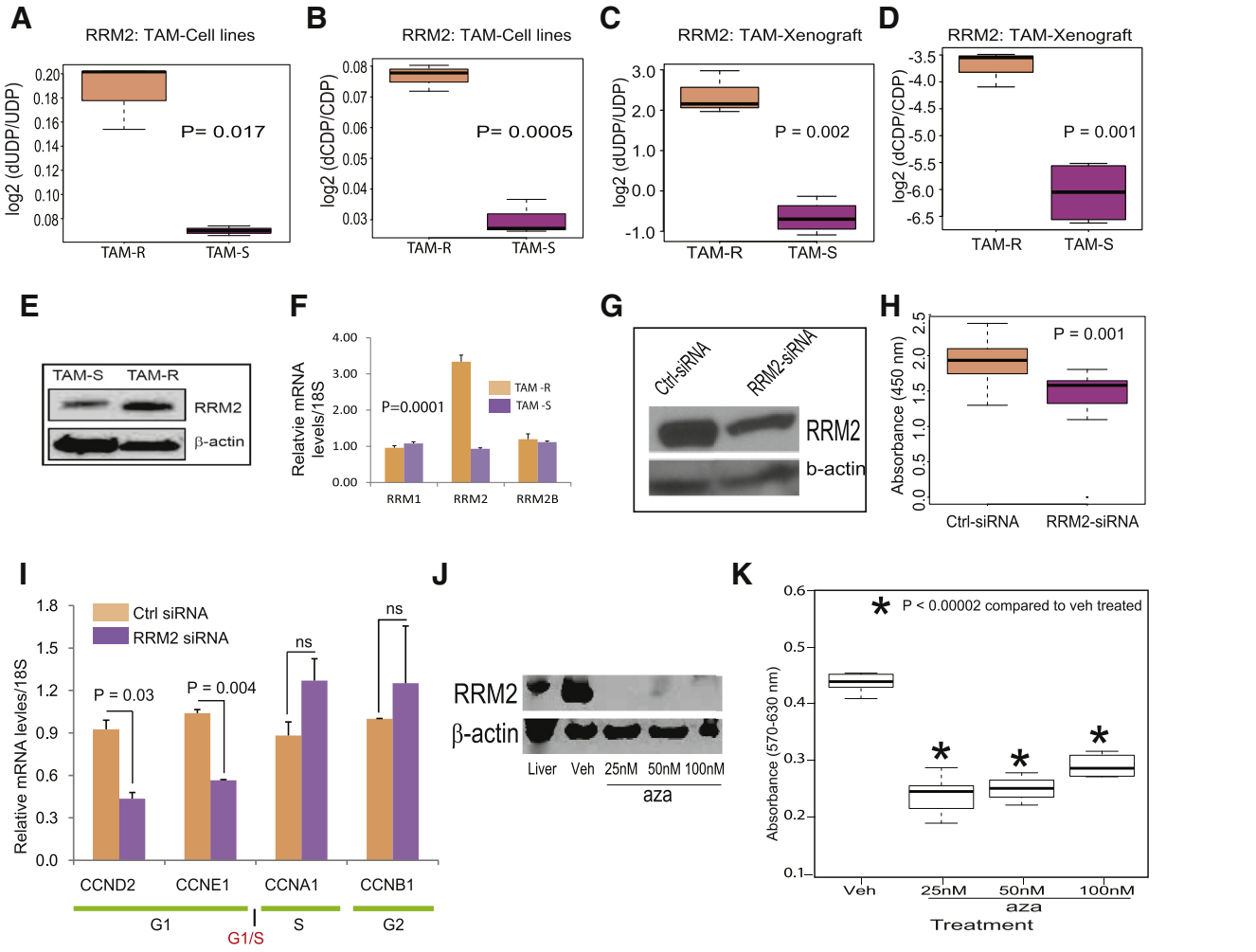


Figure 4. Elevated expression of RRM2 is associated with acquired tamoxifen resistance in BCa and could be targeted to sensitize the tumors to tamoxifen treatment. (A–D) Box plot shows the relative ratios of product:substrate for RRM2, i.e., dUDP/UDP and dCDP/CDP, in TAM-R ($n = 3$ biologic replicates) and TAM-S (each $n = 3$ biologic replicates) cell lines (A and B) and xenograft tissues (C and D), (TAM-R, $n = 3$ biologic replicates; TAM-S, $n = 2$ biologic, each in $n = 2$ technical replicates). The relative ratios for the product:substrates of RRM2 were higher in tamoxifen-resistant (TAM-R) cells ($P = .04$ and .19) and tissues ($P = .002$ and .001), compared to their parental counterparts. (E) Immunoblot analysis shows levels of RRM2 protein expression in TAM-R and TAM-S cells. β-Actin was used as a loading control. (F) Transcript levels of RRM1, RRM2, and RRM2B relative to 18S RNA in TAM-R and TAM-S cells are presented. (G) Immunoblot analysis to verify RRM2 KD in TAM-R cells. β-actin was used as a loading control. (H) RRM2 KD in TAM-R cells resulted in a significant decrease in proliferation (rank sum $P = .0001$) compared to control siRNA-treated cells, as assessed by BrdU assay. (I) Transcript levels of CCND2, CCNE1, CCNA1, and CCNB1 were measured after 48 hours of post-RRM2 siRNA or Ctrl siRNA treated cells. (J) Immunoblot analysis of aza-treated TAM R cells shows reduction in RRM2 expression. (K) TAM-R cells were treated with increasing concentrations of aza, and after 72 hours, an MTT-based assay was performed to determine the cell growth and survival.

TAM-R cells (Figure 4*J*). Furthermore, aza-treated TAM-R cells showed a significant reduction in the cell proliferation as assessed using BrdU assay (Figure 4*K*).

Discussion

To delineate the biochemical processes altered in BCa, we adopted a strategy wherein we started by defining the metabolic alterations in luminal and basal-like BCa cells that have been well characterized for their subtype gene expression and routinely used in laboratory studies [33]. The metabolic profiles were analyzed to generate pathways that were then examined for their relevance using clinical specimens. The caveats that argue against using cell line models for profiling studies include the alteration in their molecular profiles caused by culture conditions [61] and the lack of the intratumoral heterogeneity in cell lines [62]. In contrast, the challenge of using clinical specimens for translational research lies in interpatient variability as well as the heterogeneous nature of tumor-associated cell types. In spite of these potential confounders, it is encouraging to find a subset of metabolic pathways that were consistent between cell lines and patient tumors and distinguish luminal from basal-like BCa. The novelty of our integrative approach lies in identifying these commonly altered biochemical pathways using a pathway-centric rank-based method that takes into account both the degree of enrichment (which is a reflection of their biologic importance) of the pathway in patient-derived omics data and its association with patient prognosis (reflection of their clinical relevance). In this regard, our methodology is distinct from the ones that

rely on either a single data type (for example, gene expression, [63]) or combined data sets using concordance-based integrative methods [64]. However, our approach is comparable to the metabolome-centric method described by Imielinski et al. [65], with refinements that incorporate a strong patient-centric connotation to calculate the CR Score. This is exemplified in our results, wherein pyrimidine was ranked as the top pathway among the 11 possible contenders on the basis of both its quantity in patient tumors and its association with patient's clinical outcome.

Pyrimidine metabolism signifying the proliferative ability of tumor cells was also reflected by its elevated presence in PAM50-defined basal, Her-2–enriched and luminal B tumors. *RRM2* is a key gene in pyrimidine metabolism and has been earlier shown to be elevated in aggressive BCa [66]. In combination with other proliferative markers, RRM2 has also been found to have prognostic relevance in BCa [67]. Consistent with this, in our study, RRM2 expression by itself was able to distinguish good *versus* poor survivors within the entire group of patients with BCa that included a significant proportion of luminal A subtype. Importantly, from the clinical standpoint, luminal A patients are typically considered to have a better survival outcome. The clinical value of molecular predictors like PAM50 [68], Oncotype Dx [54], Mammaprint [47], and other prognostics stems from their ability to distinguish a subset of aggressive tumors within this so-called clinically indolent patient population. In light of this, it is remarkable to note that the prognostic value of RRM2 alone was comparable to all the above panel of markers, setting the stage for future prospective validation of this gene in independent patient specimens. Furthermore, in the setting of tamoxifen treatment that is routinely administered to ER+ patients, RRM2 was able to distinguish patients who were resistant to treatment from those who responded to the therapy. This is verified by higher expression and activity of RRM2 in TAM-R cell lines and xenograft models [58,69]. Furthermore, a significantly larger proportion of TAM-treated patients who died of BCa showed higher RRM2 protein expression, a novel finding that was further validated using TMA. Importantly, this finding does not imply a causal role for RRM2 in the onset of TAM resistance. Furthermore, the findings on higher RRM2 expression both in tumors with poor prognosis irrespective of the subtype and in TAM-R ER+ tumors warrant additional studies to establish the role of the protein in these patient populations.

RRM2 belongs to the family of ribonucleotide reductase that has two other isoforms, RRM1 and RRM2b. The three together catalyze the conversion of uridine/cytidine containing nucleotide triphosphates to their deoxygenated counterparts, a key step in DNA synthesis. It is important to note that the expression of RRM2 but neither RRM1 nor RRM2b was consistently and significantly elevated in tamoxifen-resistant cell line and xenograft samples as well as in publically available clinical data sets (Figure 4, *E* and *F*, and Supplementary Figure 7). In light of these findings and our observation that the ratio of the dNTP: NTP (deoxynucleotide triphosphate: nucleotide triphosphate) is higher in TAM-R *in vitro* and *in vivo* samples, we allude to the possibility that TAM resistance could be a reflection of increased RRM2 activity. However, other biochemical mechanisms, including a role for RRM1 and RRM2B, cannot be ruled out and need further examination. A recent report showed the induction of RRM2 by overexpressed AKT in TAM-R cells [70]. As AKT is known to promote proliferation and cell growth in multiple cancers, elevated expression of RRM2 could portray increased rate of DNA synthesis to support AKT-induced proliferation demand. Yet another possibility regarding role of RRM2 in the context of TAM resistance stems from our *in silico* findings (Supplementary Table 9) that suggest a potential regulation by estrogen receptor 1α (ESR1). Activating mutations of ESR1 have been reported by multiple groups to be associated with endocrine resistance [71,72]. These findings point to the possibility of increased RRM2 expression as being a downstream consequence of activated ESR1 in patients, a hypothesis that needs to be validated. In addition, a recent report in melanoma suggests that RRM2 could induce cellular senescence [73] and hence create a mechanism for the tumors to escape cytotoxic effect of the therapy. This is an interesting possibility that needs to be examined in the context of TAM resistance.

Importantly, a strong association between RRM2 expression and TAM resistance led us to test potential inhibitors whose activity has been reported to reverse RRM2 expression. Azacytidine, a well-known DNA methyltransferase inhibitor, was recently reported to decrease mRNA and protein levels of RRM2 in leukemic cell lines although its mechanism of action was not reported ([60]). Consistent with their findings, RRM2 protein expression in our hands was significantly reduced on treatment of TAM-R cells with aza. In line with this, aza-treated TAM-R cells showed significantly decreased rate of proliferation in the presence of TAM, indicative of potential resensitization to the TAM treatment. Although preliminary, these findings set the stage for additional experiments to determine the optimal concentration of aza required to supplement tamoxifen and to understand the mechanism(s) that drive this synergy.

Conclusions

Taken together, we have developed a novel bioinformatics method to integrate metabolomics and gene expression data from cancer cell lines and tissues to nominate key pathways that are altered and have prognostic value. *RRM2*, a key gene in pyrimidine metabolism, was found to be associated with aggressive breast tumors as well as TAM-R luminal BCas, and its pharmacological or genetic knockdown sensitized tumors to TAM. In summary, the study nominates RRM2 as a key marker for aggressive BCa including TAM-R tumors. In light of this finding, RRM2-associated metabolites could be developed as prognostic markers for BCa. In addition, the combination of aza with TAM could be explored in a preclinical setting to treat TAM- resistant BCas.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2014.05.007.

References

[1] Jemal A, Siegel R, Ward E, Hao Y, Xu J, and Thun MJ (2009). *Cancer Statistics, 2009. CA Cancer J Clin* **59**, 225–249.
[2] Carter CL, Allen C, and Henson DE (1989). Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* **63**(1), 181–187.
[3] McKinney CD, Frierson Jr HF, Fechner RE, Wilhelm MC, and Edge SB (1992). Pathologic findings in nonpalpable invasive breast cancer. *Am J Surg Pathol* **16**(1), 33–36.
[4] O'Shaughnessy J, Miles D, Vukelja S, Moiseyenko V, Ayoub JP, Cervantes G, Fumoleau P, Jones S, Lui WY, and Mauriac L, et al (2002). Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: phase III trial results. *J Clin Oncol* **20**(12), 2812–2823.
[5] Coon JS, Marcus E, Gupta-Burt S, Seelig S, Jacobson K, Chen S, Renta V, Fronda G, and Preisler HD (2002). Amplification and overexpression of topoisomerase IIα predict response to anthracycline-based therapy in locally advanced breast cancer. *Clin Cancer Res* **8**(4), 1061–1067.
[6] Toft DJ and Cryns VL (2011). Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies. *Mol Endocrinol* **25**, 199–211.

[7] Niida A, Smith AD, Imoto S, Aburatani H, Zhang MQ, and Akiyama T (2009). Gene set-based module discovery in the breast cancer transcriptome. *BMC Bioinforma* **10**, 71.

[8] Finnegan TJ and Carey LA (2007). Gene-expression analysis and the basal-like breast cancer subtype. *Future Oncol* **3**(1), 55–63.

[9] Kang S, Kim MJ, An H, Kim BG, Choi YP, Kang KS, Gao MQ, Park H, Na HJ, and Kim HK, et al (2010). Proteomic molecular portrait of interface zone in breast cancer. *J Proteome Res* **9**, 5638–5645.

[10] Cha S, Imielinski MB, Rejtar T, Richardson EA, Thakur D, Sgroi DC, and Karger BL (2010). *In situ* proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology. *Mol Cell Proteomics* **9**, 2529–2544.

[11] Rauser S Marquardt C, Balluff B, Deininger SO, Albers C, Belau E, Hartmer R, Suckau D, Specht K, and Ebert MP, et al (2010). Classification of HER-2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *J Proteome Res* **9**, 1854–1863.

[12] He J, Shen D, Chung DU, Saxton RE, Whitelegge JP, Faull KF, and Chang HR (2009). Tumor proteomic profiling predicts the susceptibility of breast cancer to chemotherapy. *Int J Oncol* **35**(4), 683–692.

[13] Li J, Gromov P, Gromova I, Moreira JM, Timmermans-Wielenga V, Rank F, Wang K, Li S, Li H, and Wiuf C, et al (2008). Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis. *Proteomics* **8**(23–24), 5038–5052.

[14] Datta S (2008). Classification of breast cancer *versus* normal samples from mass spectrometry profiles using linear discriminant analysis of important features selected by random forest. *Stat Appl Genet Mol Biol* **7**(2) [Article7].

[15] Laronga C and Drake RR (2007). Proteomic approach to breast cancer. *Cancer Control* **14**(4), 360–368.

[16] Nakagawa T, Huang SK, Martinez SR, Tran AN, Elashoff D, Ye X, Turner RR, Giuliano AE, and Hoon DS (2006). Proteomic profiling of primary breast cancer predicts axillary lymph node metastasis. *Cancer Res* **66**(24), 11825–11830.

[17] Nagaraja GM, Othman M, Fox BP, Alsaber R, Pellegrino CM, Zeng Y, Khanna R, Tamburini P, Swaroop A, and Kandpal RP (2006). Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. *Oncogene* **25**(16), 2328–2338.

[18] Carr KM, Rosenblatt K, Petricoin EF, and Liotta LA (2004). Genomic and proteomic approaches for studying human cancer: prospects for true patient-tailored therapy. *Hum Genomics* **1**(2), 134–140.

[19] Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, and Di Leo A (2007). Metabolomics: available results, current research projects in breast cancer, and future applications. *J Clin Oncol* **25**(19), 2840–2846.

[20] Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, and Raftery D (2008). Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* **8**(5), 617–633.

[21] Oakman C, Tenori L, Claudino WM, Cappadona S, Nepi S, Battaglia A, Bernini P, Zafarana E, Saccenti E, and Fornier M, et al (2011). Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods. *Ann Oncol* **22**(6), 1295–1301.

[22] Cheng LL, Chang IW, Smith BL, and Gonzalez RG (1998). Evaluating human breast ductal carcinomas with high-resolution magic-angle spinning proton magnetic resonance spectroscopy. *J Magn Reson* **135**(1), 194–202.

[23] Sitter B, Sonnewald U, Spraul M, Fjösne HE, and Gribbestad IS (2002). High-resolution magic angle spinning MRS of breast cancer tissue. *NMR Biomed* **15**(5), 327–337.

[24] Bathen TF, Jensen LR, Sitter B, Fjösne HE, Halgunset J, Axelson DE, Gribbestad IS, and Lundgren S (2007). MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. *Breast Cancer Res Treat* **104**(2), 181–189.

[25] Sitter B, Lundgren S, Bathen TF, Halgunset J, Fjosne HE, and Gribbestad IS (2006). Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters. *NMR Biomed* **19**(1), 30–40.

[26] Yang C, Richardson AD, Smith JW, and Osterman A (2007). Comparative metabolomics of breast cancer. *Pac Symp Biocomput* , 181–192.

[27] Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, Bass AJ, Heffron G, Metallo CM, Muranen T, and Sharfi H, et al (2011). Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat Genet* **43**(9), 869–874.

[28] Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, and Jha AK, et al (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**(7360), 346–350.

[29] Putluri N, Shojaie A, Vasu VT, Nalluri S, Vareed SK, Putluri V, Vivekanandan-Giri A, Byun J, Pennathur S, and Sana TR, et al (2011). Metabolomic profiling reveals a role for androgen in activating amino acid metabolism and methylation in prostate cancer cells. *PLoS One* **6**(7), e21417.

[30] Putluri N, Shojaie A, Vasu VT, Vareed SK, Nalluri S, Putluri V, Thangjam GS, Panzitt K, Tallman CT, and Butler C, et al (2011). Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. *Cancer Res* **71**(24), 7376–7386.

[31] Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, and Li Y, et al (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**(7231), 910–914.

[32] Vareed SK, Bhat VB, Thompson C, Vasu VT, Fermin D, Choi H, Creighton CJ, Gayatri S, Lan L, and Putluri N, et al (2011). Metabolites of purine nucleoside phosphorylase (NP) in serum have the potential to delineate pancreatic adenocarcinoma. *PLoS One* **6**(3), e17177.

[33] Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, and Tong F, et al (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**(6), 515–527.

[34] Efron B and Tibshirani R (2007). On testing the significance of sets of genes. *Ann Appl Stat* , 107–129.

[35] Morrison G, Fu X, Shea M, Nanda S, Giuliano M, Wang T, Klinowska T, Osborne CK, Rimawi MF, and Schiff R (2014). Therapeutic potential of the dual EGFR/HER-2 inhibitor AZD8931 in circumventing endocrine resistance. *Breast Cancer Res Treat* **144**(2), 263–272.

[36] Zhang Y, Tseng CC, Tsai YL, Fu X, Schiff R, and Lee AS (2013). Cancer cells resistant to therapy promote cell surface relocalization of GRP78 which complexes with PI3K and enhances PI(3,4,5)P3 production. *PLoS One* **8**(11), e80071.

[37] R Development Core Team (2008). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008 [ISBN 3-900051-07-0, URL http://www.R-project.org.].

[38] Wong J (2011). Imputation: imputation. R package version 1.3. http://CRAN.R-project.org/package=imputation; 2011.

[39] Terunuma A, Putluri N, Mishra P, Mathé EA, Dorsey TH, Yi M, Wallace TA, Issaq HJ, Zhou M, and Killian JK, et al (2014). MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J Clin Invest* **124**(1), 398–412.

[40] Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57**, 289–300.

[41] Xiao Y and Yang JYH (2007). Differential Expression *via* Distance Summary for Microarray Data R package version 1.30.0; 2007 .

[42] Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**(2), 249–264.

[43] Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, and Gentry J, et al (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80.

[44] Bolker B, et al (2011). gplots: various R programming tools for plotting data. In: Warnes G, editor. R Package Version 2.10.1; 2011.

[45] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, and Lander ES, 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**(43), 15545–15550.

[46] Tibshirani R and Efron B (2010). GSA: Gene set analysis. R package version 1.03. http://CRAN.R-project.org/package=GSA; 2010.

[47] van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, and Witteveen AT, et al (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**(6871), 530–536.

[48] Kessler JD, Kahle KT, Sun T, Meerbrey KL, Schlabach MR, Schmitt EM, Skinner SO, Xu Q, Li MZ, and Hartman ZC, et al (2012). A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Sci Signal* **335**(6066), 348–353.

[49] Creighton CJ (2012). The molecular profile of luminal B breast cancer. *Biologics* **6**, 289–297.

[50] Kommagani R, Szwarc MM, Kovanci E, Gibbons WE, Putluri N, Maity S, Creighton CJ, Sreekumar A, DeMayo FJ, and Lydon JP, et al (2013).

Acceleration of the glycolytic flux by steroid receptor coactivator-2 is essential for endometrial decidualization. *PLoS Genet* **9**, e1003900.

[51] Kanehisa M and Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**(1), 27–30.

[52] Ellis MJ, Suman VJ, Hoog J, Lin L, Snider J, Prat A, Parker JS, Luo J, DeSchryver K, and Allred DC, et al (2011). Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor–rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype—ACOSOG Z1031. *J Clin Oncol* **29**(17), 2342–2349.

[53] Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, and Parker JS, et al (2009). Ki67 index, HER-2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst* **101**(10), 736–750.

[54] Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, and Park T, et al (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351**(27), 2817–2826.

[55] Slodkowska EA and Ross JS (2009). MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* **9**(5), 417–422.

[56] Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, Ellis P, Harris A, Bergh J, and Foekens JA, et al (2007). Definition of clinically distinct molecular subtypes in estrogen receptor–positive breast carcinomas through genomic grade. *J Clin Oncol* **25**(10), 1239–1246.

[57] Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, Gillet C, Ellis P, Ryder K, and Reid JF, et al (2008). Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**(1), 239.

[58] Massarweh S, Osborne CK, Creighton CJ, Qin L, Tsimelzon A, Huang S, Weiss H, Rimawi M, and Schiff R (2008). Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res* **68**(3), 826–833.

[59] Fu L, Qiu W, Yu Y, Guo Y, Zhao P, Zhang X, Liu C, Li F, Huang H, and Huang M, et al (2014). Clinical and molecular genetic study of infantile-onset Pompe disease in Chinese patients: identification of 6 novel mutations. *Gene* **535**(1), 53–59.

[60] Aimiuwu J, Wang H, Chen P, Xie Z, Wang J, Liu S, Klisovic R, Mims A, Blum W, and Marcucci G, et al (2012). RNA-dependent inhibition of ribonucleotide reductase is a major pathway for 5-azacytidine activity in acute myeloid leukemia. *Blood* **119**(22), 5229–5238.

[61] Chan R and Wong MS (2007). Differential regulation of cyclic AMP synthesis by estrogen in MCF7 cells. *Biochem Biophys Res Commun* **363**(3), 616–620.

[62] Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnick JA, DiMeo TA, Gilmore H, Jefferson DM, and Graham RA, et al (2010). Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res* **12**(5), R87.

[63] Kristensen VN, Vaske CJ, Ursini-Siegel J, Van Loo P, Nordgard SH, Sachidanandam R, Sørlie T, Wärnberg F, Haakensen VD, and Helland Å, et al (2012). Integrated molecular profiles of invasive breast tumors and ductal carcinoma *in situ* (DCIS) reveal differential vascular and interleukin signaling. *Proc Natl Acad Sci U S A* **109**(8), 2802–2807.

[64] Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, and Saito K (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **101**(27), 10205–10210.

[65] Imielinski M, Cha S, Rejtar T, Richardson EA, Karger BL, and Sgroi DC (2012). Integrated proteomic, transcriptomic, and biological network analysis of breast carcinoma reveals molecular features of tumorigenesis and clinical relapse. *Mol Cell Proteomics* **11**(6) [M111 014910].

[66] Furuta E, Okuda H, Kobayashi A, and Watabe K (2010). Metabolic genes in cancer: their roles in tumor progression and clinical implications. *Biochim Biophys Acta* **1805**(2), 141–152.

[67] Mercier I, Casimiro MC, Wang C, Rosenberg AL, Quong J, Minkeu A, Allen KG, Danilo C, Sotgia F, and Bonuccelli G, et al (2008). Human breast cancer-associated fibroblasts (CAFs) show caveolin-1 downregulation and RB tumor suppressor functional inactivation: implications for the response to hormonal therapy. *Cancer Biol Ther* **7**(8), 1212–1225.

[68] Peppercorn J, Perou CM, and Carey LA (2008). Molecular subtypes in breast cancer evaluation and management: divide and conquer. *Cancer Invest* **26**(1), 1–10.

[69] Creighton CJ, Massarweh S, Huang S, Tsimelzon A, Hilsenbeck SG, Osborne CK, Shou J, Malorni L, and Schiff R (2008). Development of resistance to targeted therapies transforms the clinically associated molecular profile subtype of breast tumor xenografts. *Cancer Res* **68**(18), 7493–7501.

[70] Shah KN, Mehta KR, Peterson D, Evangelista M, Livesey JC, and Faridi JS (2014). AKT-induced tamoxifen resistance is overturned by RRM2 inhibition. *Mol Cancer Res* **12**(3), 394–407.

[71] Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, and Hodges L, et al (2013). Activating *ESR1* mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**(12), 1446–1451.

[72] Fuqua SA, Gu G, and Rechoum Y (2014). Estrogen receptor (ER) α mutations in breast cancer: hidden in plain sight. *Breast Cancer Res Treat* **144**(1), 11–19.

[73] Aird KM, Zhang G, Li H, Tu Z, Bitler BG, Garipov A, Wu H, Wei Z, Wagner SN, and Herlyn M, et al (2013). Suppression of nucleotide metabolism underlies the establishment and maintenance of oncogene-induced senescence. *Cell Rep* **3**(4), 1252–1265.