Combined Label-Free Quantitative Proteomics and microRNA Expression Analysis of Breast Cancer Unravel Molecular Differences with Clinical Implications

Angelo Gámez-Pozo¹, Julia Berges-Soria¹, Jorge M. Arevalillo², Paolo Nanni³, Rocío López-Vacas¹, Hilario Navarro², Jonas Grossmann⁴, Carlos A. Castaneda⁴, Paloma Main⁵, Mariana Díaz-Almirón⁶, Enrique Espinosa⁷, Eva Ciruelos⁸, and Juan Ángel Fresno Vara¹

Abstract

Better knowledge of the biology of breast cancer has allowed the use of new targeted therapies, leading to improved outcome. High-throughput technologies allow deepening into the molecular architecture of breast cancer, integrating different levels of information, which is important if it helps in making clinical decisions. microRNA (miRNA) and protein expression profiles were obtained from 71 estrogen receptor–positive (ER⁺) and 25 triple-negative breast cancer (TNBC) samples. RNA and proteins obtained from formalin-fixed, paraffin-embedded tumors were analyzed by RT-qPCR and LC/MS-MS, respectively. We applied probabilistic graphical models representing complex biologic systems as networks, confirming that ER⁺ and TNBC subtypes are distinct biologic entities. The integration of miRNA and protein expression data unravels molecular processes that can be related to differences in the genesis and clinical evolution of these types of breast cancer. Our results confirm that TNBC has a unique metabolic profile that may be exploited for therapeutic intervention.

Introduction

Breast cancer is a major health issue in developed countries. Early diagnosis and the use of adjuvant therapies have contributed to improved survival, but still more than 95,000 women are expected to die of breast cancer in the European Union in 2015 (1). In the last decades, the death rate from breast cancer has decreased gradually, due to early diagnosis and to the availability of adjuvant chemotherapy and hormonal therapy. Early diagnosis allows the detection of in situ or very small infiltrating tumors, which have an excellent prognosis, whereas adjuvant therapy decreases the rate of relapse. For this reasons, factors determining the use of such therapies are key elements in patient management. Knowledge of the molecular biology of breast cancer has recently challenged the way in which oncologists make decisions on adjuvant chemotherapy. A number of techniques, such as gene profiling, proteomics, and microRNA (miRNA) analysis, have been used to study this disease.

Gene profiling in breast cancer has received increasing attention in recent years as a way to provide prognostic and predictive information. Since the description of a 70-gene profile in 2002, this and other gene signatures, such as Recurrence Score, have made their way into the clinic very rapidly, because they provide clinically useful information (2).

Proteomics would enable the unbiased comparison of different cellular states in biology and medicine at a systems-wide level (3). Proteome analysis heavily relies on mass spectrometry (MS). MS-based proteomics is starting to mature and to deliver through a combination of developments in instrumentation, sample preparation, and computational analysis. These advances allow the identification of thousands of proteins from tissue amounts compatible with clinical routine, which is relevant for the study of complex diseases. For example, thousands of mutations have been described associated with cancer (4), but the exact relationship between genomic variation and the resulting phenotype in each individual tumor remains unknown. Deep proteomics analyses are necessary to characterize the complete scenario of signaling pathways and biologic processes altered as a result of the specific mutation set produced in each tumor.

¹Molecular Oncology and Pathology Lab, Instituto de Genética Médica y Molecular–INGEMM, Instituto de Investigación Hospital Universitario La Paz-IdiPAZ, Madrid, Spain. ²Department of Statistics, Operational Research and Numerical Analysis, University Nacional Educación a Distancia (UNED), Madrid, Spain. ³Functional Genomics Centre Zurich, University of Zurich/ETH Zurich, Zurich, Switzerland. ⁴Departamento de Investigación, Instituto Nacional de Enfermedades Neoplásicas, Lima, Surquillo-Lima, Peru. ⁵Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid, Spain. ⁶Biostatistics Unit, Instituto de Investigación Hospital Universitario La Paz-IdiPAZ, Madrid, Spain. ⁷Medical Oncology Service, Instituto de Investigación Hospital Universitario La Paz-IdiPAZ, Madrid, Spain. ⁸Medical Oncology Service, Instituto de Investigación Hospital Universitario Doce de Octubre–I12, Madrid, Spain.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Juan Ángel Fresno Vara, Instituto de Genética Médica y Molecular (INGEMM), Hospital Universitario La Paz-IdiPAZ, Paseo de la Castellana, 267-28046, Madrid, Spain. Phone: 34-91-2071010, ext. 255; Fax: 34-91-2071040; E-mail: juanangel.fresno@salud.madrid.org
doi: 10.1158/0008-5472.CAN-14-1937
©2015 American Association for Cancer Research.

www.aacrjournals.org
miRNAs play an important role as regulators of gene expression, controlling many biologic processes such as growth, development, differentiation, and apoptosis. This has led to considerable interest in the development of miRNA expression profiles as a new class of biomarkers in cancer, cardiovascular, and autoimmune diseases. Measuring miRNA expression can also be useful for systems-level studies of gene regulation, especially when miRNA measurements are combined with mRNA profiling and other genome-scale data (5).

A major limitation for the widespread use of high-throughput techniques is tissue availability. Fresh or fresh-frozen specimens are ideal for molecular profiling but rarely available in the clinic. In recent years, we and others have demonstrated that formalin-fixed, paraffin-embedded (FFPE) tissue can be readily analyzable by MS-based technologies (6, 7). On the other hand, previous studies demonstrate the possibility of performing experiments to analyze the expression of miRNAs in FFPE tissues (8). FFPE samples also constitute the basis for routine workup in pathology departments all over the world.

In this article, we combined quantitative proteomics with miRNA expression analyses in a series of breast cancer samples with appropriate clinical information. We demonstrate that it is possible to perform differential protein expression analyses by LC/MS-MS on tens of FFPE tumors. Protein patterns confirm that estrogen receptor–positive (ER+) and triple-negative breast cancer (TNBC) subtypes are distinct biologic entities (9). Also, we were able to profile miRNA expression in the same series of samples, which led us to conduct a probabilistic graphical models analysis to integrate miRNA and protein expression data. This systems-level study sheds light on the molecular processes differentially regulated in these two main subtypes of breast cancer.

Materials and Methods
Study design and sample description
Our primary objectives were to develop a method for the study of both proteins and miRNA expression pattern from a systems biology point of view, and to identify differences in biologic processes between the two main groups of breast cancer patients: ER+ and TNBC. One hundred and six FFPE samples from patients diagnosed of breast cancer were retrieved from 1+12 Biobank (RD09/0076/00118) and from IdiPaz Biobank (RD09/0076/00073), both integrated in the Spanish Hospital Biobank Network (RetBioH; www.redbiobancos.es). The histopathologic features of each sample were reviewed by an experienced pathologist to confirm diagnosis and tumor content. Eligible samples had to include at least 50% of tumor cells. Approval from the Ethical Committees of Hospital Doce de Octubre and Hospital Universitario La Paz (Madrid, Spain) was obtained for the conduct of the study.

Total protein preparation
Proteins were extracted from FFPE samples as previously described (10). Briefly, FFPE sections were deparaffinized in xylene and washed twice with absolute ethanol. Protein extracts from FFPE samples were prepared in 2% SDS buffer using a protocol based on heat-induced antigen retrieval (6). Protein concentration was determined using the MicroBCA Protein Assay Kit (Pierce-Thermo Scientific). Protein extracts (10 μg) were digested with trypsin (1:50) and SDS was removed from digested lysates using Detergent Removal Spin Columns (Pierce). Peptide samples were further desalted using ZipTips (Millipore), dried, and resolubilized in 15 μL of a 0.1% formic acid and 3% acetonitrile solution before MS analysis.

Liquid chromatography–mass spectrometry analysis
Samples (4 μL) were analyzed on a LTQ-Orbitrap Velos hybrid mass spectrometer (Thermo Fischer Scientific) coupled to NanoLC-Ultra system (Eksigent Technologies). Peptide separation was performed on a self-made column (75 μm × 150 mm) packed with Magic RP C18 AQ, 200A, 3-μm beads (Bischoff GmbH) at a flow rate of 250 nL/min. Solvent composition was 0.1% formic acid for channel A, and 0.1% formic acid and 99.9% acetonitrile for channel B. The peptides were eluted with a gradient of 5% to 30% acetonitrile in 95 minutes. The mass spectrometer was operated in data-dependent mode (DDA), acquiring a full-scan MS spectra (300–1,700 m/z) at a resolution of 30,000 at 400 m/z after accumulation to a target value of 1,000,000, followed by collision-induced dissociation (CID) fragmentation on the 20 most intense signals per cycle. CID spectra were acquired using normalized collision energy of 35 and a maximum injection time of 50 milliseconds. The automatic gain control was set to 10,000 ions. Charge state screening was enabled and singly and unassigned charge states were rejected. Precursor masses previously selected for MS/MS measurement were excluded from further selection for 45 seconds, and the exclusion window was set at 10 ppm. The size of the exclusion list was set to a maximum of 500 entries. The samples were acquired using internal lock mass calibration on m/z 429.088735 and 445.120025.

Label-free protein expression data processing
The acquired raw MS data were processed by MaxQuant (version 1.2.7.4), followed by protein identification using the integrated Andromeda search engine. Spectra were searched against a UniProtKB/Swiss-Prot database for human (73,849 entries), including a set of 260 common protein contaminants (NCBI taxonomy ID 9606, downloaded on December 13, 2011). Reversing the protein sequences was chosen as decay option in MaxQuant. Carbamidomethylation of cysteine was set as fixed modification, while oxidation (M), deamidation (N, Q), and N-terminal protein acetylation were set as variable. Enzyme specificity was set to trypsin/P allowing a minimal peptide length of seven amino acids and a maximum of two missed cleavages. Precursor tolerance was set to 20 ppm, while fragment was set to 0.5 Da. The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for proteins.

Label-free quantification was performed by setting a 2-minute window for match between runs. The protein abundance was calculated on the basis of the normalized spectral protein intensity (LFQ intensity). Quantifiable proteins were defined as those detected in at least 75% of samples in at least one type of sample (either ER+ or TNBC) samples showing two or more unique peptides. Only quantifiable proteins were considered for subsequent analyses. Protein expression data were log-transformed and missing values were replaced using data imputation for label-free data using default values. Finally, protein expression values were z-score transformed. Batch effects were estimated and corrected using ComBat (11).

All the MS raw data files acquired in this study may be downloaded from Chorus (http://chorusproject.org) under the project.
name Breast Cancer Proteomics. The peptides output file from the MaxQuant analysis is provided in the Supplementary Data.

RNA extraction and miRNA expression

Selected FFPE tumor specimens were cut into serial sections with a thickness of 10 μm. Total RNA was then isolated using the RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies). Purified RNA quality control for quantity and purity was assessed using an ND-1000 NanoDrop spectrophotometer (Thermo Fisher Scientific). A preliminary analysis was performed to compare miRNA expression measurements between paired FF and FFPE samples, as described previously (12). miRNA expression profiling was performed using a custom TaqMan Array MicroRNA Card (Life Technologies) containing 95 FFPE-reliable assays, including four housekeeping miRNAs identified using NorMean (12), plus one mandatory control. RNA concentration was adjusted to 166.7 ng/μL. For miRNA cDNA synthesis, 500 ng of total RNA was reverse-transcribed using the TaqMan MicroRNA Reverse Transcription Kit combined with the stem-loop Megaplex Primer Pools according to the manufacturer’s protocol (Life Technologies). All PCR reactions were performed on the ABI Prism 7900HT Sequence detection system (Life Technologies). Average cycling threshold (Cq) values were obtained using SDS 2.2 software (Applied Biosystems) using automatic baseline settings and a threshold of 0.2. The maximum Cq value was set at 40. Cq values were normalized using two out of four reference miRNAs (hsa-let-7d and hsa-let-7g) selected as previously described (12). One miRNA was excluded from subsequent analyses because no detection was found in most of the samples. The relative expression level of each target gene was expressed as ΔCq = Cqgene − Cqref. Cqref was obtained calculating the geometric mean of two housekeeping miRs (let-7d and let-7g). Reference-normalized expression measurements were adjusted by defining the lowest expression value as 0, with subsequent 1-unit increases reflecting an approximate doubling of the RNA. Finally, values were z-scored.

Assessing molecular differences between ER+ and TNBC samples

Significance analysis of microarrays (SAM) was carried out to find differentially expressed proteins and miRNAs between ER+ and TNBC samples with a FDR below 5%. Hierarchical clusters were constructed with the differentially expressed proteins or miRNAs between ER+ and TNBC samples identified by SAM, using Pearson Correlation and the average-linkage method. Then, we used miRWalk (March 15, 2013 version; ref. 13) to find validated targets of these miRNAs among the proteins differentially expressed between ER+ and TNBC samples.

Network construction

Statistical analyses were conducted to associate miRNAs and proteins. As a first approach to describe associations present in our database, we choose probabilistic graphical models compatible with high dimensionality. The result is an undirected graphical model with local minimum Bayesian Information Criterion (BIC; ref. 14) obtained after executing the next steps: first the spanning tree with maximum likelihood is found and then (I), a forward search is performed by successively adding edges that reduce the BIC and still preserve the decomposability (15) of the initial graph (II). In the first stage, in order to learn a Markov tree structure from a random sample of a supposed multidimensional normal population, we used the extension of the Chow–Liu solution (16), according to which, for categorical data, the maximum likelihood structure is given by the maximum weight spanning tree (17) with empirical mutual information quantities (18) as edge weights. In the Gaussian case, a similar reasoning applies, but now the mutual information value is \(-\frac{1}{2}\log(1 - r^2)\), where \(r\) is the empirical correlation coefficient between the two variables (nodes) joined by the edge. Given that the algorithm is invariant under monotone transformations of the variables, \(r^2\) can be used as a weight. In the second phase, we introduce the BIC criterion that penalizes more complex models and then, simpler graphs are generated. This is a fundamental objective in high-dimensional problems. Both methods are implemented in the open-source statistical programming language R (19). In particular, the functions minForest and stepw, in the gRapHD package (20), are used for phases I and II, respectively.

Gene Ontology analyses

Protein-to-gene ID conversion was performed using Uniprot (http://www.uniprot.org) and DAVID. Gene Ontology analyses were performed using the functional annotation chart tool provided by DAVID. We used homo sapiens as background list and selected only GOTERM-FAT Gene Ontology categories and Biocarta, KEGG, and Panther pathways.

Functional node identification and activity measurement

To identify functional nodes within the probabilistic graphical models, we split it in several branches. Then, we used Gene Ontology analyses to investigate which function or functions were over-represented in each branch. To measure the functional activity of each node, we calculated the mean expression of all the proteins included in one branch related with a concrete function. Differences in node activity between ER+ and TNBC samples were assessed by class comparison analyses.

Orthogonal verification

We used published array expression data of 1,296 primary breast carcinomas from two previously published works (21, 22). Batch effects between datasets were estimated and corrected using ComBat (11). After protein-to-gene ID conversion, all probes in dataset for each gene were retrieved. Probes with higher coefficient of variation were selected when multiple probes were found for a single gene. Differences in node activity between ER+ and TNBC samples were assessed using a two-sample t test analysis comparing expression. A multivariate permutations test was computed on the basis of 1,000 random permutations, allowing a confidence level of FDR assessment of 90% and a maximum allowed proportion of false-positive genes of 0.05.

Cell culture

The ER+ cell lines MCF7 and T47D and the TNBC cell lines MDA-MB-231 and MDA-MB-468 were cultured in RPMI-1640 medium with phenol red (Gibco) supplemented with 10% heat-inactivated fetal bovine serum, 100 mg/mL penicillin, and 100 mg/mL streptomycin. The TNBC cell line BT-20 was cultured in Minimum Essential Medium supplemented with 10% heat-inactivated fetal bovine serum, 100 mg/mL penicillin, and 100 mg/mL streptomycin. All cell lines were cultured at 37°C in a humidified
atmosphere with 5% (v/v) CO₂ in air. MCF7, T47D BT-20, and MDA-MB-231 cell lines were kindly provided by Dra. Nurta Vilabloa (Hospital Universitario La Paz, previously obtained from the ATCC in January 2014). MDA-MB-468 cell line was obtained from the ATCC (July 2014). Cell lines were routinely monitored in our laboratory and authenticated by morphology and growth characteristics, tested for Mycoplasma and frozen, and passaged for fewer than 6 months before experiments.

**Glucose and lactate measurements**

Cells (250,000) were plated in 4 mL of fresh medium, and cultivated for 72 hours. Afterward, supernatants were centrifuged to eliminate cellular debris. Lactate and glucose concentrations in cell supernatants and fresh media were measured on an amperometric electrode using the enzymes lactate oxidase and glucose oxidase, respectively, in an ABL90 FLEX gas analyzer (Radiometer) with integrated gas, electrolyte, metabolite, and CO₂-exometry measurements. At least measurements from three different experiments are provided. Glucose consumption is the difference between glucose concentration in fresh medium and glucose concentration in cell supernatant after 74 hours of cell culture. Lactate production is the difference between lactate concentration in cell supernatant after 72 hours of cell culture and lactate concentration in fresh medium. Differences were assessed by the Kruskall–Wallis test.

**Statistical analyses and software suites**

BRB-ArrayTools, SPSS v16software package, GraphPad Prism 5.1 and R v2.15.2 (with the Design software package 0.2.3) were used for all statistical analyses. All P values were two-sided and P < 0.05 was considered statistically significant. Expression data and network analyses were performed in MeV and Cyto-scape software suites. BRB-ArrayTools has been developed by Dr. Richard Simon (Biometric Research Brand, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, MD) and BRB-ArrayTools Development Team.

**Results**

**Study goal and patient’s characteristics**

Our study aims to develop a method to study both proteins and miRNA expression pattern from a systems biology point of view, to discover differences in biologic processes between the two main groups of breast cancer patients: ER⁺ and TNBC. One hundred and six patients were included in the study. Clinical characteristics from these patients are provided in Table 1. All patients were node positive; all of them were negative when tested for Her2 amplification using immunohistochemistry and fluorescent in situ hybridization when needed; and all of them received adjuvant chemotherapy (either anthracycline-based or not) plus hormonal therapy in the ER⁺ group.

**Protein extraction and MS analysis**

We have previously demonstrated that our workflow is high reproducible (10). One hundred and two FFPE samples yielded enough protein to perform the MS analyses (typically, more than 0.5 mg of total protein was recovered from each sample). After MS workflow, 96 samples gave useful protein expression data (25 TNBCs and 71 ER⁺ tumors). A total of 3,239 protein groups were identified using Andromeda (Supplementary Table S1), of which 1,095 presented at least two unique peptides and detectable expression in at least 75% of the samples in at least one type of sample (either ER⁺ or TNBC samples; Supplementary Table S2). No one decoy protein passed through these additional filters. Ten samples were excluded from the study; four of them did not yield enough protein to perform the MS analyses, and six did not reach the "mean minus twice the standard deviation" – threshold in the number of unique peptides identified.

**Label-free proteomics analysis**

Label-free quantification was performed using MaxQuant (Supplementary Table S1). Expression values for each protein were analyzed using SAM. We found 100 proteins differentially expressed between ER⁺ and TNBC samples with a FDR < 5% (Supplementary Table S3). Then, we built a hierarchical cluster using the expression values of these 100 proteins (Fig. 1). Gene Ontology analyses from these 100 proteins showed a relevant enrichment in extracellular matrix (ECM) proteins (Supplementary Table S4).

**Orthogonal verification of differences between ER⁺ and TNBC samples**

We used previously published gene expression data as verification of the results obtained using SAM (21, 22). After protein-to-gene ID conversion, 90 out of 100 proteins could be matched by this method. We found that the differences identified at the protein level were confirmed for 74 genes, which showed a significantly different expression in ER⁺ and ER⁻ samples with a FDR < 5%. A group of TNBC samples could not be defined in this verification series because HER2 status was not reported for most samples.

**miRNA expression analysis**

We analyzed the expression of 90 miRNAs in samples from the study with enough available tissue (Supplementary Tables S5 and S6). Expression values for each miRNA were analyzed using SAM. We found 19 miRNAs differentially expressed

| Table 1. Clinical characteristics of the patients included in the study |
|-----------------|----------------|----------------|
| Age at diagnosis (median) | 54.6 (32–83) | 54.2 (32–83) | 61.2 (37–78) |
| Tumor size | | |
| T1 | 33 (31%) | 28 (35%) | 5 (9%) |
| T2 | 61 (58%) | 42 (53%) | 19 (73%) |
| T3 | 10 (9%) | 8 (10%) | 2 (8%) |
| T4 | 1 (1%) | 1 (1%) | 0 (0%) |
| Multifocal | 1 (1%) | 1 (1%) | 0 (0%) |
| Lymph node status | | |
| N0 | 0 (0%) | 0 (0%) | 0 (0%) |
| N1 | 71 (67%) | 54 (68%) | 17 (65%) |
| N2 | 35 (33%) | 26 (32%) | 9 (35%) |
| Chemotherapy | | |
| No anthracyclines | 34 (32%) | 23 (29%) | 11 (42%) |
| Anthracyclines | 63 (59%) | 21 (64%) | 12 (46%) |
| Anthracyclines + taxanes | 9 (9%) | 6 (7%) | 3 (12%) |

NOTE: Clinical criteria are provided according to TNM classification. (http://www.cancer.gov/cancertopics/pdq/treatment/breast/healthprofessional/page3). Tumor grade is the description of a tumor based on how abnormal the tumor cells and the tumor tissue look under a microscope, for determination procedure, see http://www.cancer.gov/cancertopics/factsheet/detection/tumor-grade.)
between ER$^+$ and TNBC samples with an FDR < 5% (Supplementary Table S7). miR-20a, miR-19a, miR-106a, miR-18a, and miR-155b expression were higher in TNBC samples, whereas miR-190b, miR-375, miR-449a, miR-139-5p, miR-205, miR-342, miR-149, miR-625, miR-193b, miR-199a-3p, miR-30a, and miR-30e were more expressed in ER$^+$ samples. Then, we built a hierarchical cluster using the expression values of these 19 miRNAs (Fig. 2). As can be observed, TNBC samples cluster together.

**Systems biology of breast cancers**

Both label-free protein quantification and miRNA expression data were available for 79 tumors (16 TNBC and 63 ER$^+$ tumors). Once we identified proteins and miRNAs differentially expressed between ER$^+$ and TNBC samples, we checked whether any of these proteins had been previously described as a regulated target of any of these miRNAs using the validated targets module of miRWalk. We found that only five proteins were previously identified as targets of six miRNAs. This lack of information prompted us to explore new ways to generate relations, if not between miRNAs and proteins, between miRNAs and biologic processes and pathways.

Protein and miRNA expression data from all samples were used in the probabilistic graphical models analyses, with no other a priori information. Then, the resulting graph was processed (Fig. 3) looking for a functional structure, that is, if the proteins and miRNAs included in each branch of the tree had some relationship regarding their function. Thus, we divided our graph in 12 branches and a core, and performed Gene Ontology analyses. We isolated branches based on the structure of the probabilistic graphical model starting from the outside, with a minimum of 40 proteins included in each branch to allow Gene Ontology analyses in each branch and the core independently. The structure of the probabilistic graphical model had a strong biologic function basis, as 10 out of 12 branches showed a significant enrichment in proteins related with one or two specific biologic functions (Fig. 3). Gene Ontology analyses also showed that the core included proteins related with most of the categories identified in the branches, but no category was over-represented.

The next step was to identify which functions are differentially represented between ER$^+$ and TNBC tumors. To do this, we established the level of expression of each function node using the mean expression of all the proteins included in a given branch.
that belong to a single functional group, building 12 protein functional nodes (Fig. 3 and Supplementary Table S8). Then, we performed class comparison analyses to assess which functional nodes are differentially activated between ER$^+$ and TNBC samples. Seven nodes showed significant differences between both groups. Metabolism B, Proliferation, Protein Synthesis, and Mitochondria B functional nodes showed increased activity in TNBC samples, whereas Mitochondria A, Metabolism A, and ECM nodes showed increased activity in ER$^+$ tumors (Fig. 4 and Supplementary Table S8).

Then, we compared the proteins identified as differential between SAM and the probabilistic graphical model functional approach. Only 25 proteins from the SAM analysis were also in the 239 proteins from the seven functional nodes differentially expressed between ER$^+$ and TNBCs. Ultimately, we performed an orthogonal verification of the activity of these functional nodes in our validation dataset. Six out of seven nodes showed a significant difference of activation between ER$^+$ and ER$^-$ samples.

Regarding the miRNAs, our first observation was that most miRNAs were included in two big miRNA groups, meaning that relations at the expression level are stronger between these miRNAs themselves than the relation between miRNAs and proteins. One of these groups contains every measured miRNA from the human mir-17-92 cluster, whereas the other contains...
Figure 4.
Probabilistic graphical model showing median expression values of proteins and miRNAs in ER⁺ (A) and TNBC (B) samples. Red, high expression; green, low expression. The magnified area corresponds to the metabolism B node in ER⁺ (C) and TNBC (D). Big circles correspond to miRNAs.

more that 50% of all the miRNAs analyzed in this study (Fig. 3). The most outstanding observation was that 15 miRNAs were included in branches composed mainly by proteins (Supplementary Table S8).

**Function: proliferation**

We found two different functions mixed in the same branch: (i) proliferation and (ii) ECM, and two miRNAs embedded within: miR-301a and miR-331-5p. Thirteen proteins related with regulation of cell proliferation belong to the proliferation node. STAT1 is more expressed in TNBC samples than in ER⁺ ones, whereas PURA expression is decreased in TNBC samples accordingly with SAM analyses. Measurement of the functional activity of this node showed a higher proliferation activity in our TNBC sample series (P < 0.01), as well as in the validation cohort (P < 0.001).

**Function: extracellular matrix**

We also found 39 proteins related with ECM and focal adhesion. Eight of the components of this functional node were also found as differential between ER⁺ and TNBC samples in SAM analysis. All of them (Decorin, VCAN, Mimecan, Biglycan, COMP, Lumican, Prolargin, and Asporin) showed lower expression in TNBC samples. Thus, it is not surprising that the ECM functional node activity highly differs between ER⁺ and TNBC samples, both in the proteomics analysis (P < 0.01) and in the orthogonal validation (P < 0.05).

**Function: metabolism**

We found two different branches related with metabolism, thus we evaluated these two functional metabolism nodes independently. Metabolism A functional node included 33 proteins. Among them, seven were identified as differentially expressed by SAM analyses. PHGDH, LDHB, and P5CS (encoded by the ALDH1A1 gene) are more expressed in TNBC samples than in ER⁺, whereas DHPR, FBP1, and MCCB are decreased in TNBC samples. The functional activity of this node was higher in ER⁺ samples (P < 0.01) and was also significant in the orthogonal validation (P < 0.001). Three miRNAs were included in this functional node: miR-135b, miR-190b, and miR-224. Differential expression analysis showed that miR-135b expression was higher in TNBC samples, whereas miR-190b was less expressed in TNBC samples. Metabolism B includes proteins related with glutamine metabolism, such as PC5S, IDH1, and GLULD1. GLULD1 protein expression is slightly lower in TNBC tumors (fold change = 0.66, P < 0.05).

Metabolism B functional node comprised 13 proteins, including GAPDH, PGK1, HSP90, LDHA, and pyruvate kinase, among others. When we measured the functional activity of this node, it appeared increased in TNBC both in our proteomics data (P < 0.01) as in the orthogonal validation (P < 0.001). However, only one protein (UPGA) was included among the 100 proteins identified as differentially expressed between ER⁺ and TNBC. Only one miRNA (miR-449a) was included in this functional node, and it was also detected by SAM as downregulated in TNBC samples. Therefore, we identified a functional node with increased activity and a miRNA related to it with decreased expression (Fig. 4), showing a negative correlation (r = −0.45; P < 0.001; r² = 0.20; Supplementary Fig. S1). Surprisingly, miR-449a expression was not significantly correlated to any of the 13 proteins included in this functional node.

**Function: mitochondrial**

We found two different branches related with mitochondrial activity, thus we evaluated these functional mitochondrial nodes independently. Mitochondria A functional node included 37 proteins. Among them, four were identified as differentially expressed by SAM analyses. CYBB is more expressed in TNBC than in ER⁺, whereas IVD, DCXR, and HSD17B8 expression is decreased in TNBC samples. The functional activity of this node was higher in our ER⁺ samples (P < 0.05), and was also confirmed by the orthogonal validation (P < 0.001). Two miRNAs (miR-23b and miR-375) were included in this functional node. miR-375 expression was also found increased in
Metabolic analyses in breast cancer cell lines

We measured both the glucose consumption and lactate production in five breast cancer cell lines: two ER\(^+\) cell lines (MCF7 and T47D) and three TNBC cell lines (MDA-MB-231, BT-20, and MDA-MB-468). TNBC cell lines showed higher glucose consumption than ER\(^+\) cell lines (Fig. 5A). Lactate production was also higher in TNBC cell lines compared with ER\(^+\) ones (Fig. 5B).

Discussion

In this work, we demonstrate that it is technically feasible to analyze up to a hundred of clinical samples by high-throughput proteomics, which could help to place MS at the same level than gene microarrays, or massively parallel sequencing technologies, in its capability to analyze enough samples and features to generate clinically useful molecular profiles. Ongoing studies are trying to complete the molecular characterization of breast cancers (23). One key aspect of these studies is that they will analyze hundreds of samples to extract meaningful information, both from the medical and biologic point of view.

Our proteomics pipeline allowed us to detect 3,239 protein groups in FFPE breast tissues. We then applied stringent filters to avoid false detections and used only 1,095 proteins in subsequent analyses. It is remarkable that our label-free quantification approach can detect clinically relevant proteins such as HER2, STAT1, or PARP1. However, other clinically important markers, such as ER, PR, or Ki67, were not detected using this technique. One obvious limitation of proteomics when compared with genomics is that genomics can measure the expression of all known genes in the same experiment, whereas proteomics only provides a measurement of peptides that are both detected and identified. This is directly related to the wide dynamic range of protein expression, which means that the amount of some proteins can be orders of magnitude higher than others. Unfortunately, ER and PR are among these proteins showing low relative abundance in cells. We compared the abundance of ESR1, PGR, and MKI67 with some proteins identified in our MS experiments using available data from PaxDB (24). ESR1 and PGR abundance is 10 times lower than the less abundant protein identified in our MS experiments. Ki67 abundance is comparable with other detected proteins. We did not detect/identify any single peptide from this protein in our MS runs, thus with our experimental conditions, Ki67 cannot be detected in FFPE breast cancer samples. These limitations could be overcome soon, as the latest developments in the field are promising.

Label-free proteomics analysis

SAM analysis identified 100 proteins differentially expressed between ER\(^+\) and TNBC samples. Some of these proteins with higher expression in TNBC samples have been previously described as overexpressed in these tumors, such as PHGDH (25), LDHB (26), S100A8, S100A9 (27), and GRP78 (28), whereas AGR2, AGR3 (29), PIP (30), and GPD1 (31), have been shown to be upregulated in ER\(^+\) breast cancer tumors in concordance with our data. These results are strained by sample size. For this reason, proteomic-based quantitation of potential biomarkers requires further validation using orthogonal techniques. It has been already demonstrated that mRNA levels largely reflect the respective protein levels (32). Consequently, the intersection between proteomic datasets and other genome-wide datasets often allows robust cross-validation. We used independent gene expression data from public datasets to corroborate our findings. It is remarkable that 74% of the differences observed in our proteomics data were confirmed using gene expression data, and are also consistent with previous reports.

miRNA expression analysis

Regarding miRNA differential expression analysis, miR-106a, miR-18a, and miR-135b expression were higher in TNBC samples, whereas miR-190b, miR-375, miR-205, miR-342, miR-125b, and miR-214 were more expressed in ER\(^+\) samples. This is consistent with previously published data. miR-342 expression has been beforehand described as significantly higher in ER\(^+\) samples, and was included in a predictor of ER status together with miR-135b and miR-190 (33). Moreover, miR-342 expression was positively correlated with ER\(\alpha\) mRNA expression in human breast cancer (34). miR-205 was significantly upexpressed in triple-negative primary breast cancers
when compared with tumor-associated normal samples. Both miR-205 and miR-342 have been proposed as potential biomarkers for diagnosis of TNBC (35). It has been previously shown that miR-125b expression is lower in TNBC tumor tissue when compared with adjacent normal tissue, and has been related with chemoresistance (36). Finally, miR-345 expression was downregulated in ER− breast cancer patients, whereas miR-18a was increased (37).

### Systems biology of breast cancers

In this work, we have defined functional nodes and measured their activity, and also performed a comprehensive investigation of miRNA-mediated regulation through an integrative analysis of the variation of miRNA and protein expression in a large collection of breast cancer samples using a probabilistic graphical model approach. The relation among multiple types of molecules defines how they are regulated. Computationally linking and then analyzing these relations at the global scale could reveal entirely new levels of cell regulation and provide insights into damaged molecular pathways in disease (38).

Integrated pathway analysis is expected to increase the precision and sensitivity of causal interpretations for large sets of observations because no single data source is likely to provide a complete picture by itself (39).

This analysis identified 12 functional nodes, 7 of which showed differential activation after a class comparison analysis when comparing ER− and TNBC samples. Orthogonal validation using external gene expression datasets confirmed these differences in six of them: ECM, proliferating, metabolisms A and B, and mitochondria A and B. Proliferation and ECM were grouped in the same branch, despite that activity analyses showed that proliferation activity is increased in TNBC, whereas ECM is decreased. It is well known that TNBCs are highly proliferative tumors with high repopulation rate, which is in concordance with our results.

Probabilistic graphical model analyses showed that two miRNAs, miR-301a and miR-331-5p, are related with these functional nodes. It has been described that miR-301a promotes breast cancer proliferation, invasion, and tumor growth (40). It has also been described that miR-331 regulates SOCS1 (41), which acts as a negative feedback regulator of STAT (42). ECM node includes a number of proteins related with focal adhesion. There is a previous work showing differences in focal adhesion proteins between ER+ and ER− tumors measured using proteomics (43).

Interestingly, the activity of all metabolism and mitochondria related nodes have a differential activity between ER+ and TNBC samples. Metabolism B and mitochondria B are increased in TNBC samples as compared with ER+ samples, whereas metabolism A and mitochondria A activity are decreased. On the other hand, high expression of LHDB has been associated with glycolytic phenotype (26), whereas it has been demonstrated that loss of FBPI induces glycolysis and results in increased glucose uptake and macromolecule biosynthesis (44). Cancer cells reprogram their metabolism in order to satisfy their bioenergetics and biosynthetic requirements. The hallmark of tumor metabolism is aerobic glycolysis, or Warburg effect, which was first described more than 80 years ago (45). During recent years, interest in developing cancer metabolism inhibitors and the possibility of combining them with existing therapeutic approaches has increased dramatically, with academic and pharmaceutical groups actively pursuing this aspect of tumor physiology that could have profound anticancer effects and fundamentally change the treatment of cancer (46).

All these results together (glycolysis and the TCA cycle showing differential activity, higher LDHB expression and lower FBPI expression in TNBC) suggest that TNBC samples are highly glycolytic tumors, and confirm that TNBC have unique metabolic profiles that may be exploited for therapeutic intervention (47). To support this hypothesis, we evaluated this differential metabolic profile in five breast cancer cell lines. TNBC cell lines showed higher glucose consumption and lactate production, as has been showed previously (26, 44, 47).

Metabolism A node includes some proteins related with glutamine metabolism. Increased glutaminolysis has been described as a key feature of the metabolic changes in cancer cells compared with normal cells (48). Recent studies suggest differences in the metabolism of glutamine between ER+ and ER− disease. It has been observed that ER− tumors have a shift in the equilibrium between glutamate and glutamine toward glutamate in the ER− disease (49, 50). We found that GLUD1 expression is slightly lower in TNBC tumors, which is consistent with previous reports describing that TNBCs have the lowest GLUD1 protein expression rate of all molecular subtypes, whereas HER2-positive tumors show the highest (51). We also detected a higher protein expression of P5CS in TNBC tumors. P5SC is a mitochondrial enzyme that catalyzes the conversion of L-glutamate to P5C, a pivotal step in the biosynthesis of nonessential amino acids such as L-proline, L-ornithine, and L-arginine (52).

Although aerobic proliferating cells use glucose and glutamine for biomass production through the TCA cycle, hypoxic cells shunt glucose to lactate and rewire glutamine metabolism (53). GLUD1 and other glutamate-consuming enzymes activity is stimulated under glucose deprivation conditions, while its activity is repressed during robust glycolysis (54). Glutamine can be used to drive the TCA cycle independently of glucose, contribute to lipid synthesis via IDH1, and be donor for the synthesis of nonessential amino acids—via transaminases and P5CS—and nucleotides (53, 55). Further exploration of glutamine metabolic rewiring differences in subgroups of breast cancer is needed to validate our preliminary results.

It is remarkable that only one protein included in the Metabolism B node (UPGA) was included among the 100 proteins identified as differentially expressed between ER+ and TNBC. This result suggests that functional approaches can highlight differences in biologic processes and pathways that cannot be unraveled through the analysis of individual genes or proteins. In this regard, both approaches complement each other.

One of our main objectives was to find relations between miRNA and differential processes in two breast cancer subtypes. We observed that miR-449a expression was lower in TNBC samples, and our probabilistic graphical model analysis include it in the metabolism B node, showing a negative correlation with its activity, which is increased in TNBC samples. Surprisingly, no correlation between miR-449a expression and individual proteins from the node was found. Our results suggest that miR-449a absence is related with and increased glycolytic metabolism, and that miR449a exerts its regulatory effect over cellular metabolism indirectly, through regulation of yet unknown genes (Supplementary Fig. S2). Low expression level of miR-449a was correlated with poor prognosis of lung cancer patients (56), but there are no studies about its
role in breast cancer progression or its implication in the regulation of the metabolism. miR-449a deserves additional functional studies to investigate its relationship with metabolic rewiring in breast cancers, which is far beyond the scope of this work.

High-throughput technologies are providing a comprehensive view of the molecular changes in cancer tissues. MS-based proteomics together with miRNA profiling allowed us to overview differential processes in ER+ and TNBC samples. However, there are still a number of improvements required before MS-based proteomics could be considered at the same level than genomics-related technologies in terms of comprehensiveness and depth: substantial advances in speed, sensitivity, reproducibility, better identification methods, and standardization of protocols are still necessary.

In conclusion, the integration of different levels of biologic information can improve our knowledge of molecular processes in cancer cells. In this study, we measured the expression of 1,095 proteins and 90 miRNAs in breast cancer samples, unraveling differences in the activity of molecular processes, such as proliferation and metabolism, between ER+ and TNBC tumors that were also confirmed at the transcriptome level. Moreover, some miRNAs were shown to be related with these functional categories. These findings may provide a rational to design new treatments for TNBC.

Disclosure of Potential Conflicts of Interest

A. Gámez-Pozo and J. Á. Fresno Vara have ownership interest (including patents) in Biomedical Molecular Medicine S.L. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: A. Gámez-Pozo, J. Grossmann, E. Espinosa, E. Ciruelos, J.Á. Fresno Vara


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Gámez-Pozo, P. Nanni, C.A. Castaneda, E. Espinosa, E. Ciruelos

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Gámez-Pozo, J. Berges-Soria, J.M. Arevalillo, H. Navarro, J. Grossmann, P. Main, M. Diaz Almirón, E. Espinosa, E. Ciruelos, J.Á. Fresno Vara

Writing, review, and/or revision of the manuscript: A. Gámez-Pozo, P. Nanni, J. Grossmann, E. Espinosa, E. Ciruelos, J.Á. Fresno Vara

Study supervision: J.M. Arevalillo, E. Espinosa, E. Ciruelos, J.Á. Fresno Vara

Acknowledgments

The authors particularly acknowledge the patients in this study for their participation and to IdiPAZ and I+12 Biobanks for the generous gifts of clinical samples used in this study. The authors also thank Dr. Rubén Gómez Rioja for his help with glucose and lactate measurements.

Grant Support

This study was financially supported by grants PS09/01597 and PI12/00444 from Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry, Spain. This work was also supported by the PRIME-XX project, grant agreement 262067, funded by the European Union 7th Framework Programme. A. Gámez-Pozo and R. López-Vacas are supported by Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry grants CA12/00258 and CA12/00264, respectively. IdiPAZ and I+12 Biobanks are supported by Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry (RD09/0076/00073 and RD09/0076/00118, respectively) and Farmaindustria, through the Cooperation Program in Clinical and Translational Research of the Community of Madrid.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 3, 2014; revised March 3, 2015; accepted March 12, 2015; published OnlineFirst April 16, 2015.


Combined Label-Free Quantitative Proteomics and microRNA Expression Analysis of Breast Cancer Unravel Molecular Differences with Clinical Implications

Angelo Gámez-Pozo, Julia Berges-Soria, Jorge M. Arevalillo, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-14-1937

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2015/04/17/0008-5472.CAN-14-1937.DC1

Cited articles
This article cites 53 articles, 12 of which you can access for free at:
http://cancerres.aacrjournals.org/content/75/11/2243.full#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/75/11/2243.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/75/11/2243.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.