

A New Gene Expression Signature for Triple-Negative Breast Cancer Using Frozen Fresh Tissue before Neoadjuvant Chemotherapy

Sandra K Santuario-Facio,^{1,7*} Servando Cardona-Huerta,^{2*} Yadira X Perez-Paramo,³ Victor Trevino,⁴ Francisco Hernandez-Cabrera,⁵ Augusto Rojas-Martinez,^{1,7} Grecia Uscanga-Perales,⁶ Jorge L Martinez-Rodriguez,^{2,8} Lizeth Martinez-Jacobo,⁶ Gerardo Padilla-Rivas,⁶ Gerardo Muñoz-Maldonado,⁹ Juan Francisco Gonzalez-Guerrero,¹⁰ Javier Valero-Gomez,² Ana L Vazquez-Guerrero,² Herminia G Martinez-Rodriguez,⁶ Alvaro Barboza-Quintana,² Oralia Barboza-Quintana,¹¹ Raquel Garza-Guajardo,¹¹ and Rocio Ortiz-Lopez^{1,7}

¹Universidad Autonoma de Nuevo Leon, Centro de Investigación y Desarrollo en Ciencias de la Salud, Monterrey, Nuevo Leon, Mexico; ²Tecnologico de Monterrey, Hospital San Jose, Centro de Cancer de Mama, Monterrey, Nuevo Leon, Mexico; ³Pharmaceutical Sciences Department, College of Pharmacy, Washington State University, Spokane, Washington, United States of America; ⁴Tecnológico de Monterrey, Escuela Nacional de Medicina, Grupo de Investigacion en Bioinformatica, Monterrey, Nuevo Leon, Mexico; ⁵Universidad Autonoma de Nuevo Leon, Facultad de Ciencias Fisico Matematicas, Centro Investigación en Ciencias Físico Matemáticas, Monterrey, Nuevo Leon, Mexico; ⁶Universidad Autonoma de Nuevo Leon, Facultad de Medicina, Departamento de Bioquímica y Medicina Molecular, Monterrey, Nuevo Leon, Mexico; ⁷Tecnologico de Monterrey, Escuela Nacional de Medicina, Monterrey, Nuevo Leon, Mexico; ⁸Universidad Autonoma de Nuevo Leon, Hospital Universitario Dr. Jose Eleuterio Gonzalez, Centro Universitario Contra el Cancer, Servicio de Oncologia, Monterrey, Nuevo Leon, Mexico; ⁹Universidad Autonoma de Nuevo Leon, Hospital Universitario Dr. Jose Eleuterio Gonzalez, Servicio de Cirugia General, Monterrey, Nuevo Leon, Mexico; ¹⁰Universidad Autonoma de Nuevo Leon, Hospital Universitario Dr. Jose Eleuterio Gonzalez, Centro Universitario Contra el Cancer, Servicio de Oncologia, Monterrey, Nuevo Leon, Mexico; and ¹¹Universidad Autonoma de Nuevo Leon, Hospital Universitario Dr. Jose Eleuterio Gonzalez, Servicio de Anatomia Patologica y Citopatologia, Monterrey, Nuevo Leon, Mexico

Triple-negative breast cancer (TNBC) is an aggressive subtype of breast cancer tumors. Comparisons between TNBC and non-triple-negative breast cancer (nTNBC) may help to differentiate key components involved in TNBC neoplasms. The purpose of the study was to analyze the expression profile of TNBC versus nTNBC tumors in a homogeneous population from northeastern Mexico. A prospective study of 50 patients (25 TNBC and 25 nTNBC) was conducted. Clinic parameters were equally distributed for TNBC and nTNBC: age at diagnosis (51 versus 47 years, $p = 0.1$), glucose level (107 mg/dl versus 104 mg/dl, $p = 0.64$), and body mass index (28 versus 29, $p = 0.14$). Core biopsies were collected for histopathological diagnosis and gene expression analysis. Total RNA was isolated and expression profiling was performed. Forty genes showed differential expression pattern in TNBC tumors. Among these, nine overexpressed genes (*PRKX/PRKY*, *UGT8*, *HMG A1*, *LPIN1*, *HAPLN3*, *FAM171A1*, *BCL141A*, *FOXC1*, and *ANKRD11*), and one underexpressed gene (*ANX9*) are involved in general metabolism. Based on this biochemical peculiarity and the overexpression of *BCL11A* and *FOXC1* (involved in tumor growth and metastasis, respectively), we validated by quantitative polymerase chain reaction the expression profiles of seven genes out of the signature. In this report, a new gene signature for TNBC is proposed. To our knowledge, this is the first TNBC signature that describes genes involved in general metabolism. The findings may be pertinent for Mexican patients and require evaluation in other ethnic groups and populations.

Online address: <http://www.molmed.org>

doi: 10.2119/molmed.2016.00257

*SKS-F and SC-H should be considered as co-first authors, as both contributed equally to this manuscript.

Address correspondence to Rocio Ortiz-Lopez, Tecnológico de Monterrey, Escuela Nacional de Medicina. CITES 3er Piso. Av. I. Morones Prieto 3000 Poniente. Los Doctores. 64710 Monterrey, N.L. MEXICO. Telephone: + 52 81 8888 2000. Email: rortizl@itesm.mx
Submitted December 17, 2016; Accepted for Publication April 17, 2017; Published Online (www.molmed.org) May 4, 2017.

INTRODUCTION

Triple-negative breast cancer (TNBC) is a breast cancer (BC) subtype characterized by aggressive behavior (1,2), a characteristic immunohistochemistry (IHC) with negative pattern for estrogen (ER) and progesterone (PR) receptors and human epidermal growth factor receptor 2 (HER2). Since TNBC lacks targeted therapies, treatment options are limited to nonselected chemotherapy (3). These tumors represent 12–17% of all BCs, and predominantly affect premenopausal women (<50 years). It is recognized that TNBC is more frequent in African American and Hispanic women belonging to low-income groups with increased hormonal exposure and obesity (4,5). These characteristics merge as potential etiologic factors influencing the tumor's nature.

Mexico is a country with a high prevalence of obesity, where women 40–59 years old have an average body mass index (BMI) >30 kg/m² (6). The BC rate in Mexico is also increasing, becoming the primary cause of diagnosed cancer nowadays (7,8). The association between obesity and the risk of BC in postmenopausal women has been consistently observed in epidemiological studies (9,10). Regardless, most studies of premenopausal women have not found a similar relationship between BC and obesity. The prognosis for obese pre- and postmenopausal BC patients is substantially worse. Increasing evidence suggests that such associations may be related to sex hormones, insulin, and certain adipokines (10,11). Most of the studies have evaluated European-American Caucasian women. Investigations in African American and Hispanic women with BC are particularly important due to higher TNBC prevalence and health disparities in these groups.

General BC gene expression profiles have been useful to investigate tumor profiles and to define prognosis and therapy. These studies have discovered genes associated with different tumor subtypes and their outcomes (12,13). Several gene expression profile analyses

have also been proposed for TNBC to identify biomarkers and targets for personalized therapy (2,14–17). Despite some progress, large heterogeneity in the gene profiles has been noticed in studies for TNBC, and there is a lack of a useful gene signature for this tumor subtype (1,14,17–21).

The purpose of the study was to define a distinctive gene signature for TNBC using a homogeneous population from northeastern Mexico, which will be useful in characterizing biological pathways involved in TNBC. To our knowledge, this is the first report describing a differential gene signature for TNBC.

MATERIALS AND METHODS

Patient Population and Tumor Tissue Samples

The study was approved by the Ethics Committee of the Universidad Autonoma de Nuevo Leon (No. BI11-005), in accordance with the Declaration of Helsinki. Informed written consent was obtained from all patients participating in this study. A total of 149 biopsy samples were collected prospectively (from July 2011 to October 2014), with suspicious diagnosis (clinical, radiological, or both) of BC (tumor >2 cm, palpable ipsilateral lymph nodes, or ulceration). Out of 149 collected biopsies, 29 were pathologically diagnosed as TNBC (19.5%). In all, 27 TNBC samples were selected for inclusion in this study based on whole RNA quantity and quality. An additional 27 samples diagnosed as nTNBC were paired to the TNBC samples per clinical stage and were used in the expression profile analysis comparing TNBC versus nTNBC. The study design is illustrated in Figure 1.

Tumor Sample Collection and RNA Isolation

Ultrasound-guided core biopsies were obtained using a 12 Fr gauge biopsy needle (Bard®). Six to eight core samples were obtained from each patient. Four core samples were used for histopathological analysis, including hematoxylin

and eosin and IHC (PR/ER/HER2, Ki67, and p63). Two or three core samples were preserved in RNAlater solution for gene expression analysis (microarrays and quantitative polymerase chain reaction [qPCR] validation).

RNA extraction from biopsies was done using RNeasy Fibrous Tissue Mini Kit (Qiagen, Germantown, MD, USA), which allowed for obtaining high-quality RNA in small tissue biopsies. RNA quality was assessed by capillary electrophoresis using the Experion™ Automated Electrophoresis Station (Bio-Rad, Hercules, CA, USA). RNA concentration was determined by NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

Microarrays Analysis (Expression Profiles)

From the selected RNA, sample processing, microarray hybridization, and gene expression analysis were conducted using the GeneChip 3' IVT Express Kit (Affymetrix, Santa Clara, CA, USA). The hybridization mixture was prepared and applied to the GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix), measuring >43,000 transcripts representing >20,000 human genes. Washing and scanning processes were realized in the Fluidics Station 400 and GeneChip Scanner 3000 7G, respectively, and preliminary data analysis was completed using Microarray Suite software version 5.0.0.032.

Microarray Data Processing

Normalization was performed using robust microarray analysis and quantile normalization. Four samples showed poor quality in microarray data and were removed from the analysis. The analysis was performed using 50 patients. Probes whose mean expression (log scale) was <4 (in logarithm scale resulting from robust microarray analysis) were removed. Molecular classification was done using the PAM50 algorithm (22).

Gene expression signature between 25 TNBC and 25 nTNBC was performed using *t* test and Kolmogorov test, with a

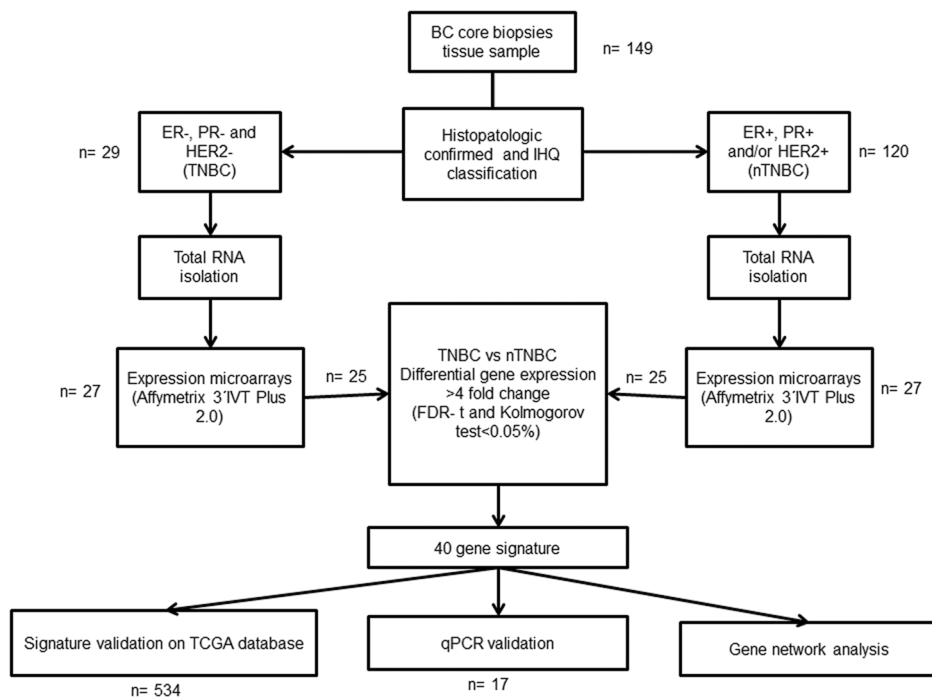


Figure 1. Study design. This diagram shows sampling and workflow of experiments and analysis of data.

multiple comparison correction using the false discovery rate (FDR) method (23). We considered as positive those probes having significant p values in both tests (FDR t and Kolmogorov test $<0.05\%$). These analyses were performed in R software package (24).

Signature Validation on the Cancer Genome Atlas Database

The Cancer Genome Atlas Database (TCGA) (<http://cancergenome.nih.gov/>), located at SurvExpress (25), contains clinical and genetic information on 17,814 genes in 534 tissue samples. This database was used to compare TNBC versus nTNBC expression profiles and to evaluate the matching with the 40 gene expression signatures found in our study. Statistical analysis was replicated in TCGA to compare profiles between TNBC and nTNBC patients from this study. All TCGA data and figures were accessed, analyzed, and generated using R language. All TCGA data included in this manuscript agree with the TCGA publication guidelines.

Gene Network

An analysis of gene network was generated using GeneMANIA (<http://www.genemania.org/>) (26) to find a relationship between genes of our signature. GeneMANIA used a very large set of functional association data, including protein and genetic interactions, pathways, coexpression, colocalization, and protein domain similarity.

Real-Time qPCR Validation

To validate microarray data, seven genes were selected: six overexpressed (*HMGAI*, *ANKRD11*, *BCL11A*, *FOXC1*, *LPIN1*, and *UGT8*) and one underexpressed (*ANXA9*). Expression analysis of these genes, including one endogenous control (*GRAMD1A*), was assessed using predesigned hydrolysis probes (*HMGAI*: Hs.PT.58.38699366; *FOXC1*: Hs00559473_s1; *LPIN1*: Hs00299515_m1; *UGT8*: Hs00409961_m1; *ANKRD1*: Hs00946580_m1; *BCL11A*: Hs01093197_m1; *ANXA9*: Hs.PT.58.1414783; and *GRAMD1A*: Hs.PT.5840681431) (Thermo-Fisher Scientific and IDT). Total RNA samples used for microarrays assays were analyzed by

qPCR using the Lightcycler 480 instrument (Roche). Ct mean for each gene was used for dCt (problem minus endogenous). The 2^{-dCt} analysis was done using calculated dCt for all genes. To compare gene expression of TNBC versus nTNBC groups, the relative expression 2^{-dCt} was evaluated from qPCR data of all genes (*HMGAI*, *ANKRD11*, *BCL11A*, *FOXC1*, *LPIN1*, *UGT8*, and *ANXA9*) after normalization with *GRAMD1A*. Two sample Kolmogorov-Smirnov tests were used in both datasets, since the number of samples was distinct in each group and in unknown distributions. Moreover, the Kolmogorov-Smirnov test is sensitive to shape, position, and dispersion behavior of distributions to determine whether two independent samples belong to the same group (homogeneity test), or whether they have one statistical significant difference. We used $p < 0.05$ to establish this difference. The results of expression values are represented in box plots for microarray and qPCR data. OriginPro software (version 9.2, OriginLab, Northampton, MA, USA) was used for statistical data analysis.

All supplementary materials are available online at www.molmed.org.

RESULTS

Patients

A total of 149 core biopsies were prospectively collected. Out of 149, 29 tumors were diagnosed as TNBC (19.5%). From the 29 TNBC samples, RNA quality for microarray was obtained in 27 biopsies. Paired nTNBC samples were selected according to clinical stage. Clinicopathological characteristics were similar in TNBC versus nTNBC patients (Table 1). Age at diagnosis, high BMI (obese or overweight), and glucose levels (>100 mg/dl) were similar in both groups.

Molecular Classification Using PAM50

According to immunohistochemistry (IHQ) analysis, out of 25 TNBC patient samples, 18 (72%) were basal-like subtype, 5 (20%) were normal-like subtype, and

Table 1. Patient and tumor characteristics.

	All patients, n = 54		TNBC, n = 25		nTNBC, n = 25		P value
Age at diagnosis (years)	49	26–71	51	39–71	47	26–61	0.1 ^e
Body mass index (kg/m ²)	28.7	20.7–40.1	27.58	20.78–38.01	29.31	21.83 a 39.65	0.14 ^e
<25	9	17%	4	16%	4	16%	0.67 ^f
≥25	38	70%	17	68%	19	76%	0.75 ^f
Menopause status							
Pre	27	50%	12	48%	13	52%	1 ^f
Post	27	50%	13	52%	12	48%	1 ^f
Background cancer ^a							
Yes	26	48%	14	56%	10	40%	0.4 ^f
No	28	52%	11	44%	15	60%	0.4 ^f
Diabetes mellitus ^b							
Yes	6	11%	5	20%	1	4%	0.2 ^f
No	48	89%	20	80%	24	96%	0.2 ^f
Glucose levels (mg/dl)	104.84	82–193	106.62	88–183	103.67	82–193	0.64 ^e
Number of children ^c							
Nulliparous	4	7%	1	4%	3	12%	0.6 ^f
1–2	16	30%	8	32%	7	28%	1 ^f
>3	33	61%	16	64%	15	60%	1
Lactation ^d							
Yes	22	41%	10	40%	11	44%	1 ^f
No	15	28%	8	32%	6	25%	0.75 ^f
Smoking							
Yes	8	15%	2	8%	5	20%	0.42 ^f
No	45	83%	22	88%	20	80%	0.7 ^f
TNM							
T1	1	2%	0	0%	1	4%	1 ^f
T2	25	46%	12	48%	11	44%	1 ^f
T3	16	30%	8	32%	6	25%	0.75 ^f
T4	11	20%	4	16%	7	28%	0.75 ^f
N0	8	15%	3	12%	5	20%	0.5 ^f
N1	32	59%	14	56%	16	64%	0.77 ^f
N2	12	22%	6	24%	4	16%	0.72 ^f
N3	2	4%	2	8%	0	0%	0.47 ^f
M0	54	100%	25	100%	25	100%	1 ^f
Clinical stage							
I	1	2%	0	0%	1	2%	1 ^f
II	25	46%	12	48%	12	48%	0.78 ^f
III	28	52%	13	52%	12	48%	1 ^f

^aAt least one relative (first, second, or third grade).

^bConfirmed at BC diagnosis.

^cNumber of pregnancies.

^dAt least 6 months.

^eT test, confidence level at 95%.

^fZ test, confidence level at 95%.

2 (8%) were HER2 subtype, based on molecular subtypes (Figure 2).

TNBC Expression Profile

To identify a TNBC signature, 27 TNBC and 27 nTNBC biopsy expression profiles

were compared. After normalization of the microarray expression performance, 25 samples from each group met the adequate sample quality for data analysis.

A supervised classification was applied based on TNBC versus nTNBC according

to IHC status. Data analysis for 50 patients showed 49 differentially expressed probes (Supplementary Table S1) corresponding to 40 gene signatures with FDR < 0.0005. A heat map for these 40 genes was generated, representing their relative

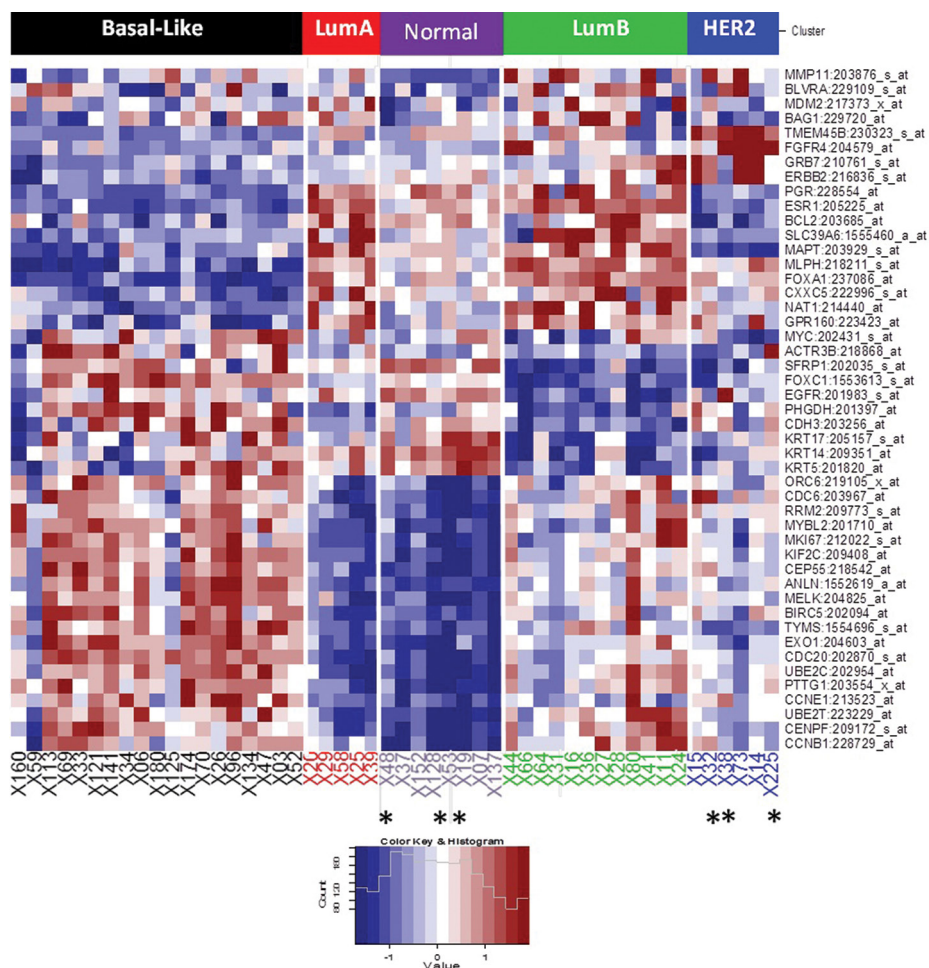


Figure 2. Molecular classification using PAM50 algorithm. Heat map from 50 samples (TNBC and nTNBC). Five different molecular subtypes are at the top of the heat map. Blue represents low gene expression, while red represents high gene expression. In the heat map, each column is for a different patient and each row is for a gene. Asterisk (*) represents TNBC samples according to IHC analysis that are not in the basal-like subtype.

expression values (Z score = -1.5 to 1.5) (Figure 3A). The upper left side shows the group of nine overexpressed genes (*FOXCl*, *PRKX/PRKY*, *UGT8*, *BCL11A*, *HMGA1*, *LPIN1*, *FAM171A1*, *HAPLN3*, and *ANKRD11*). Conversely, there are 31 underexpressed genes in TNBC at the bottom left side. The center of the heat map shows a not quite defined area shared by four TNBC and five nTNBC patients with an undifferentiated expression pattern.

External Database Validation

To compare our results against other databases, a differential expression

analysis of TNBC versus nTNBC tumors was done using the TCGA database. Sample data from this database were selected and classified as TNBC or nTNBC (using expression data ER, PR, and HER2). The analysis was performed using two approaches. In the first one, we used the TCGA gene expression database and compared TNBC versus nTNBC. We obtained 87 gene signatures with high significance ($q < 1e-40$) (Supplementary Figure S1), among which only nine (*ERBB4*, *DNAJC12*, *TBC1D9*, *MLPH*, *ANXA9*, *NAT1*, *BCAS1*, *TFF1*, and *BCL11A*) overlapped our signature

(Figure 3B). The TCGA gene expression database contains only 33 genes from the 40 genes in our signature. As a second approach, we compared the differential expression of TNBC versus nTNBC using only these 33 genes. This second analysis showed that the genes had similar behavior (overexpressed or underexpressed) that they had in our signature (Supplementary Figure S2), with a smaller statistical significance ($p < 1e-8$).

Gene Network

Figure 4 shows the networking of the 40 genes included in our TNBC signature. In this gene network, the main functions are related to chromatin (red) or phosphatidylcholine (blue) biosynthetic process (26). Table 2 shows the main functions of the networked processes including the genes of the signature. In this, the biosynthetic and metabolic processes concerning phosphatidylcholine have the lowest FDR ($1.83e-4$ and $8.71e-3$).

Real-Time qPCR Validation

Validation by qPCR was performed in 35 of the 50 samples analyzed by microarray (17 nTNBC and 17 TNBC) due to remaining amounts of RNA. A comparison between microarray expression and 2^{-dCt} qPCR levels (after normalization with *GRAMD1A* gene) demonstrates similar expression values for all genes (*HMGA1*, *ANKRD11*, *BCL11A*, *FOXCl*, *LPIN1*, *UGT8*, and *ANXA9*) (Figure 5). qPCR expression levels (A) are higher than those observed in the microarray expression analysis ($p < 0.05$) (B). Remarkable overexpression of *FOXCl*, *UGT8*, and *BCL11A* in TNBC is observed by qPCR, which is followed by expression levels of *LPIN1*, *ANKRD11*, *ANXA9*, and *HMGA1*. This clearly differs with the gene expression levels displayed by nTNBC samples. *GRAMD1A* was used as a normalizing gene, based on the low variability observed in all samples tested by the microarray analysis. *GAPDH* and *BACTIN* showed high variability among samples.

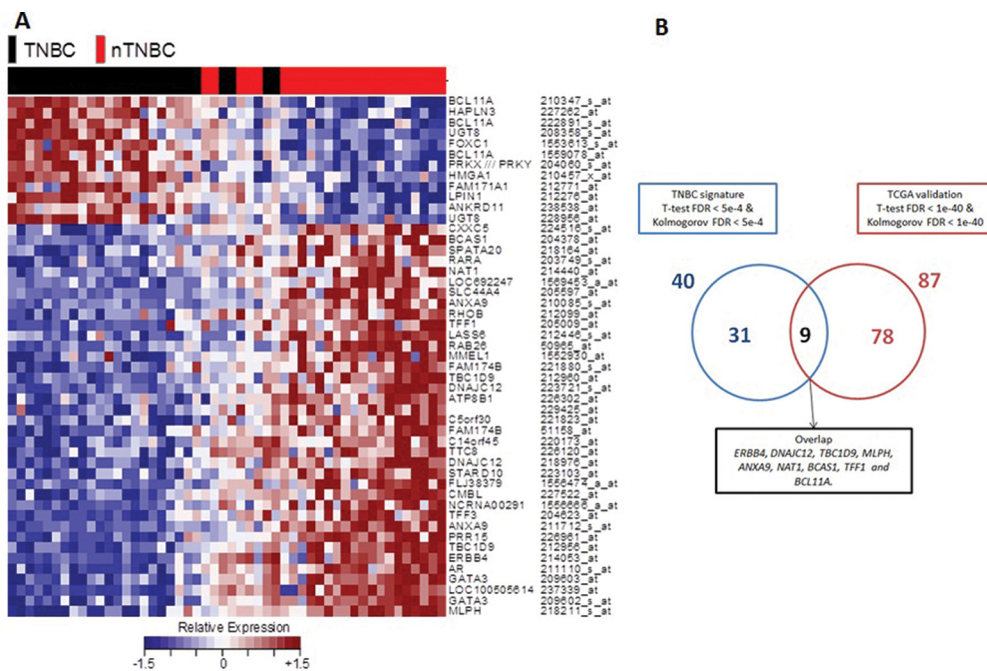


Figure 3. TNBC signature. (A) The heat map shows 49 differentially expressed probes (corresponding to 40 genes) in TNBC versus nTNBC. At the bottom of the figure is a color code that shows expression intensity (Z score). Blue represents low gene expression, red high gene expression. At the top of the figure is a color code for TNBC (black) and nTNBC (red) patients. In the heat map, each column is for a patient and each row is for a single probe. The right side of the heat map shows official symbol gene and probe identification. (B) Venn diagram of TNBC signature and TCGA validation. Blue circle represents 40 genes of our TNBC signature. Red circle represents 87 genes from TCGA data. Overlap of both circles (blue and red) represents nine genes that are present in both sample data using same analysis.

DISCUSSION

Molecular characterization of BC by gene expression profiling (gene signature) has positively influenced the diagnosis, prognosis, and therapies for this tumor, having a clear impact on the disease-free survival and quality of life of affected patients. These developments have paved the way for the translation of molecular tests to the clinical care of oncologic patients (12,13).

Regardless of advances in the molecular biology of BC, TNBC, an aggressive form of this neoplasm, remains undecipherable to current molecular tests, and patients are still in need of rational therapies. After several efforts, there are no gene expression profiles available to characterize TNBC, because studies with this purpose have demonstrated heterogeneous and nonconclusive results (2,14–17). This heterogeneity may result from differences in study design (i.e.,

group comparisons as TNBC versus normal breast tissue), data expression sets (i.e., microarray data from fresh tissue versus public tumor-expression databases), social stratifications by factors like ethnicity and income status, and so on. These studies were valuable to define sets of genes participating in tumorigenesis. They have demonstrated involvement of common pathways related to cell differentiation, epithelial mesenchymal transition (EMT), immune function, androgen receptor (AR) pathways, inflammation, and angiogenesis (1,14,17–21). However, the collection of genes identified in each study is different, and they are not useful to discriminate TNBC from nTNBC (1,14,17–21). Thus, the search for a useful gene signature for TNBC continues.

The purpose of this study was to identify a gene signature for TNBC with hallmark genes profiling this tumor subtype. To achieve this, we decided to compare

TNBC against the rest of BC subtypes (nTNBC). Emphasis was considered in the selection of samples. Our study population (TNBC and nTNBC) was composed of therapy-naïve women with BC who came from the same geographical region, sharing similar sociodemographic and clinical-pathological characteristics. The sampling process was the same for all patients, and biopsied tumors were >2 cm and/or had angiolymphatic invasion.

Gene expression profiles were assessed in tumor samples of 50 northeastern Mexican women. Sample collection was performed consecutively (n = 149) until we reached the TNBC population for this study. Subsequently, nTNBC samples were paired according to clinical stage. We observed a higher TNBC prevalence (19.5%) than reported for Caucasian women (10–13%) (5). Contrastingly, this prevalence is lower than that reported for the central Mexican population (23%)

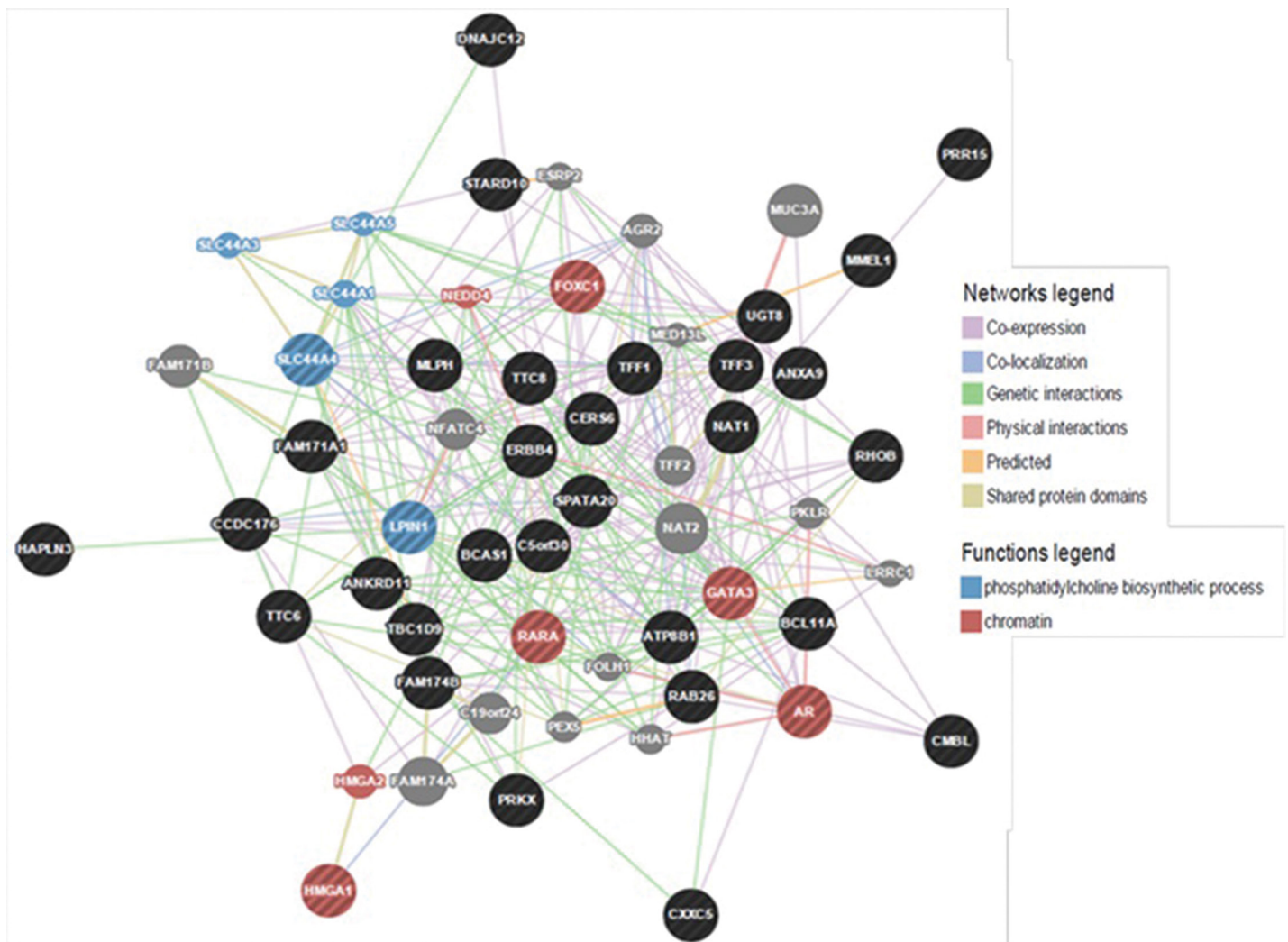


Figure 4. TNBC signature's gene network. The left side represents the gene network from the TNBC signature. Each black circle signifies a gene from the TNBC signature, while gray ones symbolize other genes not included in TNBC signatures. Lines denote a relationship between circles (genes), and line thickness correlates to relationship intensity. The right side shows color codes for lines that connect circles (genes) and functions of some genes from the gene network. Main functions are in blue (phosphatidylcholine biosynthetic process) and red (chromatin).

(27). Other similarities that caught our attention were age at cancer diagnosis, BMI (patients were overweight or obese), and glucose levels in both groups. This allowed us to pair patients with homogeneous clinical characteristics, differing only by tumor subtype (nTNBC or TNBC). We consider this crucial for identifying an intrinsic signature between these two groups of tumors.

BC classification was performed according to IHQ and microarray analysis.

IHQ was the first criterion for tumor classification. However, molecular classification by PAM50 displayed a different BC subtype in six tumor samples (Figure 2). Four of these were originally classified as TNBC by IHQ, but as a different BC subtype by the microarray analysis. The two remaining samples were reclassified inside the nTNBC group. Interestingly, these six samples are located at the center of the heat map for our gene signature (Figure 3A), implying the

heterogeneity of these tumors, probably containing a mixture of cancer clones representing different BC subtypes.

The signature described here is defined by overexpression of nine genes (*FOXC1*, *PRKX/PRKY*, *UGT8*, *BCL11A*, *HMGA1*, *LPIN1*, *FAM171A1*, *HAPLN3*, and *ANKRD11*) and downregulation of an additional 31 genes. Some of these deregulated genes are clearly involved in tumor pathways. Genes like *HMGA1*, *FOXC1*, and *BCL11A* have been inde-

Table 2. Gene network.

Feature	FDR	Genes in network	Genes in genome
Phosphatidylcholine biosynthetic process <i>SLC44A4</i> <i>LPIN1</i>	1.83e-04	5	25
Phosphatidylcholine metabolic process <i>SLC44A4</i> <i>LPIN1</i>	8.71e-03	5	60
Organic hydroxy compound biosynthetic process <i>SLC44A4</i> <i>LPIN1</i> <i>GATA3</i>	1.11e-02	6	125
Ethanolamine-containing compound metabolic process <i>SLC44A4</i> <i>LPIN1</i>	1.11e-02	5	74
Chromatin <i>FOXC1</i> <i>GATA3</i> <i>RARA</i> <i>AR</i> <i>HMGAI</i>	1.11e-02	7	212
Cellular amine metabolic process <i>SLC44A4</i> <i>LPIN1</i> <i>GATA3</i>	2.87e-02	6	171
Chromatin DNA binding <i>RARA</i> <i>FOXC1</i>	2.87e-02	3	15
Amine metabolic process <i>SLC44A4</i> <i>LPIN1</i> <i>GATA3</i>	2.87e-02	6	177
Alcohol biosynthetic process <i>SLC44A4</i> <i>LPIN1</i>	2.87e-02	5	101
Cellular biogenic amine metabolic process <i>SLC44A4</i> <i>LPIN1</i>	3.97e-02	5	113

pendently studied and considered of clinical utility for TNBC as diagnostic and metastatic biomarkers, and some have been considered as potential therapeutic targets. For example, upregulation of *FOXC1* promotes tumor growth, EMT, and metastatic behavior (28). *BCL11A* overexpression affects p53-mediated apoptosis and plays a key role in TNBC (16). On the contrary, downregulation of *NAT1*, which codifies for the n-acetyl transferase 1, is implied in the signature. *NAT1* has been reported as underreg-

ulated in breast tumors and associated with tamoxifen resistance. In TNBC cells, downregulation of this gene is associated with overexpression of *SNAIL* and downregulation of N-cadherin and β -catenin, which regulate the EMT behavior (29).

Most interestingly, several of the overregulated and downregulated genes from this signature are involved in the metabolism of lipids. Some of these metabolic abnormalities may suggest some peculiarities of the biology and poor

prognosis of TNBC, which have been independently tied to aggressive BC. For example, upregulation of *HMGAI* is associated with insulin resistance and metabolic syndrome, with increased levels of cholesterol and triglycerides and low levels of high-density lipoprotein (30–32). This gene is observed as overexpressed in breast cancer, particularly in TNBC and its metastases (33). *LPIN1* codes for an enzyme directly involved in the synthesis of triglycerides and phospholipids. This gene is crucial for the homeostasis of the endoplasmic reticulum and membrane biogenesis. Recently, it has been demonstrated that overexpression of this gene is critical for TNBC cell survival, and thus it is considered a poor prognosis marker. *LPIN1* is an attractive target for personalized cancer therapy (34). *UGT8* encodes for an enzyme involved in the synthesis of galactosyl-ceramides and sphingolipids (35). Upregulation of this gene increases the cell proliferative index and decreases apoptosis, favoring metastatic processes. In addition, overproduction of this enzyme increases resistance to doxorubicin (36). On the other hand, other genes are also involved in lipid metabolism and cancer in their underregulated state. An example is *LASS6*, which encodes for the ceramide synthase 6. This enzyme was recently implicated in EMT. Downregulation of this gene induces decreased autophagy and increased drug resistance (37). *STARD10* codifies for the StAR-related lipid transfer protein 10 (or PCTP-like protein), a phospholipid transfer protein. Downregulation of *STARD10* is associated with defective insulin secretion and has been considered a biomarker of TNBC aggression (38). *ATP8B1* gene encodes for the ATPase phospholipid transporting 8B1 enzyme, which translocates phospholipids from the outer to the inner bilayer membrane and is crucial in maintaining the cell lipid balance. Downregulation of this gene is observed in liver cholestasis and in liver and colorectal cancer (39). Finally, the *AR* gene encoding for the androgen receptor protein was also downregulated in the signature. This gene is overexpressed in some TNBC

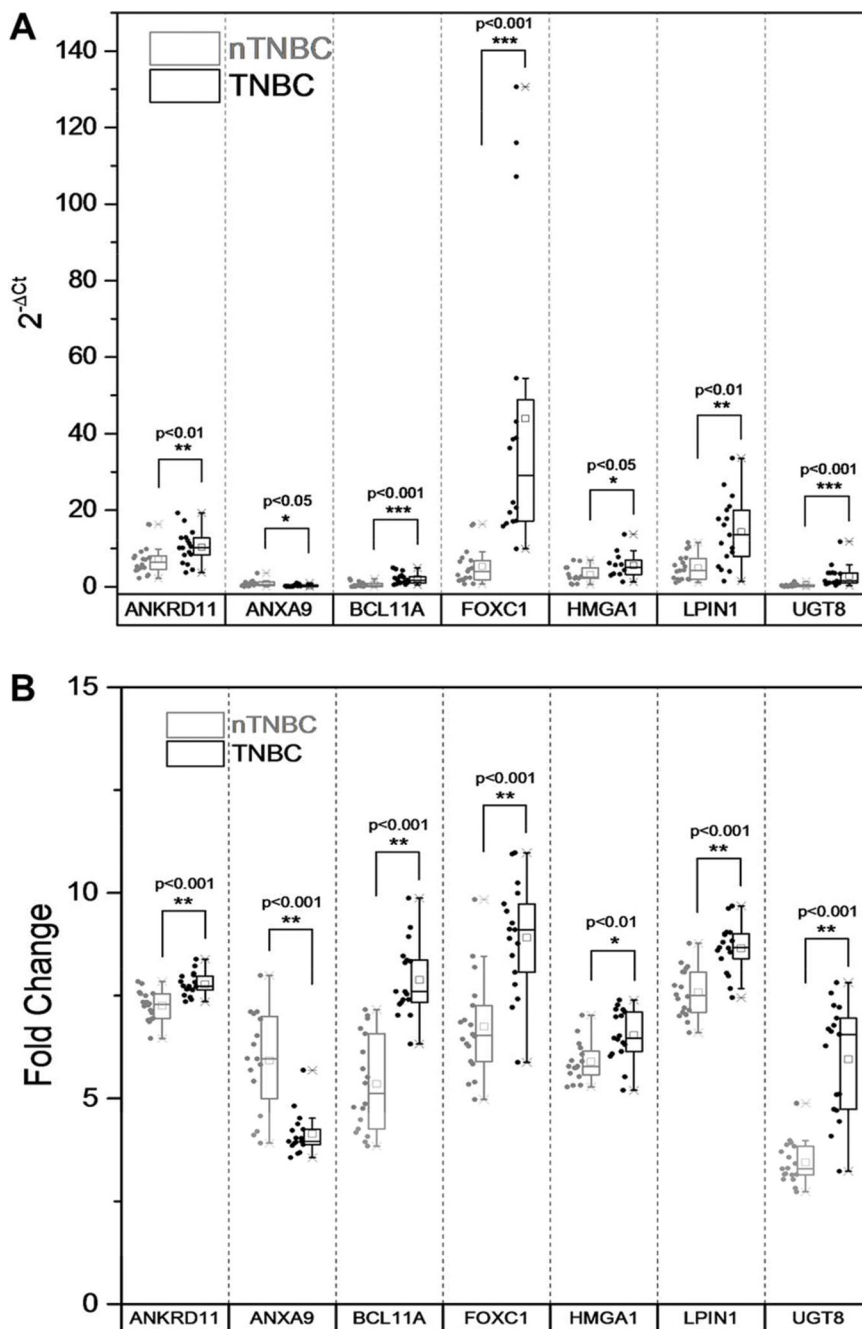


Figure 5. Real-time qPCR validation. Black and gray color show box plots for TNBC (black) and nTNBC (gray). (A) qPCR data ($2^{-\Delta Ct}$ values) from *HMGA1*, *ANKRD11*, *BCL11A*, *FOXC1*, *LPIN1*, *UGT8*, and *ANXA9* genes (using *GRAMD1A* for normalization). (B) Microarray data (fold change) from *HMGA1*, *ANKRD11*, *BCL11A*, *FOXC1*, *LPIN1*, *UGT8*, and *ANXA9* genes (using *GRAMD1A* for normalization).

patients, making the protein a candidate for targeted therapy, but its prognostic role is not yet clear (40). Recently, down-regulation of *AR* has been associated with

diabetes mellitus in TNBC patients, and it has been suggested that hyperglycemia directly regulates *AR* levels. This observation is of interest for the present study, in

which several patients showed high glucose levels (20% were diabetic) and an increased average BMI. This study remarks on the implication of lipid metabolism in the physiopathology of TNBC and, to some extent, the possible implications of increased insulin resistance, overweight, and metabolic syndrome for this aggressive breast cancer subtype. However, further studies are required to understand these plausible associations.

As mentioned above, previous reports analyzed global profiles for TNBC. However, the methodological approach is different in all of them. Kreike *et al.* compared TNBC frozen fresh tissues versus nTNBC data available from public databases (1). The androgen receptor gene (*AR*) is the only gene shared with our signature. Al-ejeh *et al.* performed an *in silico* expression profile comparison for TNBC versus nTNBC using public databases. They selected eight genes, two of them involved in the estrogen receptor pathway (*MAPT* and *MYB*) and six in chromosomal instability (*MELK*, *MCM10*, *CENPA*, *EXO1*, *TTK*, and *KIF2C*). None of these eight genes were identified in our signature (29). Rody *et al.* studied publicly available TNBC expression data to evaluate 5-year recurrence-free survival. Sixteen genes from the metagene analysis were differentially expressed. *AR* is again the only gene that matches our signature (2). Cascione *et al.* analyzed profiles of mRNA and microRNAs for TNBC paraffin-embedded tissues, comparing tumor- and nontumor-adjacent tissue to evaluate cancer survival. They identified genes related to overall survival and distant disease-free survival (17). Interestingly, *ERBB4* and *AR* genes were underexpressed, which are also underexpressed in our signature. Karn *et al.* selected TNBC data from public databases and identified a predicted signature of 264 genes not described in previous signatures, none of which match our study (20). Komatsu *et al.* described 301 overexpressed and 321 underexpressed genes in 30 TNBC and 13 normal ductal epithelial cells, all isolated by single-cell microdissection. They found

subexpression of *TBC1D9*, *TFF3*, *ERBB4*, *GATA3*, and *MLPH*, which also matches this study (14).

These reports summarize the heterogeneity of the samples, the data, and the different approaches used to define gene signatures for TNBC. However, these studies define some underexpressed genes that were also discovered in our signature.

Considering the epidemiologic characteristics and the genetic background of our analyzed population, we decided to verify whether the signature obtained in this study was consistent in other populations. To achieve this, we looked for differences in TNBC versus nTNBC in microarray gene expression datasets stored in the TCGA. This comparison showed an overlap of nine genes (*ERBB4*, *DNAJC12*, *TBC1D9*, *MLPH*, *ANXA9*, *NAT1*, *BCAS1*, *TFF1*, and *BCL11A*) with our signature. Of these, eight were underexpressed and one overexpressed. These nine genes could represent general pathways of carcinogenesis in TNBC, while the rest (31 nonoverlapping genes) could represent dysfunctional metabolism in this study (high BMI and glucose levels), probably affecting carcinogenic pathways. Still, the findings described in this article may be pertinent for Mexican patients and require evaluation in other ethnic groups and populations.

CONCLUSION

This study describes a genomic signature of 40 genes useful to discriminate TNBC, few of them independently associated with TNBC by other researchers. Interestingly, some of the genes in our signature are involved in metabolic processes related to fats and carbohydrates. A comparison of TNBC versus nTNBC in the TCGA database revealed a match of only nine genes with our signature. Thus, these genes should have important implications for the TNBC phenotype. To our knowledge, this is the first TNBC signature describing genes involved in general metabolism. However, the biological significance and molecular interactions should be further researched to

determine the intrinsic pathophysiology of this BC subtype.

ACKNOWLEDGMENTS

This work was supported by CONACYT-Mexico through an approved grant (SALUD-CONACYT-2011-C01-162301). We are very grateful to the personnel of the Unidad de Genómica, Centro de Investigación y Desarrollo en Ciencias de la Salud, Universidad Autónoma de Nuevo León for their technical assistance. SKSF and YXPP had scholarships from CONACYT (CVU nos. 369726 and 217104, respectively).

DISCLOSURE

The authors declare they have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

REFERENCES

- Kreike B, et al. (2007) Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res.* 9:R65.
- Rody A, et al. (2011) A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res.* 13:R97.
- Hudis CA, Gianni L. (2011) Triple-negative breast cancer: an unmet medical need. *Oncologist.* 16 Suppl1:1–11.
- Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer.* 109:1721–8.
- Carey LA, et al. (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA.* 295:2492–502.
- Barquera S, Campos-Nonato I, Hernandez-Barrera L, Pedroza A, Rivera-Dommarco JA. (2013) [Prevalence of obesity in Mexican adults 2000–2012]. *Salud publica de Mexico.* 55 Suppl2:S151–60.
- de la Vara-Salazar E, Suarez-Lopez L, Angeles-Llerenas A, Torres-Mejia G, Lazcano-Ponce E. (2011) [Breast cancer mortality trends in Mexico, 1980–2009]. *Salud publica de Mexico.* 53:385393.
- Knaul FM, et al. (2009) [Breast cancer in Mexico: an urgent priority]. *Salud publica de Mexico.* 51 Suppl2:s335–44.
- Swisher A, et al. (2015) Exercise and dietary advice intervention for survivors of triple-negative

- breast cancer: effects on body fat, physical function, quality of life, and adipokine profile. *Support Care Cancer.* 23:2995–3003.
- Coughlin SS, Smith SA. (2015) The Insulin-like Growth Factor Axis, Adipokines, Physical Activity, and Obesity in Relation to Breast Cancer Incidence and Recurrence. *Cancer Clin. Oncol.* 4:24–31.
- Seo BR, et al. (2015) Obesity-dependent changes in interstitial ECM mechanics promote breast tumorigenesis. *Sci. Transl. Med.* 7:301ra130.
- van de Vijver MJ, et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347:1999–2009.
- Paik S, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* 351:2817–26.
- Komatsu M, et al. (2013) Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. *Int. J. Oncol.* 42:478–506.
- Al-Ejeh F, et al. (2014) Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer. *Oncogenesis.* 3:e100.
- Khaled WT, et al. (2015) BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nat. Commun.* 6:5987.
- Cascione L, et al. (2013) Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. *PLoS One.* 8:e55910.
- Lehmann BD, et al. (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* 121:2750–67.
- Hennessy BT, et al. (2009) Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res.* 69:4116–24.
- Karn T, et al. (2011) Homogeneous datasets of triple negative breast cancers enable the identification of novel prognostic and predictive signatures. *PLoS One.* 6:e28403.
- Tchou J, et al. (2012) Human breast cancer associated fibroblasts exhibit subtype specific gene expression profiles. *BMC Med. Genomics.* 5:39.
- Parker JS, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27:1160–7.
- Benjamini Y, Hochberg Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57:289–300.
- R Core Team. (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Aguirre-Gamboa R, et al. (2013) SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis. *PLoS One.* 8:e74250.

26. Warde-Farley D, *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38:W214–20.
27. Lara-Medina F, *et al.* (2011) Triple-negative breast cancer in Hispanic patients: high prevalence, poor prognosis, and association with menopausal status, body mass index, and parity. *Cancer.* 117:3658–69.
28. Tkocz D, *et al.* (2012) BRCA1 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers. *Oncogene.* 31:3667–78.
29. Smid M, *et al.* (2006) Genes associated with breast cancer metastatic to bone. *J. Clin. Oncol.* 24:2261–7.
30. Chiefari E, *et al.* (2013) A polymorphism of HMGA1 is associated with increased risk of metabolic syndrome and related components. *Sci. Rep.* 3:1491.
31. Foti D, *et al.* (2005) Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. *Nat. Med.* 11:765–73.
32. Pullinger CR, *et al.* (2014) Evidence that an HMGA1 gene variant associates with type 2 diabetes, body mass index, and high-density lipoprotein cholesterol in a Hispanic-American population. *Metab. Syndr. Relat. Disord.* 12:25–30.
33. Shah SN, *et al.* (2013) HMGA1: a master regulator of tumor progression in triple-negative breast cancer cells. *PLoS One.* 8:e63419.
34. He J, *et al.* (2017) Lipin-1 regulation of phospholipid synthesis maintains endoplasmic reticulum homeostasis and is critical for triple-negative breast cancer cell survival. *FASEB J.* 31:1–12.
35. Sprong H, *et al.* (1998) UDP-galactose:ceramide galactosyltransferase is a class I integral membrane protein of the endoplasmic reticulum. *J. Biol. Chem.* 273:25880–88.
36. Owczarek TB, *et al.* (2013) Galactosylceramide affects tumorigenic and metastatic properties of breast cancer cells as an anti-apoptotic molecule. *PLoS One.* 8:e84191.
37. Schiffmann S, *et al.* (2009) Ceramide synthases and ceramide levels are increased in breast cancer tissue. *Carcinogenesis.* 30:745–52.
38. Murphy NC, *et al.* (2010) Loss of STARD10 expression identifies a group of poor prognosis breast cancers independent of HER2/Neu and triple negative status. *Int. J. Cancer.* 126:1445–53.
39. Eldai H, *et al.* (2013) Novel genes associated with colorectal cancer are revealed by high resolution cytogenetic analysis in a patient specific manner. *PLoS One.* 8:e76251.
40. Rahim B, O'Regan R. (2017) AR Signaling in Breast Cancer. *Cancers.* 9:21.

Cite this article as: Santuario-Facio SK, *et al.* (2017) A new gene expression signature for triple-negative breast cancer using frozen fresh tissue before neoadjuvant chemotherapy. *Mol. Med.* 23:101–111.