



The Transcriptomic Portrait of Locally Advanced Breast Cancer and Its Prognostic Value in a Multi-Country Cohort of Latin American Patients

Andrea Sabina Llera^{1†}, Eliana Saul Furquim Werneck Abdelhay^{2†}, Nora Artagaveytia³, Adrián Daneri-Navarro⁴, Bettina Müller⁵, Carlos Velazquez⁶, Elsa B. Alcoba⁷, Isabel Alonso⁸, Daniela B. Alves da Quinta^{1,9}, Renata Binato², Alicia Inés Bravo¹⁰, Natalia Camejo³, Dirce Maria Carraro¹¹, Mónica Castro¹², Juan M. Castro-Cervantes¹³, Sandra Cataldi¹⁴, Alfonso Cayota¹⁵, Mauricio Cerda¹⁶, Alicia Colombo¹⁷, Susanne Crocama¹⁸, Alicia Del Toro-Arreola⁴, Raúl Delgado-Cisterna¹³, Lucía Delgado³, Marisa Dreyer-Breitenbach¹⁹, Laura Fejerman²⁰, Elmer A. Fernández^{21,22}, Jorge Fernández²³, Wanda Fernández²⁴, Ramón A. Franco-Topete²⁵, Carolina Gabay¹², Fancy Gaete²⁶, Adriana Garibay-Escobar⁶, Jorge Gómez²⁷, Gonzalo Greif¹⁵, Thomas G. Gross²⁸, Marisol Guerrero²⁹, Marianne K. Henderson²⁸, Miguel E. Lopez-Muñoz⁶, Alejandra Lopez-Vazquez⁶, Silvina Maldonado¹⁰, Andrés J. Morán-Mendoza³⁰, María Aparecida Nagai³¹, Antonio Ocegüera-Villanueva³², Miguel A. Ortiz-Martínez³³, Jael Quintero³⁴, Antonio Quintero-Ramos⁴, Rui M. Reis³⁵, Javier Retamales³⁶, Ernesto Rivera-Claisse³⁷, Darío Rocha²², Robinson Rodríguez³⁸, Cristina Rosales⁷, Efraín Salas-González³⁰, Verónica Sanchotena⁷, Laura Segovia³⁹, Juan Martín Sendoya¹, Aida A. Silva-García²⁵, Alejandra Trincherro¹⁰, Olivia Valenzuela⁶, Vidya Vedham²⁸, Livia Zagame³², United States-Latin American Cancer Research Network (US-LACRN)[†] and Osvaldo L. Podhajcer^{1†}

OPEN ACCESS

Edited by:

Carolina Panis,
Universidade Estadual do Oeste do
Paraná, Brazil

Reviewed by:

Prihantono Prihantono,
Hasanuddin University, Indonesia
Bruno Ricardo Barreto Pires,
University of Ottawa, Canada

*Correspondence:

Andrea Sabina Llera
allera@leloir.org.ar

[†]These authors share senior
authorship

[†]The complete list of collaborators of
the US-LACRN network is displayed at
the end of the manuscript

Specialty section:

This article was submitted to
Cancer Epidemiology and Prevention,
a section of the journal
Frontiers in Oncology

Received: 15 December 2021

Accepted: 14 February 2022

Published: 22 March 2022

¹ Molecular and Cellular Therapy Laboratory, Fundación Instituto Leloir-CONICET, Buenos Aires, Argentina,

² Bone Marrow Transplantation Unit, Instituto Nacional de Câncer, Rio de Janeiro, Brazil, ³ Hospital de Clínicas Manuel Quintela, Universidad de la República, Montevideo, Uruguay, ⁴ Universidad de Guadalajara, Guadalajara, Mexico,

⁵ Instituto Nacional del Cáncer, Santiago, Chile, ⁶ Universidad de Sonora, Hermosillo, Mexico, ⁷ Hospital Municipal de Oncología María Curie, Buenos Aires, Argentina, ⁸ Centro Hospitalario Pereira Rossell, Montevideo, Uruguay, ⁹ Universidad Argentina de la Empresa (UADE), Instituto de Tecnología (INTEC), Buenos Aires, Argentina,

¹⁰ Hospital Regional de Agudos Eva Perón, San Martín, Argentina, ¹¹ Laboratory of Genomics and Molecular Biology/Centro Internacional de Pesquisa (CIPE), AC Camargo Cancer Center, Sao Paulo, Brazil, ¹² Instituto de Oncología Angel Roffo, Buenos Aires, Argentina, ¹³ Hospital de Especialidades CMNO-IMSS, Guadalajara, Mexico, ¹⁴ Instituto Nacional del Cáncer, Montevideo, Uruguay, ¹⁵ Institut Pasteur de Montevideo, Montevideo, Uruguay, ¹⁶ Integrative Biology Program, Instituto de Ciencias Biomédicas (ICBM), Centro de Informática Médica y Telemedicina, Facultad de Medicina, Instituto de Neurociencias Biomédicas, Universidad de Chile, Santiago, Chile, ¹⁷ Department of Pathology, Facultad de Medicina y Hospital Clínico, Universidad de Chile, Santiago, Chile, ¹⁸ Oncology Department, Instituto Nacional de Câncer, Rio de Janeiro, Brazil,

¹⁹ Instituto de Biología Roberto Alcántara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil, ²⁰ Department of Public Health Sciences and Comprehensive Cancer Center, University of California Davis, Davis, CA, United States, ²¹ Centro de Investigación y Desarrollo en Inmunología y Enfermedades Infecciosas [Centro de Investigación y Desarrollo en Inmunología y Enfermedades Infecciosas (CIDIE) CONICET/Universidad Católica de Córdoba], Córdoba, Argentina, ²² Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Córdoba, Argentina, ²³ Instituto de Salud Pública, Santiago, Chile, ²⁴ Hospital San Borja Arriarán, Santiago, Chile, ²⁵ Organismo Público Descentralizado (OPD), Hospital Civil de Guadalajara, Universidad de Guadalajara, Guadalajara, Mexico, ²⁶ Hospital Luis Tisne, Santiago, Chile, ²⁷ Texas A&M University, Houston, TX, United States, ²⁸ Center for Global Health, National Cancer

Institute, Rockville, MD, United States, ²⁹ Hospital San José, Santiago, Chile, ³⁰ Hospital de Gineco-Obstetricia CMNO-IMSS, Guadalajara, Mexico, ³¹ Center for Translational Research in Oncology, Cancer Institute of São Paulo (ICESP), Sao Paulo University Medical School, Sao Paulo, Brazil, ³² Instituto Jalisciense de Cancerología, Guadalajara, Mexico, ³³ Hospital General Regional No. 1, IMSS, Obregon, Mexico, ³⁴ Universidad de Sonora, Cajeme, Mexico, ³⁵ Molecular Oncology Research Center, Hospital de Cáncer de Barretos, Barretos, Brazil, ³⁶ Grupo Oncológico Cooperativo Chileno de Investigación, Santiago, Chile, ³⁷ Centro Estatal de Oncología, Hermosillo, Mexico, ³⁸ Hospital Central de las Fuerzas Armadas, Montevideo, Uruguay, ³⁹ Hospital Barros Luco Trudeau, Santiago, Chile

Purposes: Most molecular-based published studies on breast cancer do not adequately represent the unique and diverse genetic admixture of the Latin American population. Searching for similarities and differences in molecular pathways associated with these tumors and evaluating its impact on prognosis may help to select better therapeutic approaches.

Patients and Methods: We collected clinical, pathological, and transcriptomic data of a multi-country Latin American cohort of 1,071 stage II-III breast cancer patients of the Molecular Profile of Breast Cancer Study (MPBCS) cohort. The 5-year prognostic ability of intrinsic (transcriptomic-based) PAM50 and immunohistochemical classifications, both at the cancer-specific (OSC) and disease-free survival (DFS) stages, was compared. Pathway analyses (GSEA, GSVA and MetaCore) were performed to explore differences among intrinsic subtypes.

Results: PAM50 classification of the MPBCS cohort defined 42.6% of tumors as LumA, 21.3% as LumB, 13.3% as HER2E and 16.6% as Basal. Both OSC and DFS for LumA tumors were significantly better than for other subtypes, while Basal tumors had the worst prognosis. While the prognostic power of traditional subtypes calculated with hormone receptors (HR), HER2 and Ki67 determinations showed an acceptable performance, PAM50-derived risk of recurrence best discriminated low, intermediate and high-risk groups. Transcriptomic pathway analysis showed high proliferation (i.e. cell cycle control and DNA damage repair) associated with LumB, HER2E and Basal tumors, and a strong dependency on the estrogen pathway for LumA. Terms related to both innate and adaptive immune responses were seen predominantly upregulated in Basal tumors, and, to a lesser extent, in HER2E, with respect to LumA and B tumors.

Conclusions: This is the first study that assesses molecular features at the transcriptomic level in a multicountry Latin American breast cancer patient cohort. Hormone-related and proliferation pathways that predominate in PAM50 and other breast cancer molecular classifications are also the main tumor-driving mechanisms in this cohort and have prognostic power. The immune-related features seen in the most aggressive subtypes may pave the way for therapeutic approaches not yet disseminated in Latin America.

Clinical Trial Registration: ClinicalTrials.gov (Identifier: NCT02326857).

Keywords: breast cancer, Latin America, PAM50 subtypes, risk of recurrence, biological pathways

INTRODUCTION

Breast cancer is a heterogeneous disease featuring distinct histological, molecular and clinical phenotypes. Although traditional classification systems leveraging clinicopathological and molecular markers are well established, most class discovery and prognostic signatures have arisen from studies including

patients with European ancestry overrepresentation. For example, a recent study of the genetic ancestry of biospecimens in the The Cancer Genome Atlas (TCGA) database confirmed that only 41 breast cancer samples were classified as having Native/Indigenous American ancestry (0.4% of total samples) (1). METABRIC, another relevant breast cancer genomic resource, is composed almost exclusively of individuals of European ancestry (2).

The diversity of Latin American populations, both at the cultural and genetic ancestry levels, has an impact on the distribution of breast cancer risk. Several studies have shown differences in breast cancer incidence and mortality in the Latin American population (3, 4), in the proportion of aggressive subtypes (5), in genetic markers of risk (6) and in other epidemiological risk factors (7). Particularly for the molecular subtype distribution, most evidence is based on small, country-based cohorts or larger US-Hispanic cohorts that may not capture the diversity of Latin American populations or the impact of sociodemographic aspects [(8) and reviewed in (5)]. Thus, confirmation of the applicability and prognostic value of available intrinsic subtype classification signatures in a cohort of Latin American women is needed.

In the present study we analyzed the transcriptomic profile and comparatively assessed the prognostic performance of PAM50-based intrinsic and immunohistochemistry-based (IHC) surrogate subtype classifications in Latin American women included in the study Molecular Profiling of Stage II and III Breast Cancer in Latin American Women Receiving Standard-of-Care Treatment, in short, the Molecular Profile of Breast Cancer Study (MPBCS). Additionally, we compared the enriched molecular pathways in MPBCS subtypes with those found in the TCGA selected cohort of stage II-III breast cancer patients. This initiative of the United States-Latin America Cancer Research Network (US-LACRN, or LACRN), launched by the Center for Global Health of the US-National Cancer Institute, NIH, USA, and institutions of Argentina, Brazil, Chile, Mexico and Uruguay, focused on the integrative profiling of stage II-III breast cancer in Latin American women (9).

MATERIALS AND METHODS

Study Setup and Patient Eligibility

After setting up consensual scientific, technical and oversight standards for research studies, LACRN launched the LACRN-MPBCS (or MPBCS) in 37 health care institutions (mostly from the public health system) and research institutes in Argentina (n=6), Brazil (n=4), Chile (n=9), México (n=13), and Uruguay (n=5), with the primary objective of characterizing the distribution of molecular profiles of stage II and III invasive breast cancer in Latin American women. The approved protocol for this study is in **Supplementary File 1**.

From 2011 to 2014, the MPBCS recruited 1449 patients (281 in Argentina, 314 in Brazil, 205 in Chile, 531 in Mexico, and 118 in Uruguay), from whom 149 were excluded due to predefined exclusion criteria or consent withdrawal, leaving 1300 eligible patients. Relevant inclusion criteria consisted of being a Latin American woman of any ethnicity residing in the recruitment countries, aged 18 years or older, with AJCC 7 clinical stage II and III breast cancer, with ECOG performance status 0-1, accessible for biopsy or candidate for primary surgery, and who have shown understanding and were willing to sign the informed consent. The exclusion criteria were prior history of non-breast malignancy up to 5 years before inclusion in the

study; bilateral invasive, *in-situ*, or inflammatory breast cancer; clinical or radiological evidence of distant metastases; a prior record of hormone therapy, chemotherapy, biological therapy, targeted therapy and/or radiation therapy for breast cancer; and pregnancy and lactation. Fertile women were required to use nonhormonal contraceptives.

This constitutes a case-only, observational cohort study that was conducted in accordance with the Declaration of Helsinki and local regulations. The local Ethics Committees in each country and the NIH Institutional Review Board approved the study (IRB #: 15CN055; iRIS Reference #: 549650). All participants signed informed consent forms. This manuscript describes the primary endpoint of the study, which was the characterization of molecular profiles of breast cancer (AJCC 7 clinical stage II or III) in Latin American women with gene-expression data (n=1071). Power calculations are thus not applicable for this descriptive primary endpoint. By the study design, which included other objectives and outcomes not pertinent to this manuscript, this cohort was hospital-based and focused on tumor stages II and III (**Supplementary Information 1**).

Clinical Procedures and Collection of Tissue Samples

Participants enrolled in the study answered a sociodemographic questionnaire and underwent a routine standard-of-care clinical, radiographic, and surgical evaluation. Clinicopathological, epidemiological and molecular data were collected in protocol-specific case report forms (CRFs) using common data elements, and according to a Manual of Operations (MOP) that was built for the study. This study used OpenClinica® (openclinica.com) as the clinical data management system.

After participants signed the written informed consent form, they were enrolled in the study, their clinical evaluation was performed, and the treating physician recommended the appropriate standard of care treatment for the participant – either primary surgery or neoadjuvant chemotherapy followed by surgery – based on tumor stage, according to local institutional guidelines. Therefore, 62% of the eligible patients were treated by surgery and 38% by neoadjuvant chemotherapy followed by surgery. During neoadjuvant chemotherapy, participants received chemotherapy for a period of 16-24 weeks prior to surgery. According to protocol, the participant received treatment with four cycles of doxorubicin (60mg/m²) and cyclophosphamide (600mg/m²), every three weeks, followed by four cycles of docetaxel (90 mg/m²) with/without 6 mg/kg trastuzumab (8 mg/kg as loading dose) every 3 weeks or, alternatively, 12 weekly cycles of paclitaxel (80 mg/m²) with/without 6 mg/kg trastuzumab every 3 weeks (8 mg/kg as loading dose) for four cycles. Every three weeks the participant underwent a routine examination as directed by the physician. This included a physical examination, clinical tumor measurement, differential blood count, blood biochemistry panel, as well as a serological pregnancy test if necessary. A MUGA or echocardiography was taken at baseline, before surgery and, for participants using trastuzumab, before the

start of trastuzumab associated to a taxane. Before surgery and after neoadjuvant chemotherapy, participants were re-examined by the surgeon in charge and were evaluated as to whether they were candidates for surgery (conservative or radical mastectomy). A preoperative mammogram (after the last chemotherapy administration) was done to each participant to evaluate regression or local clinical evolution. Surgery was performed within 42 days after completion of chemotherapy.

All participants had a medical follow-up every six months after surgery and a mammogram every year for five years after the date of surgery.

Specimen Handling and Pathology Assessment

All sample collection procedures were carried out according to the LACRN consensus standard operating procedures for frozen and formalin-fixed, paraffin-embedded (FFPE) specimens based on TCGA best practice recommendations (<https://brd.nci.nih.gov/brd/sop-compendium/show/701>). Tissue samples were obtained from core needle biopsies or from surgical resections. Participants who were candidates for neoadjuvant chemotherapy prior to surgery underwent a routine, standard-of-care diagnostic core needle biopsy procedure for histological confirmation of breast cancer. A minimum of four and a maximum of seven 14-gauge core biopsies were collected from each participant; two to three cores were fixed in formalin (fixation time strictly in the range of 16-32 hs) for diagnostic purposes and two to four core biopsies were snap-frozen in liquid nitrogen or dry ice/acetone bath for research purposes. Cold ischemia times were strictly monitored and registered in order not to exceed 30 minutes from extraction to fixation (formalin or freezing). BSI-II (bsisystems.com) was used as the biospecimen inventory management system.

All samples were processed at the local pathology laboratory at each participating institution, using the exact same lot of commercial kits produced by the same manufacturer (Dako Products, Agilent Santa Clara, CA) and following the manufacturer's protocol and instructions to reduce variability of results across the sites. FFPE samples were utilized for routine diagnostic work-up including histopathological evaluation of the tumor by hematoxylin and eosin (H&E) analysis and assessment of expression of tumor estrogen receptor - ER (ER pharmDx), progesterone receptor - PgR (PgR pharmDx), HER2 (HerceptTest), and Ki67 (FLEX Monoclonal Mouse Anti-Human Ki-67 Antigen, Clone MIB-1) status by IHC and complemented by HER2 fluorescent *in situ* hybridization - FISH (HER2 FISH pharmDx) or chromogenic *in situ* hybridization - CISH (Her2 DuoCISH pharmDx) as appropriate (for HER2 IHC 2+ cases). ER and PgR were considered as positive when 1% or more of the tumoral cells showed positive nuclear staining (10), irrespectively of intensity, although the Allred score was also recorded (11). Ki-67 results were recorded as % of tumoral cells with positive staining. HER2 status was evaluated and recorded as positive, negative or equivocal according to the protocol included in the MOP.

Quality control (QC) and assurance (QA) data for biospecimen handling and molecular procedures were thoroughly registered in CRFs. To ensure consistency and minimize variations, intracountry evaluations were performed for IHC data obtained with ER, PgR, Ki67 and HER2 and *in situ* hybridization using either FISH or CISH, as appropriate. In addition, reference centers in all countries went through an external evaluation by the College of American Pathologists for IHC (ER, PgR), and FISH and CISH HER2 testing. A specific analysis of the performance of the Ki67 determination was done to evaluate variation between centers (**Supplementary File 2 - Extended Methods**).

MPBCS Expression Data Acquisition and Pre-Processing

Two-color microarray analysis was performed using the Agilent gene expression platform (Series C Scanner, Agilent, Santa Clara, CA), processed in central molecular biology laboratories located in Argentina (Instituto Leloir), Brazil (Instituto de Câncer do Estado de Sao Paulo, Hospital do Câncer de Barretos, AC Camargo Cancer Center and Instituto Nacional de Câncer de Brasil), Mexico (University of Guadalajara and University of Sonora), Chile (Instituto de Salud Pública de Chile) and Uruguay (Instituto Pasteur). All laboratories were trained during a dedicated workshop and followed the same SOPs for total RNA isolation, quantification, labeling, hybridization, and scanning to ensure proper harmonization and minimize bias. An external quality control of results was performed by Dr Katherine Hoadley at University of North Carolina.

Snap-frozen tissue samples collected during biopsy or surgery before any chemotherapy were utilized to extract RNA. If both biopsy and surgical resection samples were available from a participant, RNA was extracted from the surgical sample. Only tissues with tumor content higher than 60%, as evaluated by histopathology, were processed.

Total tumor RNA was extracted from each tissue sample using RNeasy Mini Kit (QIAGEN, Hilden, Germany). RNAs were quantified using NanoDrop 2000 and RNA integrity was evaluated on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara CA, USA). A RIN higher than 6.0 was considered as of enough quality to perform the microarrays.

Tumor RNA and a Universal Human Reference RNA (Stratagene, San Diego CA, USA) were amplified, differentially labeled with Cy5 and Cy3, respectively, using Agilent Low Input Quick Amp Labeling Kit 2-Color, and subsequently hybridized 1:1 in mass (i.e. 825 ng each labeled RNA) with Human Gene Expression v2 4x44K (AMADID 026652) microarrays using the Agilent Gene Expression Hybridization Kit and Wash Buffer (Agilent Technologies, Santa Clara CA, USA). Feature Extraction 11.5.1.1 software (Agilent Technologies, Santa Clara CA, USA) was used by all laboratories to generate raw data. Raw data were subsequently analyzed in a single batch. Microarrays with signal intensities below 10,000 for 99% of signal distribution were considered failed. Samples with 'out of range' Feature Extraction QC evaluation metrics or with lower signal intensity

(<25,000) were flagged for later analysis. Expression data for probes was processed as log₂ ratios, probes were mapped to Entrez Gene IDs according to HsAgilentDesign026652.db R package and probes without a valid Entrez Gene ID were removed from further analysis.

A total of 229 patients (17%) failed to render reliable gene expression data due to one of the following quality issues: failure at obtaining biopsy, insufficient tumor content, poor RNA quality and/or quantity, poor array hybridization or poor quality metrics after array scanning.

TCGA Data Acquisition and Processing

TCGA clinical information and PAM50 classification limited to primary tumor samples stages II-III were obtained from cBioPortal (12) on 2020-04-22. RNA-seq raw counts from the stages II-III subset of breast cancer tumors of the TCGA RNAseq dataset (TCGARf) were downloaded with TCGA-Assembler pipeline (13) on 2018-01-12. Normalization factors were calculated using the trimmed mean of M-values (TMM (14)). For plotting and GSEA analysis, raw count data was transformed to log₂ counts per million with a prior count of 0.5. For differential expression analysis, raw counts were processed by *voom* (15) previous to *limma* (16) analysis. Gene symbols were obtained from the available Entrez IDs using *org.Hs.eg.db* R library (Carlson M (2019). *org.Hs.eg.db*: Genome wide annotation for Human. R package version 3.8.2.), and genes for which no valid symbol could be found were removed from the databases.

PAM50 Subtyping and Risk of Recurrence Score (ROR) Calculation

The expression log₂ ratios of the PAM50 genes were centered according to their balanced medians. To obtain the balanced medians, the dataset was subsampled to match the ER distribution of that in the training set used for PAM50. Balanced ER-subsets were identified in each country; for each country we identified the maximum number of high-quality ER-negative samples (the limiting factor), and then randomly selected a similar number of ER-positive samples. A total of 248 ER-negative and 248 ER-positive samples were selected to determine an ER-balanced median for each of the 50 genes used in the PAM50 subtyping scheme. This included pairs from each country (Argentina 59, Brazil 46, Chile 32, Mexico Guadalajara 73, Mexico Sonora 22, and Uruguay 16 pairs). The balanced medians were then used to center the gene expression data of the whole cohort and the PAM50 algorithm was ran as suggested (17) (generic code - <https://genome.unc.edu/pubsup/breastGEO/PAM50.zip>). Risk of Recurrence score – subtype only (ROR-S) was calculated as described in (17).

Pathway Enrichment and Differential Expression Analysis

For pathway analysis, additional QC and normalization of gene expression data were performed using the R package *Agi4x44.2c* (18). For the MPBCS cohort a batch correction was done due to a technical bias introduced by the year in which the microarray

analysis was performed (**Supplementary File 2** – Extended Methods). For this, prior to averaging probe expression by gene symbol, two batches based on processing dates (i.e. before or after September 2017) were defined for the whole cohort. Probe expression data was normalized to equalize median-absolute values and adjusted for batch effects using the ComBat function of *sva* package (v 3.32.1) with default settings and the earlier batch as reference.

Pathway enrichment was studied by GSEA and GSVA using a single list of 1200 curated gene sets comprising MSigDB Hallmark, KEGG and REACTOME, that contain between 15 and 500 genes, identified by Entrez ID (19, 20). GSEA was performed using GSEA desktop software v 4.0.3 with standard settings, considering a *P*-value <0.05 as significant. For microarray data, GSVA was applied on the expression data normalized to equalize median-absolute values, using the Gaussian method to estimate the cumulative distribution function. For RNA-seq data, GSVA was applied on raw counts data, using the Poisson method to estimate the cumulative distribution function. All comparisons with their GSEA normalized enriched score (NES) or GSVA coefficient, *P*- and adjusted *P*-values are detailed in **Supplementary Table 1**.

Collapse of enriched terms (gene sets) was performed for each contrast by 1) selecting the top ten significant (i.e. *p*<0.05) gene sets with highest GSEA NES absolute values, 2) verifying that those terms were also significant in GSVA (i.e. adjusted *P*-value <0.05, 3) reviewing the rest of significant terms in GSEA analysis using as keywords genes already deemed relevant in the literature for the corresponding intrinsic subtype contrast, and 4) manually identifying common terms and pathways in the resulting list of terms (**Supplementary Table 1**, see References worksheet).

Pathway analysis with differentially expressed genes were performed using MetaCore™ software (Clarivate Analytics, USA) (<https://portal.genego.com/>). First, normalized gene expression values of TCGA and MPBCS cohorts were used to obtain a list of differentially expressed genes (DEGs) for each contrast (e.g. luminal A vs basal-like and so on) using *limma* (16). A batch identification variable was included in the model as a covariate. DEGs for each contrast were defined as those with log fold-change (logFC) >0.5 and a Benjamini-Hochberg adjusted *P*-value<0.05. With both MPBCS and TCGA lists of DEGs, we identified those DEGs common to both cohorts and those exclusive to each cohort using <http://bioinformatics.psb.ugent.be/webtools/Venn/>. Lists of common, MPBCS-exclusive and TCGA-exclusive DEGs were then used for independent MetaCore analyses. Hierarchical clustering with Spearman distance with selected genes was performed using *ComplexHeatmap* R package (21).

Transcription Factor Activity Analysis

Transcription factors (TF) activities per sample were inferred from gene expression data using DoRothEA in combination with the statistical method VIPER as it incorporates the mode of regulation of each TF-target interaction (22). DoRothEA is a gene set resource containing signed TF-target interactions curated and collected from different types of evidence.

Each interaction is accompanied by an interaction confidence level ranging from A to E. We used the level A as it is the highest reliable level according to the authors.

The multi-unpaired comparisons between PAM50 subtypes were computed using the *limma* R package. The TF activities per sample obtained from DoRothEA-VIPER were used to fit a linear model for each TF and the *eBayes* test used to obtain the corresponding adjusted *P* values and the logFC as described in (23).

Determination of the Cytolytic Score (CYT)

CYT was calculated from the microarray batch-normalized expression matrix as the geometric mean of granzyme A (*GZMA*) and perforin (*PRF1*) expression (24).

Classification of TNBC Tumors in Lehmann's Molecular Subtypes (TNBC-6) and the Derived TNBC-4 and Bareche-2018

TNBC tumors were defined as those HR-/HER- (n=170). Lehmann's molecular subtypes (TNBC-6) were assigned to the TNBC samples using the TNBCtype online subtyping tool (<http://cbc.mc.vanderbilt.edu/tnbc/>) (25). Samples in which ER expression was greater than the 75 percentile at transcriptome level were removed, leaving a total of 145 samples for analysis. Each TNBC sample was assigned to a TNBC molecular subtype based on the highest Pearson correlation (centroid) and lowest p-value. Those samples with low correlation with all centroids (i.e. correlation coefficient <0.1 or P-value <0.05) or whose correlation coefficients were similar between subtypes (difference of two largest correlation coefficients <0.05) would be considered unclassified (UNS). A total of 128 tumors were assigned to any of the TNBC-6 classes.

For the TNBC-4 reclassification, IM and MSL samples were re-assigned to the second highest correlated centroid (26). For Bareche-2018 classification, BL2 samples were re-assigned to the second highest correlated centroid (27).

Statistical Analysis

Association among different clinicopathological variables was tested for significance with the Chi-square test, and Cramer's V test provided a measure of the strength of the association (28). Survival analyses were performed using the Kaplan-Meier estimator (*survival* R package), comparing each pair of curves with the log-rank test. Overall cancer-related survival (OSC) was defined as the interval from informed consent signature to the date of death. Disease-free survival (DFS) comprises the interval from date of surgery to first recurrence with medical confirmation, respectively. Disease-free patients were censored at the time of the last follow-up. For patients known to have a progression but missing an exact progression date, the date of the last medical examination was used. For each risk classification (i.e. PAM50, IHC, IHC-St. Gallen, etc) and each survival response (OSC and DFS), univariate and multivariate Cox proportional-hazards models were fitted with all available data points, and p-value and Harrell's correlation index were calculated.

Comparison of medians was performed using the Kruskal-Wallis test with Dunn's post test for sample pair comparison. *P* values were corrected for multiple comparisons using the Benjamini-Hochberg adjustment.

RESULTS

Clinical Features of MPBCS Samples According to the PAM50 Classifier

From the 1300 eligible MPBCS patients, a total of 1071 tumors were successfully characterized by gene expression microarrays and constituted the core of this work (Figure 1). A unique characteristic of this study was that each site performed gene expression analysis independently with their own equipment and personnel. The harmonization of all procedures related to sample obtention, storage, RNA extraction and microarray analysis was critical to minimize batch effects in microarray analysis. Using the PAM50 genes, principal component analysis (PCA) showed no significant bias by country, the arm of the study, type of sample (i.e., biopsy or surgical specimen) or year of microarray hybridization, confirming the data reliability (Supplementary Figure S1) and allowing to classify patients according to gene expression-derived PAM50 intrinsic subtypes without any batch correction.

PAM50 classification defined 456 tumors (42.6%) as luminal A (LumA), 228 (21.3%) as luminal B (LumB), 142 (13.3%) as HER2-enriched (HER2E) and 178 (16.6%) as basal-like (Basal). Sixty-seven tumors (6.3%) were classified as normal-like and were therefore excluded from some of the analyses. Normal-like tumors are considered tumors with very little tumor content and

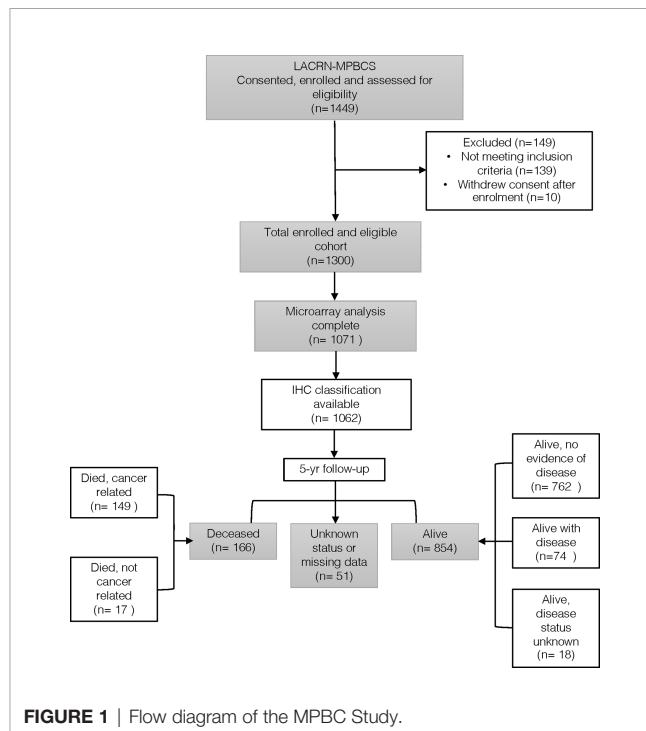


FIGURE 1 | Flow diagram of the MPBC Study.

with no prognostic value (17, 29). Moreover, as there is no surrogate category for Normal-like tumors, it was not possible to perform any comparison with other prognostic classifications.

Table 1 shows a summary of the clinical characteristics of the 1071 tumor samples and their distribution according to PAM50 subtypes. Age was significantly associated with the PAM50 subtype ($p < 0.001$), as a significantly lower mean age at diagnosis was seen for Basal (50.4 \pm 10.8 years) cases than for the rest of the subtypes (except normal-like). All clinical and pathological data were associated to a certain degree with PAM50 subtyping. As shown by the color shades in **Table 1**, histological type, clinical stage, tumor size and lymph node status, albeit significantly associated, showed relatively low correlations with PAM50 subtypes. On the other hand, tumors larger than 50 mm in size and/or with lymph node metastasis were enriched in non-LumA subtypes. IHC-based ER, PgR and FISH-based HER-2 status were strongly correlated with PAM50 subtypes while Ki67 values above the 20% threshold were moderately correlated with non-LumA subtypes (**Table 1**).

Comparative Prognostic Analysis of PAM50 and the IHC-Based Classifiers

IHC-based surrogate subtypes were calculated according to both classic IHC and a modification of the St Gallen 2013 criteria (St Gallen IHC). A total of 673/1062 samples (63.4%) were HR+/HER2-, of which 309 were classified as highly proliferative tumors (Ki67 > 20%) (**Figure 2A**). Of the remaining samples, 130 (12.2%) were HR+/HER2+, 89 (8.4%) were HR-/HER2+ and 170 (16.0%) were triple negative (TNBC) (**Figure 2A**).

Only ~60% of PAM50 assigned patients (622/995) could be classified into an equivalent IHC surrogate subtype (**Figure 2A**). The most consistent category was TNBC, which included 73.0% (130/178) of Basal cases (**Figures 2B, C**). The most heterogeneous subtype was the HR+ subtype, which included LumA, LumB and HER2E tumors in varying proportions (**Figure 2B, C**). HER2E tumors were distributed in almost equal proportions between HR+ and HR- classes (**Figures 2B, C**). The correlation between IHC and PAM50 subtypes was statistically significant ($p < 0.001$) with a Cramer's V of 0.57 and of 0.61 for classic IHC and St Gallen IHC, respectively (**Figures 2B, C**).

Survival curves according to the different classifications are shown in **Figure 3**, and their corresponding hazard ratios (HRs) against the class with the best survival are included in **Supplementary Table S2**. As all clinical and pathological variables showed associations with subtypes (**Table 1**), a univariate Cox model was selected for comparisons. Of note, multivariate analysis including age and clinical stage as covariables confirmed the prognostic value of subtypes (**Supplementary Table S2**).

PAM50 subtypes were prognostic for OSC and DFS (**Figure 3A**). Patients with LumA tumors had the best outcome, and those with Basal tumors had the worst outcome [HR 4.7 (confidence interval (CI) 3.0-7.5)]. Patients with either LumB [HR 2.5 (CI 1.6-4.1)] or HER2E [HR 2.7 (CI 1.6-4.7)] tumors showed similar intermediate survival. Consistent

with that, ROR-S clearly discriminated patient survival into high [HR 6.5 (CI 3.9-10.7)], intermediate [HR 3.5 (CI 2.0-5.8)] and low (reference) risk groups (**Figure 3B**). The ROR-S obtained the highest correlation index for both OSC and DFS (**Figure 3**).

Classic IHC was only able to discriminate TNBC patient outcome from the rest (**Figure 3C** and **Supplementary Table S2**). However, when Ki67% was used to distinguish highly proliferative HR+ tumors (St Gallen IHC), the difference in survival between HR+HER2-high Ki67 cases [HR 2.8 (CI 1.6-4.8)] and HR+HER2-low Ki67 cases became apparent (**Figure 3D** and **Supplementary Table S2**).

Enriched Pathways in the Intrinsic Subgroups of the MPBCS Cohort

For a deeper understanding of our Latin American cohort in molecular terms, we studied which biological pathways were enriched in each intrinsic subtype using different approaches. For enrichment analysis, we applied batch correction to the gene expression matrix, which accounted for a slight variation introduced by the year of array hybridization seen in the third PCA component (**Supplementary File 2** – Extended Methods). Both GSEA and GSVA showed a general concordance in significantly enriched pathways ($p < 0.05$) in the comparisons among subtypes of the MPBCS cohort (**Supplementary Table S1**). As an example, **Figure 4** shows the top ten significant terms found enriched by GSEA in each comparison. **Table 2** summarizes the most relevant enriched pathways in GSEA and/or GSVA according to the condition in which they showed positive enrichment.

Even when all genes evaluated were used for these enrichment analyses, the pathways that stood out in every contrast included those represented by the PAM50 genes. Indeed, all luminal contrasts were enriched mainly on the ER pathway activation. Basal tumors exhibited enrichment of terms related to proliferation and DNA replication and repair (including *TP53*- and *MYC*-related pathways) in all contrasts but LumB. Cell proliferation was not enriched in the LumB vs Basal or in LumB vs HER2E comparisons indicating that LumB tumors in our cohort do not differ from other aggressive tumors in terms of proliferation capacity (**Figure 4** and **Table 2**). Enhanced proliferation in LumB tumors was also the most significant distinction between LumB and LumA tumors, confirming the association of LumB with high Ki67 levels (**Figure 4** and **Table 2**).

Terms associated with fatty acid metabolism and peroxidation stood out as characteristics of the luminal phenotype, as were significant both in the LumA vs Basal and LumB vs Basal comparisons (**Figure 4** and **Table 2**).

Terms related to *ERBB2* signaling appeared significantly enriched with respect to both LumA and LumB only on GSVA (**Table 2** and **Supplementary Table S1**). The LumA vs HER2E comparison showed that HER2E tumors were more proliferative, while LumA were enriched in different growth factor pathways including the PI3K cascade. The comparison between LumB and HER2E indicated an enrichment of DNA repair pathways in the LumB subtype and several terms related to extracellular matrix

TABLE 1 | Summary of clinicopathological characteristics in the MPBCS cohort and their association with PAM50 subtypes[#].

Clinical parameter	Whole cohort n (%)	PAM50					Cramer's V/AOV
		LumA	LumB	Her2	Basal	Normal	
Number of subjects (%)	1071	456 (42.6%)	228 (21.3%)	142 (13.3%)	178 (16.6%)	67 (6.3%)	
Age at diagnosis - mean (SD)	54.5 (12.2)	56.4 (12.4)a	55.4 (13.4)a	54.3 (11)a	50.4 (10.8)b	50.3 (9.6)b	p < 0.001
Histological type							0.1 (low)
Invasive (infiltrating) lobular carcinoma	62 (5.8%)	40	10	2	2	8	
Invasive (infiltrating) ductal carcinoma	919 (85.8%)	372	197	134	162	54	
Invasive mammary carcinoma (NOS)	19 (1.8%)	10	6	0	3	0	
Invasive mixed ductal and lobular carcinoma	24 (2.2%)	13	7	0	1	3	
Other	47 (4.4%)	21	8	6	10	2	
Histological grade							0.26 (medium)
Low	128 (12%)	100	10	5	4	9	
Intermediate	473 (44.2%)	240	113	48	43	29	
High	431 (40.2%)	98	94	85	130	24	
Missing	39 (3.6%)	18	11	4	1	5	
Clinical stage							0.14 (low)
II A	383 (35.8%)	209	80	42	34	18	
II B	324 (30.3%)	137	69	41	57	20	
III A	199 (18.6%)	39	48	38	59	15	
III B	98 (9.2%)	40	21	14	15	8	
Missing/Other	67 (6.3%)	31	10	7	13	6	
Lymph node status							0.15 (low)
Positive	556 (51.9%)	191	120	84	123	38	
Negative	482 (45%)	250	102	52	50	28	
Missing	33 (3.1%)	15	6	6	5	1	
Tumor size at diagnosis							0.12 (low)
< = 20mm	113 (10.6%)	60	22	10	13	8	
> 20-50mm	666 (62.2%)	311	140	83	98	34	
> 50mm	257 (24%)	68	59	44	62	24	
Missing	35 (3.3%)	17	7	5	5	1	
ER status							0.52 (high)
Positive	800 (74.7%)	442	223	56	33	46	
Negative	269 (25.1%)	13	5	86	145	20	
Missing/Indeterminate	2 (0.2%)	1	0	0	0	1	
PgR status							0.45 (high)
Positive	670 (62.6%)	398	175	37	18	42	
Negative	397 (37.1%)	56	53	104	160	24	
Missing/Indeterminate	4 (0.4%)	2	0	1	0	1	
HER2 status							0.39 (high)
Positive	219 (20.4%)	45	33	107	15	19	
Negative	843 (78.7%)	407	191	35	163	47	
Missing/Equivocal	9 (0.8%)	4	4	0	0	1	
Ki67 (%) - median=30, range=0-100							0.32 (medium)
< = 20%	422 (39.4%)	284	54	38	19	27	
> 20%	607 (56.7%)	159	161	100	155	32	
Missing/Indeterminate	42 (3.9%)	13	13	4	4	8	

[#]The whole cohort distribution is shown as absolute number of patients and percentage of the total, while distribution according to subtypes are shown in number of patients. AOV: analysis of variance. For age at diagnosis among different PAM50 subtypes, categories with significant differences (Tukey HSD, alpha = 0.05) are indicated by a change on the superindex letter (i.e. a vs b are significantly different). For the remaining variables, the association between each of them and PAM50 subtypes was always significant (chi-squared test, p < 0.001 in all cases). Hue represents the sign of the standardized chi-squared residual; red hue indicates higher observed than expected counts, and blue indicates lower observed than expected counts. Color saturation represents the absolute value of the standardized chi-squared residuals; more saturation indicates a larger deviation from the expected counts. Low, medium and high refer to the strength of the association seen between subtypes and the clinicopathological characteristics. A Cramer's V value of 0.2 or less indicates a weak association, between 0.2 and 0.3 a moderate association and higher than 0.3 a strong association.

and cell adhesion enriched in HER2E. In general, both HER2E and Basal tumors showed upregulation of pathways associated with cell-cell communication when compared with luminal groups. The paired comparison between HER2E and Basal groups once again showed the strong activation of the cell cycle (also including the *TP53* and *MYC* pathways) seen in Basal tumors, while HER2E tumors showed enrichment of *ERBB2* and redox pathway terms in the GSEA analysis (Supplementary Table S1), as well as certain differential

hormone dependency indicated by the regulation of androgen response (Figure 4 and Table 2).

Analysis of the Activity of Transcription Factors (TF) Between PAM50 Subtypes

To search specifically for TF networks differentially activated in each PAM50 subtype comparison, we analyzed the MPBCS expression matrix using DoRothEA and then search for differentially activated TF between each unpaired subtype

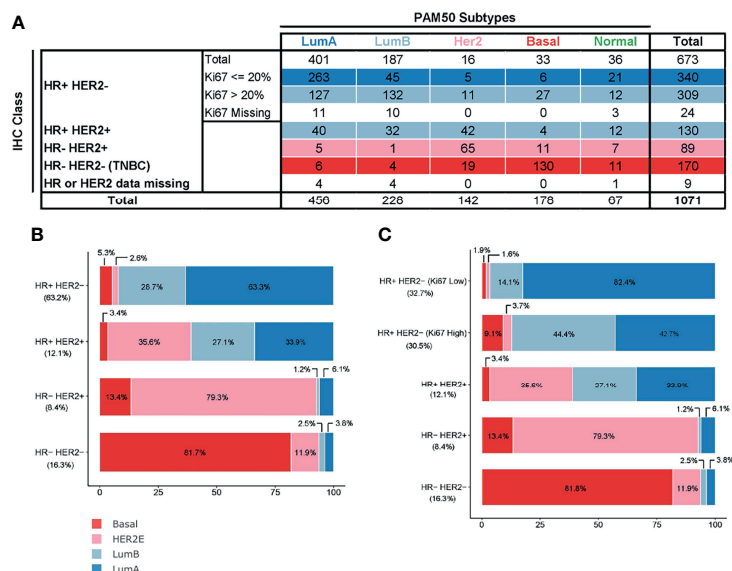


FIGURE 2 | Comparison of PAM50 subtypes (excluding normal-like) and surrogate, immunohistochemistry-based subtypes present in MPBCS patients.

(A) Frequency table for each category. Row colors represent the conventional correspondence between intrinsic and surrogate subtypes. **(B)** Classic IHC surrogate subtypes as defined by ER and PgR hormone receptors (HRs) and HER2 receptors ($n = 996$). **(C)** St Gallen IHC surrogate subtypes as defined by HR, HER2 receptors and high (more than 20%) or low (equal or less than 20%) levels of Ki67-positive cells. ($n = 975$). Patients with missing Ki67 values are excluded. Percentages in the x-axis and within each bar correspond to those of PAM50 subtypes in each surrogate group. Percentages in the y-axis correspond to the proportion of each surrogate subtype in the patients' total. The total number of patients used for each panel excludes patients with missing values in any of the IHC determinations and those who were labeled as normal-like by PAM50.

comparison. Among the most differentially regulated, **Figure 5** shows the E2F and MYC families of proliferation-related TF more activated in Basal samples in all comparisons. The JUN/FOS transcription complex, essential for MAPK-associated proliferation, is also prevalent in the more aggressive Basal and HER2E tumors. On the contrary, *PGR*, *ESR1* and *AR* transcriptional activities are enhanced in hormonal-driven LumA and LumB tumors. *FOXA1* and *GATA3* TF are both activated in luminal subtype, as expected, given that both *FOXA1* and *GATA3* are part of the luminal molecular signature. *NFKB1*, a master TF of aggressiveness and inflammation, and its counterpart *RELA* are highly activated in Basal and HER2E samples. STAT networks, involved in antigen presentation and immune response, are also activated in Basal samples, as well as *TWIST*, a TF associated to the epithelial-mesenchymal transition.

Subtype-Specific Enriched Pathway Comparison Between MPBCS and TCGA

To further search for unique molecular characteristics among subtypes of the Latin American cohort, we performed the GSEA and GSVA contrasts between PAM50 subtypes with the publicly available data from TCGA and then compared the enriched pathways found in each cohort. For this comparison we selected only stage II-III TCGA breast cancer patients with PAM50 subtype assignment (TCGA r_f) that rendered 19547 genes in common with our dataset. As expected, we observed a high degree of concordance between cohorts (**Supplementary Figure S2**). The NES for significantly enriched terms in MPBCS

correlated in general with those seen for the same terms in TCGA r_f , in all contrasts. Only a few isolated significant terms showed an opposite sign (**Supplementary Figure S2**, dots in red, term names within each graphic). The full list of TCGA enriched pathways and their NES scores with p-values is also included in **Supplementary Table S1**.

Differential Analysis of Components of the Immune System

MetaCore pathway analysis of the two cohorts using DEGs showed additional significant information beyond those already found by the gene set enrichment analyses, such as the enrichment in cell-surface receptors associated with the crosstalk between the epithelial malignant cell compartment and the tumor stroma (e.g., IGF1-R, FGF3-R, HGF-R, IBP-4) in LumB tumors and the activation of Wnt signaling in HER2E, with overexpression of several members of the Wnt family and Frizzled. Of note, terms related to the most relevant innate and adaptive immune functions dominated in all Basal comparisons only when using DEGs exclusive of the MPBCS cohort and not among the TCGA DEGs (**Supplementary Table S3**). Indeed, among those MPBCS DEGs that were not found differentially expressed in the TCGA contrasts we could identify several genes encoding proteins involved in immunological synapses, such as the costimulatory CD40 (*CD40* gene), the immunosuppressive PD-L1 (*CD274* gene), and members of the signaling lymphocyte activation molecule (SLAM) family, which were elevated in Basal tumors compared to LumA tumors. In this comparison, genes

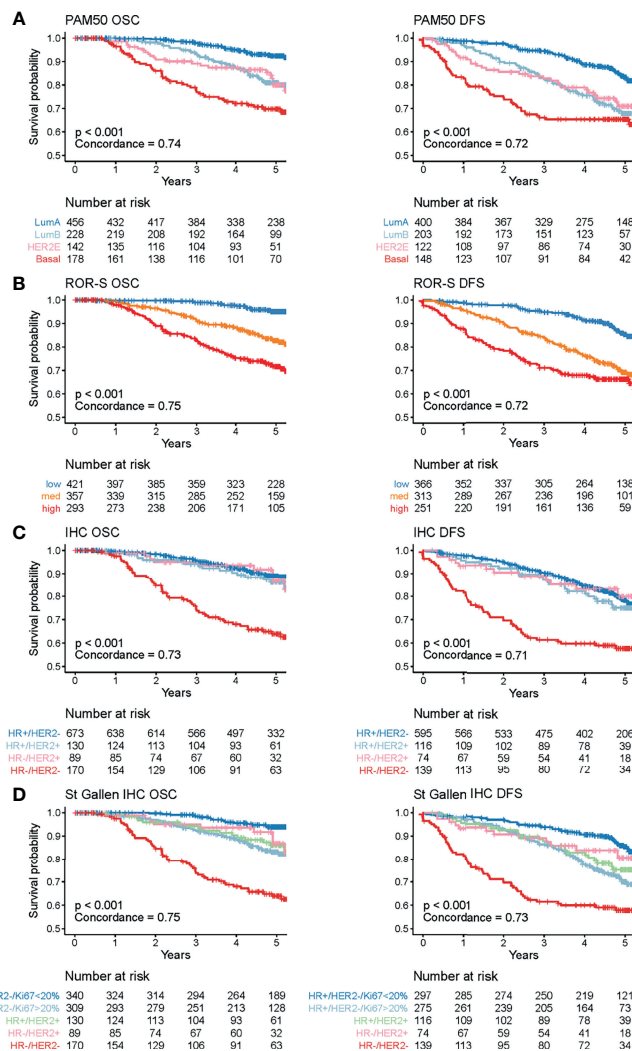
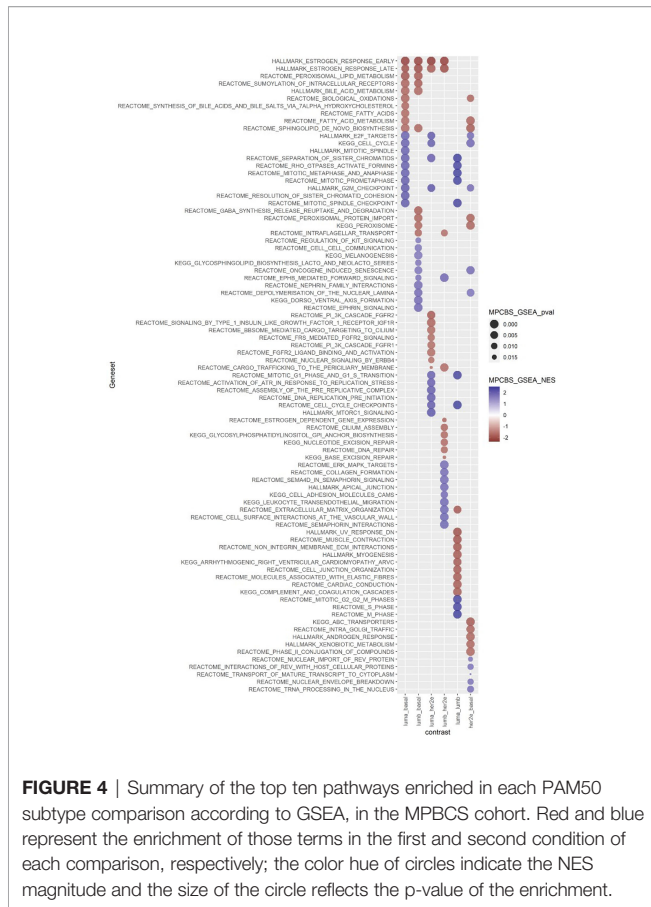


FIGURE 3 | Cancer-related survival (OSC) and disease-free survival (DFS) of the MPBCS cohort according to different classifications. Each graph shows the Kaplan-Meier survival curves for each classification. **(A)** PAM50, **(B)** ROR-S, **(C)** classic IHC, **(D)** St Gallen IHC. Each class color is defined at the table of number at risk, below each graph. The p-value included in the graphs corresponds to the significance value of the log-rank test for all groups. Correlation (i.e., C-index) for each of the classifiers are also shown.

related to T-cell and B-cell modulation such as IL-12, IL-21, TNF β , and IFN γ were also overexpressed in Basal tumors, as well as several proinflammatory chemokines (e.g., GRO-1 and 2, CCL7, IL-6) and chemokine receptors (e.g., CCR5, CXCR3) associated with lymphocyte and neutrophil presence and/or activity. Furthermore, a family of histones identified by Metacore as relevant in the formation of neutrophil extracellular traps (NETs) were upregulated in proliferative tumors compared to LumA tumors, with their expression highest in LumB tumors.

As several of the MetaCore immune-related differential terms were found in more than one paired comparison, we selected only the MPBCS DEGs contained in those terms and used them for a hierarchical clustering (**Figure 6**). Two main clusters of genes become apparent, a smaller one with genes overexpressed

in luminal samples (Cluster 1, **Figure 6**) and one with genes overexpressed in Basal and most HER2E samples (Cluster 2, **Figure 6**). Genes belonging to the adaptive immune system were almost exclusively in the second cluster with high expression in Basal and HER2E samples. On the other side, different set of genes involved in inflammatory pathways were overexpressed along all subtypes (Clusters 1 and 2, **Figure 6**). However, those with broader activities, that is, with important non-immune functions as well, were the ones enriched in LumA and LumB tumors (Cluster 1, **Figure 6**). Among them we found *PRRT2*, a tumor suppressor gene that inhibits proliferation, and genes involved in the DNA damage response and/or growth factor signaling cascades (e.g. *CDKN1B*, *MAPK8*, *AKT1* and *WNT4*). Of note, immune-related pathways were seldom observed among the top ten TCGA-exclusive differential terms of any contrast



(Supplementary Table S3), although the presence of the adoptive immune system in TCGA and other breast cancer cohorts has been widely reported.

The cytolytic score (CYT) has been used as indicator of the presence of tumor infiltrating lymphocytes (TILs). We calculated the CYT for all samples, and their distribution among the different subtypes is shown in **Figure 7A**. Significantly higher CYTs were seen in Basal samples, followed by HER2E samples. Luminal A and B samples have the lowest scores, although the shape of the boxplot may indicate a bimodal distribution in those samples. Next, we explored if tumors with high CYT (defined as CYTs above the median) showed differences in survival compared to those with low CYTs. At the level of the whole MPBCS cohort, survival analysis showed a slight tendency for better OSC and DFS in those patients with high CYTs, although the difference was not significant (not shown). When Basal and LumA samples were separately analyzed (with their own median value as threshold), CYT did not show OSC or DFS prognostic value in LumA samples (**Figure 7B, D**), while a tendency (still non-significant) for better prognosis with higher CYTs was seen for Basal samples (**Figures 7C, E**). The same by-subtype tendency but with worse p-values was seen using the whole cohort median as the threshold of low and high CYTs (not shown).

In recent years several molecular signatures exclusive for TNBC has been described, most of them converging in discrete molecular subtypes that differ in their expression of proliferation, immune and mesenchymal genes. We have applied Lehmann’s signature to classify the MPBCS TNBC cohort in the original 6 subgroups (TNBC-6) (25), in Lehmann’s refined TNBC-4

TABLE 2 | Summary of top enriched gene sets by GSEA and GSVA (using all genes) in the MPBCS cohort.

	Enriched in first condition	Enriched in second condition
LumA vs Basal	Estrogen response Peroxisoma/fatty acid metabolism Synthesis of glycosylphosphatidylinositol Biological oxidations	Cell cycle and mitosis (including TP53 activity) DNA replication MYC targets
LumB vs Basal	Estrogen response Peroxisoma/fatty acid metabolism Sumoylation of intracellular receptors Biological oxidations	Control of morphogenic processes Cell-cell communication Signaling by EGFR
LumA vs HER2E	Estrogen response Growth factor signaling (ERBB4, FGFR, ILGF, PI3K cascades)	Cell cycle and mitosis (including TP53 activity) DNA replication MYC targets PTEN regulation of stability and activity
LumB vs HER2E	Estrogen response DNA repair Synthesis of glycosylphosphatidylinositol	Extracellular matrix organization/collagen assembly Signaling by ERBB2 Signaling by EGFR MAPK signaling pathway Cell adhesion and migration
LumA vs LumB	Extracellular matrix organization Cell-cell and cell-matrix adhesion PI3K-AKT signaling in cancer MAPK signaling pathway Myogenesis and muscle contraction	Cell cycle and mitosis (including TP53 activity) DNA replication MYC targets
HER2E vs Basal	Peroxisome/Fatty acid metabolism/steroid biosynthesis Biological oxidations Xenobiotic metabolism Downregulation of ERBB2 signaling Androgen response	Cell cycle and mitosis (including TP53 activity) MYC targets Intracellular transport

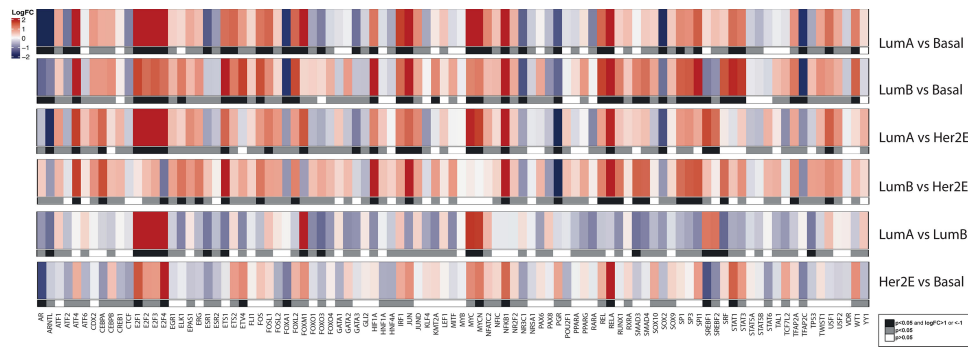


FIGURE 5 | Differentially activated TF among PAM50 subtypes in the MPBCS cohort. The differential activities found for each TF (columns) in each PAM50 subtype contrast (rows) are shown in this heatmap. Z-scores are shown for the second term in the comparison (e.g. the red hue in the first row correspond to a TF activated in Basal samples). Below each row of the heatmap, a square in a gradient from white to black represent the p-value and logFC of the differential activity for each TF in each comparison; white correspond to non-significant comparisons.

subtypes (26) and in Lehmann’s derived Bareche-2018 five-class groups (27). **Supplementary Figure S3** shows the relative distribution of subtypes within TNBC for each molecular signature (A), and their corresponding OSC (B) and DFS (C) survival curves. Confirming the results obtained with other cohorts, no significant differences could be found among any of the subtypes in any of the versions of the signatures (26).

DISCUSSION

The present study is the first multicountry, multicentric effort in Latin America, in which each participating institution generated a local biobank, performed molecular profiling, and registered its own clinical and transcriptional data by following strict consensus protocols and standard operating procedures.

This resulted in a homogeneous dataset of high-quality, state-of-the-art clinical, pathological, and molecular data with minimum interinstitutional variation and a low percentage of missing data. This model of collaboration should be encouraged to lead to more diverse databases.

The clinicopathological and molecular features of the MPBCS cohort appear similar to those described in other hospital-based Latin American studies (8, 30–34). Non-LumA intrinsic subtypes accounted for more than half of the tumors in the MPBCS cohort, indicating the enrichment of this cohort in more aggressive molecular subtypes. However, compared with the LACE cohort, in which Sweeney et al. (8) analyzed the distribution of PAM50-derived subtypes in the Hispanic/Latino cases, the relative distribution of non-LumA subtypes within the MPBCS cohort slightly favored the Basal subtype over the HER2E cases and LumB cases (i.e., Basal/HER2E ratio 1.2 in MPBCS vs 0.7 in Hispanic LACE; Basal/LumB ratio 0.8 in

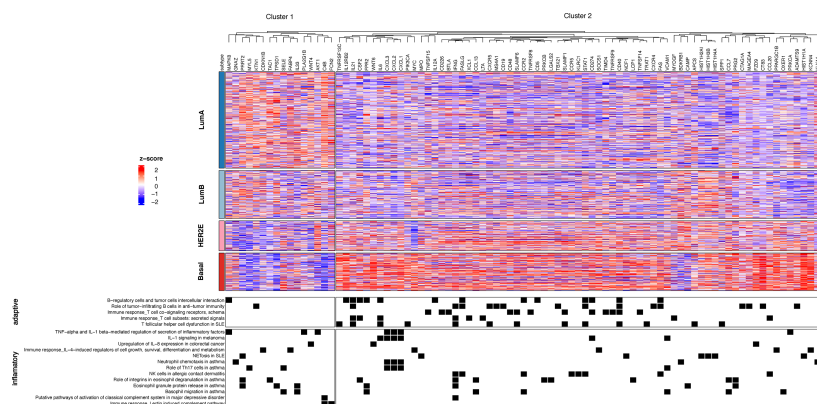


FIGURE 6 | Clustering of differentially expressed genes related to immune terms in all PAM50 subtypes of the MPBCS cohort. The top panel shows the heatmap derived from a hierarchical clustering ($k = 2$); genes included in the enriched MetaCore pathways are in columns and tumors (grouped by PAM50 subtypes) are in rows. The bottom panel lists the immune-related enriched MetaCore pathways; black squares indicate the presence of a selected gene in any of the MetaCore terms.

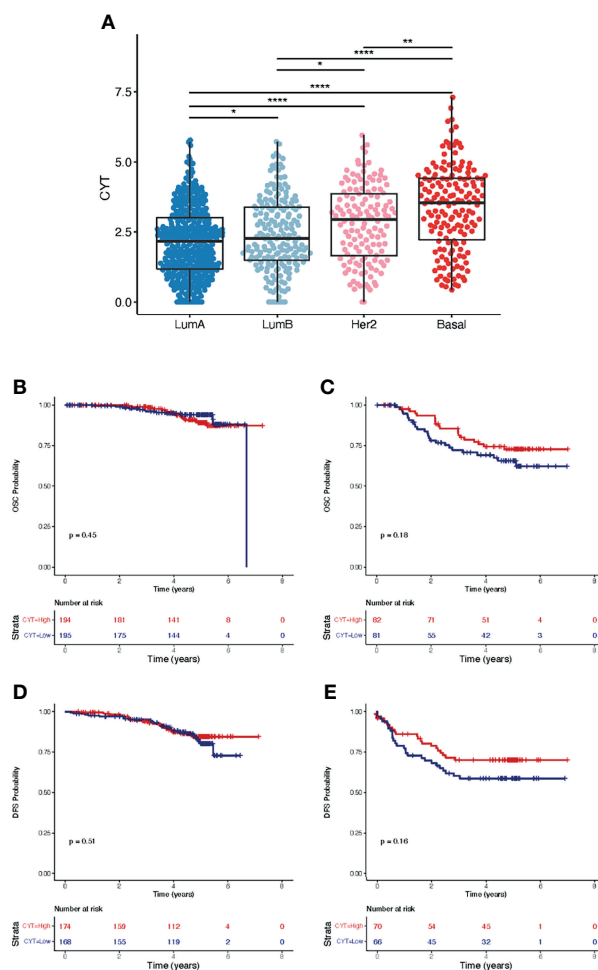


FIGURE 7 | Cytolytic score CYT and survival analysis according to PAM50 subtypes in the MPBCS cohort. **(A)**, distribution of CYT among PAM50 subtypes. Medians were compared using the Kruskal-Wallis test ($p=2.2e-16$). Adjusted p -values for the paired comparisons using Dunn's test are defined as follows: * <0.05 , ** <0.01 , **** <0.0001 . **(B, C)**, Kaplan-Meier curves and survival analysis for cancer-related overall survival (OS) according to a binary assignment of high (i.e. above median) and low CYT to LumA **(B)** and Basal **(C)** tumors. **(D, E)**, Kaplan-Meier curves and survival analysis for disease-free survival (DFS) according to the same binary assignment of high (i.e. above median) and low CYT to LumA **(D)** and Basal **(E)** tumors.

MPBCS vs 0.5 in Hispanic LACE). Although this shift towards the Basal subtype most likely reflects the inclusion criteria of locally advanced stage (II/III) tumors (5, 35), we cannot disregard the fact that the diversity of our cohort might mask the effect of the ancestry in this distribution. Further work on the MPBCS will address this point.

For the first time, the 5-year prognostic ability of PAM50 could be established in a large Latin American-based cohort. The prognostic power of PAM50 subtypes is compatible with what has been reported for node-negative and node-positive disease in other cohorts (17, 36–38). Albeit the prognostic power of IHC St Gallen for all the survival variables is acceptable and the C-indexes are comparable to those of PAM50, PAM50 still performs better at discriminating risk, as was already described for other cohorts (37, 39). Furthermore, PAM50-derived ROR-S (36) successfully classifies tumors into low, intermediate and

high-risk groups with an even better performance than PAM50. Ongoing studies are focused in the exploration of other factors (40) and molecular signatures (37, 41) that can be used for further improving the prognostic value in this cohort.

Gene expression-based pathway analysis showed a general concordance of differentially expressed pathways between intrinsic subtypes with those previously described in the literature (42–45). GSEA/GSVA analyses prove that the major molecular characteristics of stage II and III breast cancer in patients from this Latin American cohort are similar to those of stage-matched TCGA patients. Proliferation, cell cycle control and DNA damage repair pathways were related to LumB, HER2E and Basal tumors, while LumA tumors showed a strong dependence on the estrogen pathway activation. Master TF activated in each contrast analysis were consistent with these main driver pathways, as already noted in the literature (42).

Currently there is a general consensus in that each breast cancer intrinsic subtype represent an average phenotype resulting from an heterogeneous composition of cells within a tumor, and that the aggressiveness of a breast tumor is dictated by the different combination of cells present in the tumor (46–48). For this cohort, our multiple molecular analyses support this hypothesis and acknowledges the continuum of molecular phenotypes that coexists within each intrinsic subtype.

Among the findings that we consider worth to further explore, we observed upregulated fatty acid metabolic pathways in HR+ tumors. This finding, although already noted by the seminal work of Perou et al. (49) and others, can be reanalyzed in the light of recent studies that named fatty acid metabolism as a critical pathway related to cancer cell capacity to metastasize (50). The reprogramming of lipid metabolism is a hallmark of many cancers, including breast cancer (51). Our evidence points to the capacity of HR+ tumors for *de novo* fatty acid synthesis, mobilization and oxidation, in contrast to TNBC, which is more dependent on the uptake and storage of exogenous fatty acids (52). Further analysis will be done to explore the relationship between obesity, fatty acid metabolism and recurrence in luminal tumors, as the epidemiological and ancestral characteristics of the MPBCS cohort may bear subtler differences with other cohorts.

Interestingly, MetaCore analyses highlighted GO terms related to inflammatory/immune response that were only prominent in the MPBCS cohort and not in the TCGA samples; nevertheless, this is not a unique feature of MPBCS, as other more specific immunogenomic analysis of the TCGA database has already associated most Basal and some of HER2+ and LumB breast cancers with an IFN- γ dominant class, with a high content of CD8-T cells and of TCR diversity (53). The strong presence of the adaptive immune system in Basal or TNBC tumors has been extensively described in the literature and was linked to the presence of tumor-infiltrating lymphocytes (26, 27, 54–56). The score CYT described for our cohort is also consistent with these observations. Although underpowered, the tendency of high-CYT Basal (but not luminal) tumors to show better prognosis in the MPBCS cohort is consistent with previous evidence (57). Moreover, our evidence seems to be in line with the value of TIL content as a good prognostic marker in TNBC but not in luminal tumors (58, 59). These observations stress the importance of understanding the role of the immune response in the context of the different driving processes of tumor progression and therapy response in luminal and basal breast cancer.

CYT was also proposed as biomarker of response to immunotherapy in several tumors (60), although it is still not recognized as such in breast cancer (61). Another feature seen in our Basal samples, the expression of PD-L1, has been approved as biomarker for immune checkpoint blockade in TNBC (61, 62) although it turned out to be ineffective in certain settings (63). New prospective studies built under the example of the MPBCS might explore if the combination of both CYT and PD-L1 expression can be used as better predictive biomarkers of response to immunotherapy in Latin American TNBC patients.

Different TNBC classifications have been developed and their prognostic and predictive power studied (26, 27, 54, 64, 65). None of them have demonstrated a clear, robust prognostic value (26), although they have recently been shown as useful for predicting response to neoadjuvant chemotherapy in the BrightTNess trial (66). Our work proved that the classification of TNBC tumors into subtypes in the MPBCS cohort was feasible, although our study was not designed for detecting differences in survival or treatment response according to TNBC subtypes and it is inherently underpowered for these purposes.

Our TF analysis is consistent with the already observed *FOXA1*-driven immunosuppression in luminal samples (67) and the *NFKB1/STAT*-driven immune response in the Basal and, to a lesser extent, HER2E tumors (68). Other immune characteristics of our cohort are supported with less evidence in the literature but are worth to be studied further. For example, the literature suggest that B-cell markers in TNBC could be used as indicators of good prognosis (69) or that tumor-associated neutrophils (TANs) can be correlated with poor prognosis (70). In particular, neutrophil activity in NETs has been associated with sterile inflammation and tumor aggressiveness in breast cancer (71).

Interestingly, some recent reports addressed putative ancestral differences in the immune environment of TNBC tumors, with mixed results according to race (72, 73). The present immune-related findings highlight the quality of the MPBCS transcriptomic data and, along with the molecular ancestry assignment already achieved in this cohort, pave the way for deeper analyses in the search for better therapeutic strategies for Latin American patients.

This study has some limitations. As already mentioned, the hospital-based characteristic and the bias towards advanced stages of this cohort prevent the evaluation of the actual subtype distribution, as could be done in a population-based cohort. It is possible that some variations in prognostic power may be seen when by-country, by-treatment and/or by-ancestry stratification are done. Ongoing studies are based on the hypothesis that a fine tuning of the differential molecular pathways associated with tumor growth and aggressiveness may be found when stratified sample analysis by genetic ancestry is performed.

Our first studies on the high-quality data of the MPBCS cohort have given long-sought evidence that the main gene-expression features of breast cancer described for other cohorts are roughly conserved across Latin American patients and are informative on prognosis. Further analysis will explore potential country- and/or ancestry-specific differences in prognosis and treatment outcomes in this cohort.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from US-LACRN but restrictions apply to the availability of these data.

Data are however available from the authors upon reasonable request and with permission of US-LACRN.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by National Cancer Institute Ethics Committee (IRB #: 15CN055; iRIS Reference #: 549650) and the local institutional committees in each country. The patients provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization and design, JG, OP, EA, AL, NA, AD-N, BM, CV, IA, RB, AB, JF, WF, FG, MG, LS, NC, DC, MC, SCA, Aca, SCr, WF, FG, MG, MN, RMR, RR, LD, MD-B, MH, VV, and TG. Supervision, OP, AL, NA, AD-N, BM, CV, EBA, IA, AB, JF, NC, DC, MC, SCA, Aca, SCr, WF, FG, MG, MN, RMR, RR, JC-C, LS, ACo, RD-C, RF-T, AM-M, AO-V, JR, CR, ES-G, AS-G, LZ, AD-A, AQ-R, LD, MD-B, JG, TG, MH, VV, OV, MO-M, ER-C, Mce, AG-E, EAF, and EA. Data collection, AL, NA, AD-N, BM, CV, EBA, IA, RB, JF, AB, JS, NC, DC, MC, SCA, Aca, SCr, WF, FG, MG, MN, RMR, RR, LS, JC-C, ACo, RD-C, RF-T, CG, SM, AM-M, AO-V, JR, CR, VS, ES-G, AS-G, AT, LZ, AD-A, AQ-R, MO-M, JQ, ER-C, Mce, AL-V, EAF, GG, and EA. Bioinformatic and statistical analysis, DR, JS, Mce, RB, EA, EAF, GG, ML-M, DA, and AL. Visualization, DR, JS, RB, EAF, and AL. Data interpretation, DR, JS, RB, EAF, LF, NA, OP, EA, JF, Mce, AD-A, AQ-R, and AL. Writing, DR, AL, EA, and OP. Reviewing and editing, LF, OP, AL, NA, AD-N, BM, CV, JF, EBA, IA, AB, DR, NC, DC, MC, SCA, Aca, SCr, WF, FG, MG, MN, RMR, JC-C, RR, LS, ACo, RD-C, RF-T, AM-M, AO-V, JR, CR, ES-G, AS-G, LZ, AD-A, AQ-R, LD, MD-B, JG, TG, MH, VV, OV, JQ, JS, Mce, EAF, AG-E, GG, AL-V, and EA. Funding acquisition, JG, TG, VV, MH, BM, NA, EA, AD-N, CV, MD-B, AL, DR, JS, AL, EA, and OP have verified the underlying data. All authors contributed to the article and approved the submitted version.

FUNDING

Center for Global Health at the United States - National Cancer Institute at the National Institutes of Health (CGH-US(NCI), (contract award # HHSN2612010000871/NO2-PC-2010-00087), Fogarty International Center, Susan G. Komen for the Cure, Argentina, Fundación Argentina de Nanotecnología at the Ministry of Science, Technology, and Innovation, Agencia Nacional de Promoción Científica y Tecnológica, Instituto Nacional del Cáncer at the Ministry of Health, Brazil, Ministry of Health, Chile, Public Health Institute at the Ministry of Health, Mexico, Consejo Estatal de Ciencia y Tecnología de

Jalisco (COECYTJAL) (no grant number applies, Universidad de Sonora. The CGH-NCI(US) had a role in study design and data curation. The rest of the sponsors have no role in the study design, nor in the collection, analysis, and interpretation of data, the writing of the report and the decision to submit the paper for publication.

ACKNOWLEDGMENTS

We thank all personnel from Westat, in particular Vanessa Olivo, Kerry Grace Morrissey and Kathy Clingan, for their hard work coordinating data curation. We also acknowledge the personnel from Leidos, in particular Mariana González del Riego and Silvina Frech during the step of US-LACRN capacity building. We thank the enormously valuable help of Dr Mark Cosentino in biobank setup, Dr James Robb in pathology good practices and Dr Katherine Hoadley in microarray analysis and PAM50 classification. We finally want to thank all the patients that took part of this study for their generous collaboration with data and biospecimens and their commitment to the advance of breast cancer knowledge in Latin America.

We dedicate this work to Dr Ignacio Muse, the first Uruguay US-LACRN coordinator who passed away in 2012, and to the memory of Dr Victor Jimenez-Moreno, from Guadalajara, Mexico.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.835626/full#supplementary-material>

Supplementary File 1 | MPBCS Protocol.

Supplementary File 2 | Extended Methods.

Supplementary Figure 1 | Principal component analysis (PCA) from PAM50 gene expression data.

Supplementary Figure 2 | Concordance of GSEA normalized enrichment scores (NES) in significantly enriched gene sets of MPBCS and TCGA^{rf} stage II-III breast cancer tumors.

Supplementary Figure 3 | Prognostic analysis of different TNBC subtypes (TNBC-6 and its derived TNBC-4 and Bareche-2018) in the MPBCS cohort.

Supplementary Table 1 | GSEA and GSVA enriched terms for each subtype contrast in the MPBCS cohort.

Supplementary Table 2 | Hazard ratios and statistical significance of Cox survival models for each prognostic classification tested.

Supplementary Table 3 | MetaCore top ten pathways for each subtype contrast performed with all differentially expressed genes in common between TCGA and MPBCS, exclusively in MPBCS and exclusively in TCGA.

REFERENCES

- Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell* (2020) 37(5):639–54.e6. doi: 10.1016/j.ccell.2020.04.012
- Curtis C, Shah SP, Chin S-FF, Turashvili G, Rueda OM, Dunning MJ, et al. The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups. *Nature* (2012) 486(7403):346–52. doi: 10.1038/nature10983
- DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. *Cancer Epidemiol Biomark Prev* (2015) 24(10):1495–506. doi: 10.1158/1055-9965.EPI-15-0535
- Goss PE, Lee BL, Badovinac-Crnjevic T, Strasser-Weippl K, Chavarri-Guerra Y, Louis JS, et al. Planning Cancer Control in Latin America and the Caribbean. *Lancet Oncol* (2013) 14(5):391–436. doi: 10.1016/S1470-2045(13)70048-2
- Serrano-Gómez SJ, Fejerman L, Zabaleta J. Breast Cancer in Latinas: A Focus on Intrinsic Subtypes Distribution. *Cancer Epidemiol Prev Biomark* (2017) 27(1):cebp.0420.2017. doi: 10.1158/1055-9965.EPI-17-0420
- Zavala VA, Serrano-Gomez SJ, Dutil J, Fejerman L. Genetic Epidemiology of Breast Cancer in Latin America. *Genes (Basel)* (2019) 10(2):153. doi: 10.3390/genes10020153
- Stern MC, Fejerman L, Das R, Setiawan WV, Cruz-Correa MR, Perez-Stable EJ, et al. Variability in Cancer Risk and Outcomes Within US Latinos by National Origin and Genetic Ancestry. *Curr Epidemiol Rep* (2016) 3(3):181–90. doi: 10.1007/s40471-016-0083-7
- Sweeney C, Bernard PS, Factor RE, Kwan ML, Habel LA, Quesenberry CP, et al. Intrinsic Subtypes From PAM50 Gene Expression Assay in a Population-Based Breast Cancer Cohort: Differences by Age, Race, and Tumor Characteristics. *Cancer Epidemiol Prev Biomark* (2014) 23(5):714–24. doi: 10.1158/1055-9965.EPI-13-1023
- Investigators of the US–Latin America Cancer Research Network. Translational Cancer Research Comes of Age in Latin America. *Sci Transl Med* (2015) 7(319):319f50. doi: 10.1126/scitranslmed.aad5859
- Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College Of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *J Clin Oncol* (2010) 28(16):2784–95. doi: 10.1200/JCO.2009.25.6529
- Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and Predictive Factors in Breast Cancer by Immunohistochemical Analysis. *Mod Pathol An Off J United States Can Acad Pathol Inc* (1998) 11(2):155–68.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The Cbio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095
- Zhu Y, Qiu P, Ji Y. TCGA-Assembler: Open-Source Software for Retrieving and Processing TCGA Data. *Nat Methods* (2014) 11(6):599600. doi: 10.1038/nmeth.2956
- Robinson MD, Oshlack A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol* (2010) 11(3):R25. doi: 10.1186/gb-2010-11-3-r25
- Law CW, Chen Y, Shi W, Smyth GK. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol* (2014) 15(2):1–17. doi: 10.1186/gb-2014-15-2-r29
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
- Parker JS, Mullins M, Cheang M, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J Clin Oncol* (2009) 27(8):1160–7. doi: 10.1200/JCO.2008.18.1370
- Gonzalez G, Fresno C, Merino G, Llera AS, Podhajcer O, Fernández E. Agi4x44.2c: A Two-Colour Agilent 4x44 Quality Control R Library for Large Microarray Projects. *Ann Terc Congr Argentino Bioinformática Y Biol Comput* (2012).
- Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: A Desktop Application for Gene Set Enrichment Analysis. *Bioinformatics* (2007) 23(23):3251–3. doi: 10.1093/bioinformatics/btm369
- Hänzelmann S, Castelo R, Guinney J. GSEA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
- Gu Z, Eils R, Schlesner M. Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics* (2016) 32(18):2847–9. doi: 10.1093/bioinformatics/btw313
- García-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodríguez J. Benchmark and Integration of Resources for the Estimation of Human Transcription Factor Activities. *Genome Res* (2019) 29(8):1363–75. doi: 10.1101/gr.240663.118
- García-Alonso L, Iorio F, Matchan A, Fonseca N, Jaaks P, Peat G, et al. Transcription Factor Activities Enhance Markers of Drug Sensitivity in Cancer. *Cancer Res* (2018) 78(3):769–80. doi: 10.1158/0008-5472.CAN-17-1679
- Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and Genetic Properties of Tumors Associated With Local Immune Cytolytic Activity. *Cell* (2015) 160(1–2):48–61. doi: 10.1016/j.cell.2014.12.033
- Chen X, Li J, Gray WH, Lehmann BD, Bauer JA, Shyr Y, et al. TNBCtype: A Subtyping Tool for Triple-Negative Breast Cancer. *Cancer Inform* (2012) 11:147–56. doi: 10.4137/CIN.S9983
- Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS One* (2016) 11(6):e0157368. doi: 10.1371/journal.pone.0157368
- Bareche Y, Venet D, Ignatiadis M, Aftimos P, Piccart M, Rothe F, et al. Unravelling Triple-Negative Breast Cancer Molecular Heterogeneity Using an Integrative Multiomic Analysis. *Ann Oncol Off J Eur Soc Med Oncol* (2018) 29(4):895–902. doi: 10.1093/annonc/mdy024
- Cohen J. Set Correlation and Contingency Tables. *Appl Psychol Meas* (1988) 12(4):425–34. doi: 10.1177/014662168801200410
- Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S, et al. Development and Verification of the PAM50-Based Prosigna Breast Cancer Gene Signature Assay. *BMC Med Genomics* (2015) 8(1):54. doi: 10.1186/s12920-015-0129-6
- Zavala VA, Bracci PM, Carethers JM, Carvajal-Carmona L, Coggins NB, Cruz-Correa MR, et al. Cancer Health Disparities in Racial/Ethnic Minorities in the United States. *Br J Cancer* (2021) 124:315–32. doi: 10.1038/s41416-020-01038-6
- Martinez M, Wertheim BC, Natarajan L, Schwab R, Bondy M, Daneri-Navarro A, et al. Reproductive Factors, Heterogeneity, and Breast Tumor Subtypes in Women of Mexican Descent. *Cancer Epidemiol Prev Biomark* (2013) 22(10):1853–61. doi: 10.1158/1055-9965.EPI-13-0560
- Caglevic C, Anabalón J, Soza C, Milla E, Gaete F, Carrasco, et al. Triple-Negative Breast Cancer: The Reality in Chile and in Latin America. *Ecancermedicalscience* (2019) 13:893. doi: 10.3332/ecancer.2019.893
- Serrano-Gomez SJ, Sanabria-Salas MC, Hernández-Suarez G, García O, Silva C, Romero A, et al. High Prevalence of Luminal B Breast Cancer Intrinsic Subtype in Colombian Women. *Carcinogenesis* (2016) 37(7):669–76. doi: 10.1093/carcin/bgw043
- de Macêdo Andrade AC, Ferreira Júnior CA, Dantas Guimaraes B, Pessoa Barros AW, de Almeida GS, Weller M. Molecular Breast Cancer Subtypes and Therapies in a Public Hospital of Northeastern Brazil. *BMC Womens Health* (2014) 14(1):1–9. doi: 10.1186/1472-6874-14-110
- Howlander N, Cronin KA, Kurian AW, Andridge R. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer Epidemiol Biomark Prev* (2018) 27(6):619–26. doi: 10.1158/1055-9965.EPI-17-0627
- Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A Comparison of PAM50 Intrinsic Subtyping With Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor–Positive Breast Cancer. *Clin Cancer Res* (2010) 16(21):5222–32. doi: 10.1158/1078-0432.CCR-10-1282
- Lundberg A, Lindström L, Harrell CJ, Falato C, Carlson JW, Wright PK, et al. Gene Expression Signatures and Immunohistochemical Subtypes Add Prognostic Value to Each Other in Breast Cancer Cohorts. *Clin Cancer Res* (2017) 23(24):clincanres.1535.2017. doi: 10.1158/1078-0432.CCR-17-1535
- Ren JX, Gong Y, Ling H, Hu X, Shao ZM. Racial/ethnic Differences in the Outcomes of Patients With Metastatic Breast Cancer: Contributions of

- Demographic, Socioeconomic, Tumor and Metastatic Characteristics. *Breast Cancer Res Treat* (2019) 173(1):225–37. doi: 10.1007/s10549-018-4956-y
39. Cejalvo J, Pascual T, Fernández-Martínez A, Brasó-Maristany F, Gomis RR, Perou CM, et al. Clinical Implications of the non-Luminal Intrinsic Subtypes in Hormone Receptor-Positive Breast Cancer. *Cancer Treat Rev* (2018) 67:63–70. doi: 10.1016/j.ctrv.2018.04.015
 40. Kim H, Park K, Kim Y, Park S, Lee H, Lim S, et al. Discordance of the PAM50 Intrinsic Subtypes Compared With Immunohistochemistry-Based Surrogate in Breast Cancer Patients: Potential Implication of Genomic Alterations of Discordance. *Cancer Res Treat* (2018) 51(2):737–47. doi: 10.1007/s10549-018-4858-z
 41. Pareja F, Reis-Filho JS. Triple-Negative Breast Cancers — A Panoply of Cancer Types. *Nat Rev Clin Oncol* (2018) 15(6):1. doi: 10.1038/s41571-018-0001-7
 42. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* (2012) 490(7418):61. doi: 10.1038/nature11412
 43. Abba MC, Lacunza E, Butti M, Aldaz CM. Breast Cancer Biomarker Discovery in the Functional Genomic Age: A Systematic Review of 42 Gene Expression Signatures. *Biomark Insights* (2010) 5:103–18. doi: 10.4137/BMI.S5740
 44. Fan C, Prat A, Parker JS, Liu Y, Carey LA, Troester MA, et al. Building Prognostic Models for Breast Cancer Patients Using Clinical Variables and Hundreds of Gene Expression Signatures. *BMC Med Genomics* (2011) 4:3. doi: 10.1186/1755-8794-4-3
 45. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological Processes Associated With Breast Cancer Clinical Outcome Depend on the Molecular Subtypes. *Clin Cancer Res* (2008) 14(16):5158–65. doi: 10.1158/1078-0432.CCR-07-4756
 46. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A Single-Cell and Spatially Resolved Atlas of Human Breast Cancers. *Nat Genet* (2021) 53(9):1334–47. doi: 10.1038/s41588-021-00911-1
 47. Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-Omic Machine Learning Predictor of Breast Cancer Therapy Response. *Nature* (2021) 601:623–9. doi: 10.1038/s41586-021-04278-5
 48. Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C, Dykgers A, et al. A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* (2019) 177(5):1330–45.e18. doi: 10.1016/j.cell.2019.03.005
 49. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular Portraits of Human Breast Tumours. *Nature* (2000) 406(6797):747–52. doi: 10.1038/35021093
 50. Koundouros N, Poulgiannis G. Reprogramming of Fatty Acid Metabolism in Cancer. *Br J Cancer* (2020) 122(1):4–22. doi: 10.1038/s41416-019-0650-z
 51. Mashima T, Seimiya H, Tsuruo T. *De Novo* Fatty-Acid Synthesis and Related Pathways as Molecular Targets for Cancer Therapy. *Br J Cancer* (2009) 100(9):1369–72. doi: 10.1038/sj.bjc.6605007
 52. Monaco ME. Fatty Acid Metabolism in Breast Cancer Subtypes. *Oncotarget* (2017) 8(17):29487–500. doi: 10.18632/oncotarget.15494
 53. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023
 54. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua S, et al. Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-Negative Breast Cancer. *Clin Cancer Res* (2015) 21(7):1688–98. doi: 10.1158/1078-0432.CCR-14-0432
 55. Grusos T, Gigoux M, Manem VSK, Bertos N, Zuo D, Perlitch I, et al. Spatially Distinct Tumor Immune Microenvironments Stratify Triple-Negative Breast Cancers. *J Clin Invest* (2019) 129(4):1785–800. doi: 10.1172/JCI96313
 56. Ali HR, Glont SE, Blows FM, Provenzano E, Dawson SJ, Liu B, et al. PD-L1 Protein Expression in Breast Cancer Is Rare, Enriched in Basal-Like Tumours and Associated With Infiltrating Lymphocytes. *Ann Oncol* (2015) 26(7):1488–93. doi: 10.1093/annonc/mdv192
 57. Oshi M, Kawaguchi T, Yan L, Peng X, Qi Q, Tian W, et al. Immune Cytolytic Activity Is Associated With Reduced Intra-Tumoral Genetic Heterogeneity and With Better Clinical Outcomes in Triple Negative Breast Cancer. *Am J Cancer Res* (2021) 11(7):3628–44.
 58. Pellegrino B, Hlavata Z, Migali C, De Silva P, Aiello M, Willard-Gallo K, et al. Luminal Breast Cancer: Risk of Recurrence and Tumor-Associated Immune Suppression. *Mol Diagn Ther* (2021) 25(4):409–24. doi: 10.1007/s40291-021-00525-7
 59. El Bairi K, Haynes HR, Blackley E, Fineberg S, Shear J, Turner S, et al. The Tale of TILs in Breast Cancer: A Report From The International Immunology Biomarker Working Group. *NPJ Breast Cancer* (2021) 7(1):150. doi: 10.1038/s41523-021-00346-1
 60. Ye Y, Zhang Y, Yang N, Gao Q, Ding X, Kuang X, et al. Profiling of Immune Features to Predict Immunotherapy Efficacy. *Innovation* (2022) 3(1):100194. doi: 10.1016/j.xinn.2021.100194
 61. Kossai M, Radosevic-Robin N, Penault-Llorca F. Refining Patient Selection for Breast Cancer Immunotherapy: Beyond PD-L1. *ESMO Open* (2021) 6(5):100257. doi: 10.1016/j.esmoop.2021.100257
 62. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im S-A, Yusof MM, et al. Pembrolizumab Plus Chemotherapy Versus Placebo Plus Chemotherapy for Previously Untreated Locally Recurrent Inoperable or Metastatic Triple-Negative Breast Cancer (KEYNOTE-355): A Randomised, Placebo-Controlled, Double-Blind, Phase 3 Clinical Trial. *Lancet* (2020) 396(10265):1817–28. doi: 10.1016/S0140-6736(20)32531-9
 63. Miles D, Gligorov J, André F, Cameron D, Schneeweiss A, Barrios C, et al. Primary Results From IMpassion131, a Double-Blind, Placebo-Controlled, Randomised Phase III Trial of First-Line Paclitaxel With or Without Atezolizumab for Unresectable Locally Advanced/Metastatic Triple-Negative Breast Cancer. *Ann Oncol* (2021) 32(8):994–1004. doi: 10.1016/j.annonc.2021.05.801
 64. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of Human Triple-Negative Breast Cancer Subtypes and Preclinical Models for Selection of Targeted Therapies. *J Clin Invest* (2011) 121(7):2750–67. doi: 10.1172/JCI45014
 65. Liu Y-R, Jiang Y-Z, Xu X-E, Hu X, Yu K-D, Shao Z-M. Comprehensive Transcriptome Profiling Reveals Multigene Signatures in Triple-Negative Breast Cancer. *Clin Cancer Res* (2016) 22(7):1653–62. doi: 10.1158/1078-0432.CCR-15-1555
 66. Filho OM, Stover DG, Asad S, Ansell PJ, Watson M, Loibl S, et al. Association of Immunophenotype With Pathologic Complete Response to Neoadjuvant Chemotherapy for Triple-Negative Breast Cancer: A Secondary Analysis of the Brightness Phase 3 Randomized Clinical Trial. *JAMA Oncol* (2021) 7(4):603–8. doi: 10.1001/jamaoncol.2020.7310
 67. He Y, Wang L, Wei T, Xiao YT, Sheng H, Su H, et al. FOXA1 Overexpression Suppresses Interferon Signaling and Immune Response in Cancer. *J Clin Invest* (2021) 131(14):e147025. doi: 10.1172/JCI147025
 68. Garrido-Castro AC, Lin NU, Polyak K. Insights Into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov* (2019) 9(2):176–98. doi: 10.1158/2159-8290.CD-18-1177
 69. Mathews JC, Nadeem S, Levine AJ, Pouryahya M, Deasy JO, Tannenbaum A. Robust and Interpretable PAM50 Reclassification Exhibits Survival Advantage for Myoepithelial and Immune Phenotypes. *NPJ Breast Cancer* (2019) 5(1):1–8. doi: 10.1038/s41523-019-0124-8
 70. Shaul ME, Fridlender ZG. Tumour-Associated Neutrophils in Patients With Cancer. *Nat Rev Clin Oncol* (2019) 16(10):601–20. doi: 10.1038/s41571-019-0222-4
 71. Snoderly HT, Boone BA, Bennewitz MF. Neutrophil Extracellular Traps in Breast Cancer and Beyond: Current Perspectives on NET Stimuli, Thrombosis and Metastasis, and Clinical Utility for Diagnosis and Treatment. *Breast Cancer Res* (2019) 21(1):1–13. doi: 10.1186/s13058-019-1237-6
 72. Ding YC, Steele L, Warden C, Wilczynski S, Mortimer J, Yuan Y, et al. Molecular Subtypes of Triple-Negative Breast Cancer in Women of Different Race and Ethnicity. *Oncotarget* (2019) 10(2):198–208. doi: 10.18632/oncotarget.26559
 73. O'Meara T, Safonov A, Casadevall D, Qing T, Silber A, Brigid K, et al. Immune Microenvironment of Triple-Negative Breast Cancer in African American and Caucasian Women. *Breast Cancer Res Treat* (2020) 175(1):247–59. doi: 10.1007/s10549-019-05156-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Citation: Llera AS, Abdelhay ESW, Artagaveytia N, Daneri-Navarro A, Müller B, Velazquez C, Alcoba EB, Alonso I, Alves da Quinta DB, Binato R, Bravo AI, Camejo N, Carraro DM, Castro M, Castro-Cervantes JM, Cataldi S, Cayota A, Cerda M, Colombo A, Crocama S, Del Toro-Arreola A, Delgadillo-Cisterna R, Delgado L, Dreyer-Breitenbach M, Fejerman L, Fernández EA, Fernández J, Fernández W, Franco-Topete RA, Gabay C, Gaete F, Garibay-Escobar A, Gómez J, Greif G, Gross TG, Guerrero M, Henderson MK, Lopez-Muñoz ME, Lopez-Vazquez A, Maldonado S, Morán-Mendoza AJ, Nagai MA, Ocegüera-Villanueva A, Ortiz-Martínez MA, Quintero J, Quintero-Ramos A, Reis RM, Retamales J, Rivera-Claisse E, Rocha D, Rodríguez R, Rosales C, Salas-González E, Sanchotena V, Segovia L, Sendoya JM, Silva-García AA, Trincherero A, Valenzuela O, Vedham V, Zagame L, United States-Latin American Cancer Research Network (US-LACRN) and Podhajcer OL (2022) The

Transcriptomic Portrait of Locally Advanced Breast Cancer and Its Prognostic Value in a Multi-Country Cohort of Latin American Patients. *Front. Oncol.* 12:835626. doi: 10.3389/fonc.2022.835626

Copyright © 2022 Llera, Abdelhay, Artagaveytia, Daneri-Navarro, Müller, Velazquez, Alcoba, Alonso, Alves da Quinta, Binato, Bravo, Camejo, Carraro, Castro, Castro-Cervantes, Cataldi, Cayota, Cerda, Colombo, Crocama, Del Toro-Arreola, Delgadillo-Cisterna, Delgado, Dreyer-Breitenbach, Fejerman, Fernández, Fernández, Fernández, Franco-Topete, Gabay, Gaete, Garibay-Escobar, Gómez, Greif, Gross, Guerrero, Henderson, Lopez-Muñoz, Lopez-Vazquez, Maldonado, Morán-Mendoza, Nagai, Ocegüera-Villanueva, Ortiz-Martínez, Quintero, Quintero-Ramos, Reis, Retamales, Rivera-Claisse, Rocha, Rodríguez, Rosales, Salas-González, Sanchotena, Segovia, Sendoya, Silva-García, Trincherero, Valenzuela, Vedham, Zagame, United States-Latin American Cancer Research Network (US-LACRN) and Podhajcer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.