UC Davis
UC Davis Previously Published Works

Title

Ancestry as a potential modifier of gene expression in breast tumors from Colombian women.

Permalink
https://escholarship.org/uc/item/19c146c3

Journal
PloS one, 12(8)

ISSN
1932-6203

Authors
Serrano-Gómez, Silvia J
Sanabria-Salas, María Carolina
Garay, Jone
et al.

Publication Date
2017

DOI
10.1371/journal.pone.0183179

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

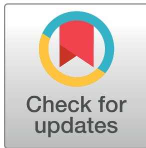
RESEARCH ARTICLE

Ancestry as a potential modifier of gene expression in breast tumors from Colombian women

Silvia J. Serrano-Gómez[1,2], María Carolina Sanabria-Salas[1], Jone Garay[3], Melody C. Baddoo[4], Gustavo Hernández-Suarez[1], Juan Carlos Mejía[5], Oscar García[6], Lucio Miele[7], Laura Fejerman[8☯], Jovanny Zabaleta[3,9☯]*

1 Grupo de investigación en Biología del Cáncer, Instituto Nacional de Cancerología, Bogotá D. C, Colombia, **2** Programa de doctorado en Ciencias Biológicas, Pontificia Universidad Javeriana, Bogotá D. C, Colombia, **3** Stanley S. Scott Cancer Center, LSUHSC, New Orleans, LA, United States of America, **4** Tulane University School of Medicine, New Orleans, LA, United States of America, **5** Grupo de Patología, Instituto Nacional de Cancerología, Bogotá D. C, Colombia, **6** Grupo de Seno y Tejidos blandos, Instituto Nacional de Cancerología, Bogotá D. C, Colombia, **7** Department of Genetics, LSUHSC, New Orleans, LA, United States of America, **8** Department of Medicine, Institute of Human Genetics, University of California San Francisco, San Francisco, CA, United States of America, **9** Department of Pediatrics, LSUHSC, New Orleans, United States of America

☯ These authors contributed equally to this work.
* jzabal@lsuhsc.edu



OPEN ACCESS

Citation: Serrano-Gómez SJ, Sanabria-Salas MC, Garay J, Baddoo MC, Hernández-Suarez G, Mejía JC, et al. (2017) Ancestry as a potential modifier of gene expression in breast tumors from Colombian women. PLoS ONE 12(8): e0183179. https://doi.org/10.1371/journal.pone.0183179

Editor: Amanda Ewart Toland, Ohio State University Wexner Medical Center, UNITED STATES

Received: May 17, 2017

Accepted: July 31, 2017

Published: August 23, 2017

Copyright: © 2017 Serrano-Gómez et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data analyzed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE101927.

Funding: Colombian National Cancer Institute (C41030610-110 to MCSS). National Cancer Institute (CA 160607 to L.F.). J.Z. has been partially supported by grants from the National Institute of General Medical Sciences (NIGMS, P20GM103501

Abstract

Background

Hispanic/Latino populations are a genetically admixed and heterogeneous group, with variable fractions of European, Indigenous American and African ancestries. The molecular profile of breast cancer has been widely described in non-Hispanic Whites but equivalent knowledge is lacking in Hispanic/Latinas. We have previously reported that the most prevalent breast cancer intrinsic subtype in Colombian women was Luminal B as defined by St. Gallen 2013 criteria. In this study we explored ancestry-associated differences in molecular profiles of Luminal B tumors among these highly admixed women.

Methods

We performed whole-transcriptome RNA-seq analysis in 42 Luminal tumors (21 Luminal A and 21 Luminal B) from Colombian women. Genetic ancestry was estimated from a panel of 80 ancestry-informative markers (AIM). We categorized patients according to Luminal subtype and to the proportion of European and Indigenous American ancestry and performed differential expression analysis comparing Luminal B against Luminal A tumors according to the assigned ancestry groups.

Results

We found 5 genes potentially modulated by genetic ancestry: *ERBB2* (log2FC = 2.367, padj<0.01), *GRB7* (log2FC = 2.327, padj<0.01), *GSDMB* (log2FC = 1.723, padj<0.01, *MIEN1* (log2FC = 2.195, padj<0.01 and *ONECUT2* (log2FC = 2.204, padj<0.01). In the

subproject # 2, P30GM114732, U54GM104940-01) and the National Institute on Minority Health and Health Disparities (NIMHD, P20MD004817, U54MD008176-01).

Competing interests: The authors have declared that no competing interests exist.

replication set we found a statistical significant association between *ERBB2* expression with Indigenous American ancestry (p = 0.02, B = 3.11). This association was not biased by the distribution of HER2+ tumors among the groups analyzed.

Conclusions

Our results suggest that genetic ancestry in Hispanic/Latina women might modify *ERBB2* gene expression in Luminal tumors. Further analyses are needed to confirm these findings and explore their prognostic value.

Introduction

Breast cancer is a complex-multifactorial disease, consisting of a highly heterogeneous group of tumors with particular molecular features, prognosis and responses to therapy [1–4]. The first gene expression-based classification of breast cancer into intrinsic subtypes was published in 2000 [5] and identified estrogen receptor (ER) positive (ER+) subtypes Luminal A and B, and ER negative (ER-) subtypes basal-like and human epidermal growth factor receptor 2-enriched (HER2-enriched) [1, 6]. Subsequent studies showed differences in the outcomes according to intrinsic subtypes [7, 8]. Based on this classification, the best outcomes are observed for Luminal A tumors while basal-like and HER2-enriched are associated with worse outcomes.

The Luminal B subtype represents 30%– 40% of breast cancers [9, 10]. Despite expressing ERα and being amenable to endocrine therapy, they tend to be clinically more aggressive and have worse prognosis compared to Luminal A tumors. For example, it has been observed that survival curves for Luminal B tumors are similar to those from basal-like tumors after 10 years of follow-up [9]. Moreover, Luminal B tumors have higher risk of *de novo* resistance to endocrine therapies [7, 11] and at the molecular level, they are characterized by increased expression of cell proliferation genes or cell cycle regulators such as *MKI67* and *AURKA* [12–14]. Luminal B tumors more frequently receive high recurrence scores based on the Oncotype Dx gene expression signature and are more likely to benefit from cytotoxic chemotherapy, reaching higher percentages of pathologic complete response (pCR) when compared to Luminal A tumors [13, 15]. However, results on this issue have not been consistent [16–18].

Although there is still controversy as to whether Luminal A and Luminal B represent two different biological entities or a single entity that changes from one status to another through the acquisition of mutations [19–21], it is well established that this group of tumors are characterized by the expression of estrogen receptor α (*ESR1* gene), progesterone receptor (*PGR* gene) and cytokeratins characteristic of luminal cells such as cytokeratin 8 and cytokeratin 18 (*KRT8*, *KRT18*) [19]. Luminal A tumors are usually low-grade malignancies that show gains in 1q and loss in 16q. Luminal B tumors have a more complex profile of copy number variations (CNV). Amplifications at 8p11 (*FGFR1* locus), 8q21, 11q13, 17q12 (*ERBB2* locus) and 20q13 have been reported [13, 22, 23]. Based on gene expression profiles, these two subtypes share molecular patterns such as the expression of the *ESR1* gene and other genes such as *FOXA1* and *BCL2*, but their main difference is the high expression of proliferation genes such as *MKI67*, the survivin gene *BIRC5*, and the cyclin B1 (*CCNB1*) gene that characterizes the Luminal B subtype [14, 24, 25].

Hispanic/Latinas is a heterogeneous group with variable proportions of European, Indigenous American (IA) and African ancestries [26]. The Colombian population is one of the most diverse of Latin America [27]. Our group has previously reported that Luminal B, as defined

by St. Gallen surrogate immunohistochemistry criteria, is the most common breast cancer intrinsic subtype among Colombian women [28]. The objective of the present work was to explore the molecular profile of Luminal tumors in Colombian women to assess the association between genetic ancestry and gene expression. We performed RNA-seq analysis in 42 formalin-fixed paraffin embedded (FFPE) tumor blocks previously classified as Luminal subtypes (21 Luminal A and 21 Luminal B) and with known genetic ancestry proportions [28]. We found 5 candidate genes (*ERBB2*, *GRB7*, *GSDMB*, *MIEN1* and *ONECUT2*) potentially modulated by genetic ancestry in Colombian-Latina patients with Luminal tumors.

Materials and methods

Patient selection

Patient's sample blocks were selected from a database of 252 breast cancer patients with known genetic ancestry and with FFPE tumor specimens available. These 252 patients are part of a database of 301 breast cancer patients from Colombia that we have previously described [28]. For deep sequencing analysis, we selected 59 samples from the Andean region according to breast cancer intrinsic subtype and to the predominance of the European or IA ancestral fraction. Intrinsic subtypes were assessed using a panel of 6 immunohistochemistry (IHC) markers that included ER, progesterone receptor (PgR), the human epidermal growth factor receptor 2 (HER2), Ki-67, the Epidermal Growth Factor Receptor (EGFR), and Cytokeratin 5/6 (CK5/6), following the recommendations of St. Gallen 2013 consensus [29]. This study was approved by the Colombian National Cancer Institute ethics committee. Since we worked with de-identified FFPE tissues collected more than 3 years before the analysis done for this work, the Colombian NCI according to the Colombian laws, considered that no informed consent was required.

Ancestry estimation

DNA was extracted from normal FFPE tissues using the RecoverAll™ Total Nucleic Acid Isolation Kit (Life Technologies, Carlsbad, CA) following the manufacturer's recommendations. A panel of 106 Single Nucleotide Polymorphisms (SNPs) previously validated as Ancestry-Informative Markers (AIMs) was used to estimate individual genetic ancestry [30]. Genotyping was performed at the University of Minnesota Genomics Center using Sequenom technology. SNPs with call rate $<90\%$ or that deviated from Hardy-Weinberg equilibrium were removed from the analysis, leaving 80 SNPs for ancestry estimation. The software STRUCTURE version 2.222 [31] was used under an admixture model fixing the number of ancestral components to $k = 3$ to estimate Indigenous American (IA), European and African proportions for each of the samples. We used a burn-in period of 10,000 iterations followed by 50,000 additional iterations. Parental populations that include 42 Europeans (Coriell's North American Caucasian panel), 37 West Africans (non-admixed Africans living in London, United Kingdom and South Carolina) and 30 Indigenous Americans (15 Mayans and 15 Nahuas) [30] were included to perform a supervised analysis of our samples.

RNA-Seq sample preparation and data analysis

Hematoxylin and eosin-stained slides were evaluated by a pathologist to estimate the percentage of tumor present in the paraffin block selected. For cases with or more than 60% of tumor content, five 10μm sections were used for RNA extraction. For cases with less than 60% of tumor content, areas that contained tumor were marked to obtain 5 tumor cores using a 1-mm punch needle. RNA extraction was done using the RecoverAll™ Total Nucleic Acid

Isolation Kit (Life Technologies, Carlsbad, CA) following the manufacturer's recommendations. RNA-seq analysis was performed at the Stanley S. Scott Cancer Center's Translational Genomics Core at LSUHSC. RNA was quantified by NanoDrop ND1000 Spectrophotometer (Thermo Scientific, Wilmington, USA) and its quality assessed with RNA 6000 Nano kit in the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Even though the samples presented some RNA degradation, they were suitable for library preparation, based on the protocols and recommendations from Illumina.

Library preparation was performed in 59 samples from 1μg of total RNA using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina Inc., San Diego, CA). Briefly, isolated RNA was depleted of ribosomal RNA using the rRNA Removal Mix provided by the kit. Random hexamers were used for cDNA synthesis. Subsequently, cDNA was subjected to end repair, adapter ligation and size selection using AMPure XP beads (Beckman Coulter Inc., Brea, CA). Fragmentation step was omitted due to the sample quality, as recommended by the protocol. Libraries were quantified by Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA) and the validation of the library size was performed in an Agilent Bioanalyzer using a DNA 1000 kit (Agilent Technologies, Santa Clara, CA) to verify the presence of a 260 base pair fragment. From the luminal tumors selected for library preparation, 42 (21 Luminal A and 21 Luminal B) had the expected size to proceed to the sequencing.

Sequencing was performed in a Genome Analyzer IIX (Illumina Inc) in a single-read 60 + 7 run (sequence plus index). For data analysis, FASTQ files were generated using CASAVA v1.8.1. FastQC software (Version 0.9.6) was used to evaluate the quality of the files. The trimming of adapter sequences from the reads was performed using fastq-mcf utility [32] and RSEM [33] was used to map single-end reads to reference transcriptome hg38 (Ensembl) and to provide read counts and normalized expression values for each case analyzed. The data analyzed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE101927.

Differential gene expression analysis

To identify ancestry-associated differentially expressed genes in Luminal tumors, we categorized patients according to the average European or IA ancestry fractions and compared Luminal B tumors vs. Luminal A tumors. We used Luminal A tumors as a reference group as they represent the most biologically similar but less aggressive breast cancer subtype compared to Luminal B. We used DESeq2 package [34] in R-studio (http://www.rstudio.com/) to perform differential expression analysis. This analysis applies a general linear model to estimate log2 fold changes (log2FC) to test if differences between groups are equal to zero. Pre-filtering was applied to the data matrix to analyze transcripts with at least 1 read count. Genes with Benjamini-Hochberg adjusted < 0.05 ($padj < 0.05$) were reported as significantly different between groups. Signaling pathway analysis was done in Metacore (Thomson Reuters) and DAVID annotation tool (http://david.abcc.ncifcrf.gov/) [35]. Venn diagrams were done using Venny 2.1 online tool (http://bioinfogp.cnb.csic.es/tools/venny/).

Real time-PCR validation

cDNA was synthesized from 100ng of total RNA using SuperScript III First-Strand Synthesis SuperMix Kit (Invitrogen) in 166 samples from Luminal tumors (42 analyzed by RNA-seq and 124 new samples), according to the manufacturer's instructions. TaqMan probes were used to quantify the levels of mRNA expression of candidate genes: *ERBB2* (Hs01001580_m1), *GRB7* (Hs00917999_g1), *ONECUT2* (Hs00191477_m1). The reaction was amplified in a QuantStudio 12 K plex Real-Time PCR machine (ThermoScientific). The $2^{-\Delta\Delta CT}$ method was used to

estimate the fold changes and *GAPDH* (Hs03929097_g1) was used as an internal calibrator. Water was used as a negative control.

Statistical analysis

All statistical analyses were performed using R project (www.r-project.org) and SPSS Inc. (Released 2007; SPSS for Windows, Version 16.0. Chicago, IL, USA). Differences in the characteristics of the patients according to intrinsic subtype were analyzed using X^2 test and differences in the mean of the ancestry fractions, age at diagnosis and tumor size were analyzed using analysis of variance test (ANOVA). *p* values less than 0.05 were considered statistical significant. Logistic regression model was used to test the association between gene expression of *TOP2A* and *CYP19A1* and presence of recurrences.

Pearson correlation was used to determine the correlation between the expression level of ERα, PgR, HER2 and Ki-67 obtained by IHC and RNA-seq. For gene expression, we used normalized values of the read counts from each gene. The expression by IHC was assessed by percentage of expression for ERα, PgR and Ki-67. HER2 measurement was semi-quantitative according to the recommendations of the American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP) guideline [36]. According to these criteria, negative cases are those with no membrane staining or weak staining for less than 10% of tumor cells (score 0), or incomplete and weak staining for more than 10% of tumor cells (score 1+). Cases with weak to moderate staining in more than 10% of tumor cells are assigned at score of 2+. Finally, HER2 positive cases have a complete and intense membrane staining in more than 10% of tumor cells and are assigned a score of 3+. We used HER2 scores by IHC to performed Pearson correlation.

We used Spearman correlation to test the correlation between the expression levels of the candidate genes obtained by RNA-seq and the fold changes calculated from the qRT-PCR. Linear regression analysis was used to test the association between expression levels of candidate genes and the intrinsic subtypes of breast cancer (Luminal A or Luminal B), and/or the genetic ancestry.

Results

Characteristics of patients

Twenty one (21) of these patients were classified as Luminal A and 21 as Luminal B (Table 1) according to the recommendations of the St. Gallen 2013 panel [29].

The mean age at diagnosis was 59.4 years and the average of the tumor size was 39 millimeters (mm). The average of European, IA and African ancestry fractions were 0.58, 0.36 and 0.06, respectively. We did not find statistical significant differences in the aforementioned characteristics between Luminal intrinsic subtypes. All patients were positive for expression of ER, meanwhile, PgR was positive in all Luminal A tumors and in 81% of Luminal B tumors. Other clinicopathological variables such as node status and the clinical stage at diagnosis, as defined by the American Joint Committee (AJCC) stage, and recurrences did not differ by Luminal intrinsic subtype.

We found statistical significant differences in HER2 expression by IHC ($p < 0.01$). All Luminal A tumors were negative for HER2 expression while for Luminal B tumors, 61.9% were positive and 38.1% were negative. We also found statistical significant differences in the tumor grade. Tumor grade 3 was found only for Luminal B intrinsic subtype when compared to Luminal A (28.6% vs. 0%, respectively). The administration of adjuvant therapy differed between the two luminal subtypes ($p = 0.032$). Patients with Luminal A tumors were more likely to receive hormonotherapy than patients with luminal B subtype (38.1% vs 4.8%); while

Table 1. Characteristic of patients analyzed by RNA-seq.

	Luminal A (n = 21)	Luminal B (n = 21)	*p*
Age, Yrs.	57.7 ± 13.9	61.1 ± 11.3	0.391
Tumor Size (mm)	78.9 ± 186.6	39.5 ± 20.07	0.871
Mean European ancestry	0.57 ± 0.13	0.58 ± 0.18	0.78
Mean IA ancestry	0.37 ± 0.13	0.36 ± 0.17	0.853
Mean African ancestry	0.06 ± 0.07	0.06 ± 0.06	0.802
PgR expression, N (%)			0.11
Positive	21 (100)	17 (81)	
Negative	0	4 (19)	
HER2 expression, N (%)			< 0.01
Positive	0	13 (61.9)	
Negative	21 (100)	8 (38.1)	
Tumor Grade, N (%)			0.008
I	3 (14.3)	0	
II	16 (76.2)	12 (57.1)	
III	0	6 (28.6)	
Unknown	2 (9.5)	3 (14.3)	
Nodes, N (%)			0.354
Positive	9 (42.9)	13 (61.9)	
Negative	12 (57.1)	8 (38.1)	
AJCC Stage, N (%)			0.328
I	3 (14.3)	1 (4.8)	
IIA/IIB	9 (42.9)	7 (33.3)	
IIIA/IIIB/IIIC	8 (38.1)	13 (61.9)	
IV	1 (4.8)	0	
Adjuvant Therapy, N (%)			0.032
Chemotherapy	1 (4.8)	0	
Hormonotherapy	8 (38.1)	1 (4.8)	
Combined*	11 (52.3)	19 (90.4)	
Not administered	1 (4.8)	0	
Unknown	0	1 (4.8)	
Cytotoxic regimen, N (%)			0.003
Anthracyclines	4 (19)	5 (23.8)	
Anthracyclines + Taxanes	2 (9.5)	1 (4.8)	
Anthracyclines + Taxanes + Trastuzumab	0	1 (4.8)	
Anthracyclines + Trastuzumab	0	1 (4.8)	
CMF regimen	0	1 (4.8)	
Taxanes	6 (28.6)	2 (9.5)	
Taxanes + Trastuzumab	0	6 (28.6)	
TC regimen	0	1 (4.8)	
Trastuzumab alone	0	1 (4.8)	
Not administered	9 (42.9)	1 (4.8)	
Unknown	0	1 (4.8)	
Hormonotherapy, N (%)			0.001
Anastrazole	2 (9.5)	1 (4.8)	
Letrozole	1 (4.8)	2 (9.5)	
Not administered	2 (9.5)	0	
Unknown	0	1 (4.8)	

(*Continued*)

Table 1. (*Continued*)

	Luminal A (n = 21)	Luminal B (n = 21)	*p*
Switch Aromatase Inhibitor	1 (4.8)	11 (52.3)	
Tamoxifen	15 (71.4)	6 (28.6)	
Recurrence, N (%)			0.439
Systemic	3 (14.3)	4 (19)	
No recurrences	18 (85.7)	16 (76.2)	
Unknown	0	1 (4.8)	

* Hormonotherapy and chemotherapy

https://doi.org/10.1371/journal.pone.0183179.t001

patients with Luminal B tumors were more likely to receive combined therapy compared to luminal A tumors (90.4% vs. 52.3%, respectively). The cytotoxic regimens were also different between the luminal subtypes (p = 0.003). Patients with luminal B subtype more frequently received cytotoxic treatments with anthracyclines (23.8%) and taxanes plus trastuzumab (28.6%) compared to 42.9% of patients with luminal A tumors who did not received cytotoxic chemotherapy. Finally, we also found statistically significant differences in the administration of hormonotherapy (p = 0.001). Patients with luminal A tumors typically received Tamoxifen (71.4%) while patients with luminal B tumors were more likely to switch to an aromatase inhibitor (52.3%).

Correlation analysis between immunohistochemistry and gene expression levels

As mentioned above, we used IHC surrogates from St. Gallen 2013 consensus to classify breast cancers into intrinsic subtypes. We performed Pearson correlations to determine whether the immunohistochemical expression of ERα, PgR, HER2 and Ki67 was associated with their gene expression profiles. We found statistically significant correlations between IHC and mRNA expression levels measured by RNA-seq for PgR ($R^2 = 0.737$, $p < 0.01$), ERα ($R^2 = 0.505$, $p = 0.02$), Ki67 ($R^2 = 0.629$, $p < 0.01$) and HER2 ($R^2 = 0.485$, $p = 0.001$) (S1A–S1D Fig). These results suggest that the approximation to Luminal subtypes by IHC is reasonable given the high correlation between protein expression and the gene expression of four of the markers used.

Gene expression profile of Luminal B tumors in Colombian women classified by St Gallen 2013 surrogates

We have previously reported that according to the St. Gallen 2013 panel surrogates, Luminal B is the most common intrinsic subtype of breast cancer in Colombian-Latinas [28]. In order to analyze the underlying molecular profile of Luminal B tumors in our population we compared tumors classified by IHC as Luminal B versus Luminal A. We found 67 differentially expressed genes (padj < 0.05) from which 39 were up-regulated and 28 down-regulated in the Luminal B subtype (Table 2).

Unsupervised hierarchical clustering showed that using these genes, most Luminal B tumors (15) clustered together and diverge from Luminal A tumors (Fig 1A). Interestingly, 6 Luminal B tumors clustered with the Luminal A group. We analyzed the expression levels of *ESR1*, *PGR*, *MKI67* and *ERBB2* at the gene expression level as these codify for the markers that we used in the IHC to distinguish Luminal B from Luminal A tumors following St. Gallen surrogates. We observed that although these 6 tumors have lower expression of *ESR1* when

Table 2. Differentially expressed genes between Luminal B and Luminal A tumors classified by St. Gallen 2013 surrogates.

Up-regulated genes between Luminal B and Luminal A			Down-regulated genes between Luminal B and Luminal A		
Gene	logFC	padj	Gene	logFC	padj
CDK1	1.432	0.004	*RALBP1*	-0.443	0.048
RP11-510N19.5	1.412	0.016	*RNU5B-1*	-0.713	0.029
AIF1L	1.400	0.001	*TTC39C*	-0.721	0.042
CYP19A1	1.325	0.030	*RCAN3*	-0.760	0.042
TOP2A	1.309	0.011	*KDM4B*	-0.762	0.018
KIF14	1.287	0.010	*INO80E*	-0.768	0.034
DSCAM-AS1	1.275	0.042	*SNORA54*	-0.798	0.024
LAD1	1.249	0.040	*STARD13*	-0.845	0.030
CD24	1.228	0.024	*HIPK2*	-0.894	0.034
CENPF	1.210	0.001	*ZNF213*	-0.913	0.018
IQGAP3	1.180	0.021	*ABAT*	-0.922	0.024
PGAP3	1.157	0.033	*RP4-734G22.3*	-0.980	0.037
CDC6	1.148	0.048	*JMJD8*	-1.002	0.023
SLC4A8	1.143	0.006	*ELOVL5*	-1.030	0.018
BCAS1	1.135	0.024	*TMEM177*	-1.046	0.018
ORMDL3	1.135	0.015	*SNHG8*	-1.059	0.018
CRABP2	1.123	0.004	*PAIP2B*	-1.062	0.029
ASPM	1.119	0.018	*SNORA76C*	-1.095	0.049
STARD3	1.113	0.029	*FGD3*	-1.096	0.026
BUB1	1.087	0.016	*RPS16P5*	-1.117	0.048
CCNA2	1.080	0.034	*FCGBP*	-1.222	0.048
IGFBP5	1.062	0.027	*RBBP8*	-1.226	0.013
MKI67	1.054	0.021	*KCND3*	-1.228	0.024
ANLN	1.048	0.048	*RNU6-36P*	-1.232	0.042
CCNB2	1.026	0.043	*BAI2*	-1.372	0.013
ELF3	1.018	0.018	*ABCA3*	-1.383	0.001
EXOC2	0.982	0.018	*SERPINA1*	-1.449	0.013
CENPE	0.976	0.033	*NTRK2*	-1.497	0.008
NT5E	0.941	0.014			
SIX4	0.918	0.048			
CLDN4	0.913	0.038			
ARF6	0.810	0.019			
CDK12	0.808	0.030			
RHOC	0.639	0.049			
RBM39	0.568	0.045			
CAMSAP2	0.545	0.034			
UTP20	0.543	0.043			
U2SURP	0.387	0.034			
CAND1	0.361	0.034			

https://doi.org/10.1371/journal.pone.0183179.t002

compared to all other luminal A tumors, their expression for the other three markers (PGR, MKI67, and ERBB2) was similar to the luminal A cluster (Table 3). This result suggests that from the molecular profile, these 6 tumors behave more similarly to luminal A tumors than to luminal B although at the protein level they are classified as luminal B. It remains to be determined whether these tumors have better outcomes than the other Luminal B cases. However,

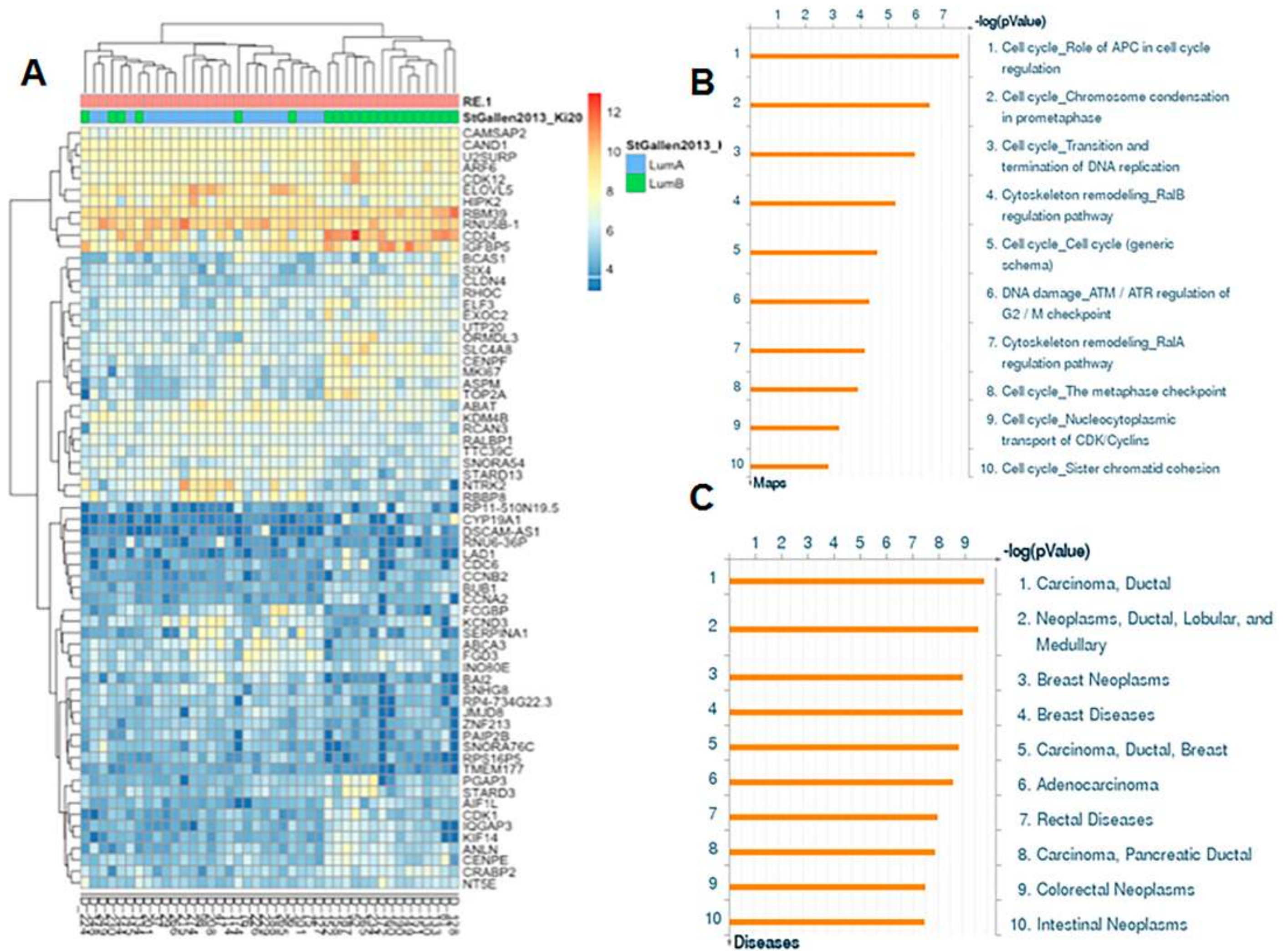


Fig 1. Gene expression profile of 42 Luminal breast cancer samples. (A) Unsupervised hierarchical clustering with 67 differentially expressed genes between IHC defined Luminal B and Luminal A tumors. (B) Most relevant signaling pathways associated with 67 differentially expressed genes in Luminal B tumors from Colombian women. (C) Diseases associated with differentially expressed genes in Luminal B.

https://doi.org/10.1371/journal.pone.0183179.g001

Kaplan-Meier analysis did not show any statistically significant difference between the groups (data not shown).

Pathway analysis showed that the top up-regulated genes participate in biological processes such as mitosis and cell cycle regulation (e.g., *CDK1*, *CDC6*, *CCNB2*, *BUB1*, *CENPF*, *ANLN*, *CENPE*, *CCNA2*, *ASPM*, *MKI67*) and down-regulated genes encode mostly phosphoproteins

Table 3. Expression of *ESR1*, *PGR*, *MKI67* and *ERBB2* in clusters identified in the unsupervised hierarchical clustering.

	ESR1	*PGR*	*MKI67*	*ERBB2*
Luminal A	9.98 ± 1.25	8.47 ± 2.22	5.73 ± 0.63	203.44 ± 116.68
*Luminal A.1**	9.18 ± 1.65	8.04 ± 2.26	5.24 ± 1.32	134.88 ± 65.27
Luminal B	9.82 ± 1.65	7.08 ± 1.87	7.10 ± 0.62	789.49 ± 1244.61

*This group corresponds to the tumors classified as luminal B by immunohistochemistry but that clustered together with luminal A tumors

https://doi.org/10.1371/journal.pone.0183179.t003

(e.g., *KCND3*, *RALBP1*, *RCAN3*, *ABCA3*, *RBBP8*, *PAIP2B*, *STARD13*, *ELOVL5*, *HIPK2*, *NTRK2*, *KDM4B*, *BAI2*, *FGD3*) (Fig 1B). The diseases associated with these differentially expressed genes include ductal carcinoma and breast neoplasms (Fig 1C), which was consistent with the origin of the tissue specimens.

Interestingly, two of the genes that we found upregulated were *TOP2A* (log2FC = 1.309, padj = 0.011) and *CYP19A1*, which codify for the aromatase gene (log2FC = 1.325, padj = 0.030). As *TOP2A* has been associated with response to anthracycline-based chemotherapy and aromatase inhibitors are widely used for breast cancer treatment, we explored if the expression of these genes and the intrinsic subtype of breast cancer could be associated with the development of recurrences. We did not find any statistical significant association (S1 Table)

Ancestry-associated differentially expressed genes in Luminal tumors

Stratified analysis by European ancestry fraction. To identify ancestry-associated differentially expressed genes in Luminal tumors, we categorized patients according to Luminal subtype (Luminal A and Luminal B) and to the proportion of European ancestry into **low European ancestry group** (European ancestry proportion below the average 0.58); and **high European ancestry group** (European ancestry proportion above the average 0.58). The averages for European, IA and African ancestry fractions according to the assigned groups are shown in the S2 Table. We then compared Luminal B against Luminal A tumors according to the assigned ancestry groups and found 27 ancestry-modulated genes in the low European ancestry group and 3 in the high European ancestry group (Fig 2).

We compared the differentially expressed genes found in the global profile of Luminal B tumors (Table 2) against the ancestry-modulated genes and we observed that 15 genes were in common with the low European ancestry group (*AIF1L*, *CYP19A1*, *CENPF*, *PGAP3*, *SLC4A8*, *ORMDL3*, *CRABP2*, *STARD3*, *BUB1*, *CDK12*, *SNORA54*, *HIPK2*, *FCGBP*, *RBBP8*, *NTRK2*). On the other hand, 12 genes were unique for this ancestry group and included *ERBB2*, *GRB7*, *MIEN1*, *ONECUT2*, *GSDMB*, *NUFIP2*, *TNFSF13*, *LRRC1*, *PSMD3*, *SLC23A3*, *ARHGAP33* and *HES1*. The high European ancestry group did not show common genes with the global profile or with the low European ancestry group and had 3 unique differentially expressed genes (*ATP8B3*, *FDXACB1*, and *RAB26*) (Fig 2 and Table 4).

Stratified analysis by Indigenous American ancestry fraction. We repeated the differential expression analysis stratifying by the IA average fraction into **low IA ancestry group** (IA ancestry fraction below 0.36), and **high IA ancestry group** (IA ancestry fraction above 0.36). The averages for the European, IA and African ancestry fractions according to assigned groups are in the S3 Table. Compared to the global profile of Luminal B tumors (Table 2), we found 5 genes in common with the high IA ancestry group (*ORMDL3*, *STARD3*, *SLC4A8*, *CDK12*, *HIPK2*) and 3 with the low IA ancestry group (*NT5E*, *SNORA76C*, *ABCA3*) (Fig 3).

In the low IA ancestry group, 6 genes were unique (*PCSK1*, *GABRA2*, *HTR1F*, *CNKSR3*, *PLCB1* y *RNU5A-1*). On the other hand, the unique genes found in the high IA ancestry group were *ERBB2*, *GRB7*, *GSDMB*, *MIEN1* and *ONECUT2*. These 5 genes were also found in the low European Ancestry group and the direction of the change was the same as that of the high IA ancestry group (Table 5).

Confirmatory analysis of candidate genes by RT-PCR

We selected *ERBB2*, *GRB7* and *ONECUT2* for validation due to their importance in the biology of breast cancer, the magnitude of the change found in the RNA-seq data analysis and the consistency between the European and IA ancestry analyses. Confirmatory analysis was

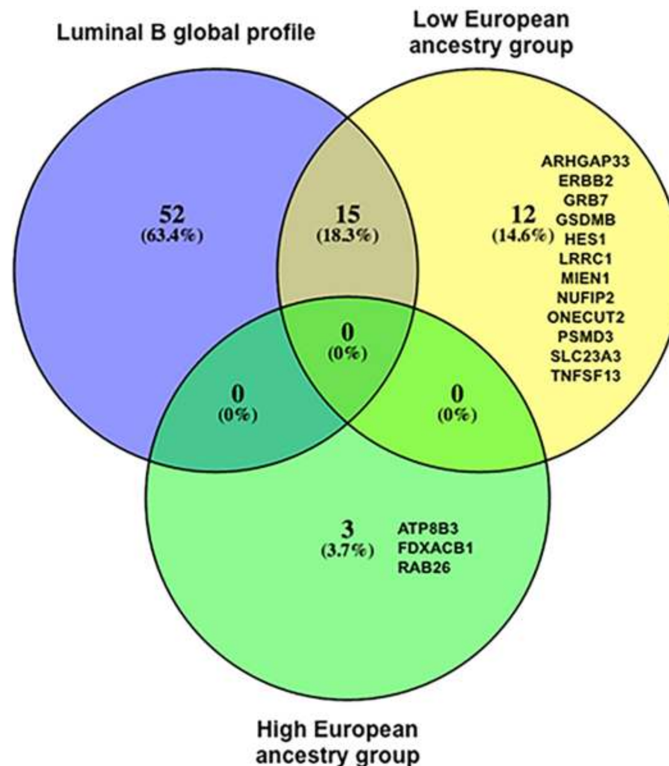


Fig 2. Differentially expressed genes according to European ancestry in IHC defined Luminal B vs. Luminal A tumors. Venn diagram shows the number of differentially expressed genes (padj < 0.05) between Luminal B and Luminal A tumors with low European ancestry and high European ancestry. "Global" refers to differentially expressed genes between Luminal tumors without stratification by ancestry.

https://doi.org/10.1371/journal.pone.0183179.g002

performed by semi-quantitative RT-PCR in 166 samples from Luminal tumors (42 analyzed by RNA-seq and 124 new samples). The characteristics of the 124 additional patients from the confirmatory analysis are in the S4 Table. Spearman correlation analysis showed statistically significant correlations in gene expression levels between RNA-seq and RT-PCR analysis for

Table 4. Differentially expressed genes for Luminal B vs. Luminal A tumors according to European ancestry group.

Differentially expressed genes unique for Luminal B tumors from the low European ancestry group			Differentially expressed genes unique for Luminal B tumors from the high European ancestry group		
Genes	**logFC**	**padj**	**Genes**	**logFC**	**padj**
ERBB2	2.367	1.48E-06	*ATP8B3*	-1.70	0.007
GRB7	2.327	3.15E-04	*FDXACB1*	-1.62	0.049
ONECUT2	2.204	1.28E-03	*RAB26*	-1.72	0.049
MIEN1	2.195	3.15E-04			
GSDMB	1.723	1.92E-03			
PSMD3	1.386	4.33E-02			
HES1	1.092	4.69E-02			
LRRC1	1.079	4.33E-02			
NUFIP2	1.071	6.27E-03			
ARHGAP33	-1.263	4.53E-02			
TNFSF13	-1.523	3.83E-02			
SLC23A3	-1.589	4.33E-02			

https://doi.org/10.1371/journal.pone.0183179.t004

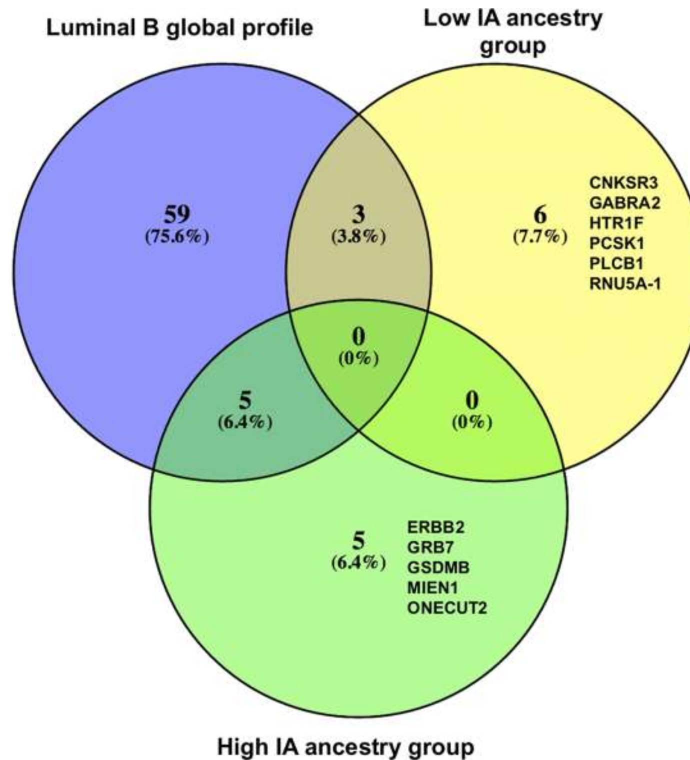


Fig 3. Differentially expressed genes according to IA ancestry in Luminal B vs. Luminal A tumors. Venn diagram shows the number of differentially expressed genes (padj < 0.05) between Luminal B and Luminal A tumors with low IA ancestry and high IA ancestry. "Global" refers to differentially expressed genes between Luminal tumors without stratification by ancestry.

https://doi.org/10.1371/journal.pone.0183179.g003

ERBB2 ($p < 0.01$, $R^2 = 0.62$), *ONECUT2* ($p = 0.014$, $R^2 = 0.62$), and *GRB7* ($p = 0.0131$, $R^2 = 0.40$) (S2A–S2C Fig).

We used a linear regression model to test if changes in expression levels of *ERBB2*, *ONECUT2* and *GRB7* were explained by genetic ancestry in Luminal subtypes. For this analysis we used log2FC values from the qRT-PCR. We tested the association of the expression levels of the candidate genes with an interaction variable between intrinsic subtype and genetic ancestry (European or IA). We found a statistically significant association between *ERBB2* expression and the IA fraction (p = 0.02, B = 3.11, CI 95% 0.43, 5.79), but not for the interaction (Table 6).

Concordantly, when we tested the association between *ERBB2* expression and the IA ancestry fraction stratified by Luminal subtype, we found an association of *ERBB2* expression and

Table 5. Differentially expressed genes in Luminal B tumors according to the Indigenous American ancestry groups.

Differentially expressed genes unique for Luminal B tumors from low IA ancestry group			Differentially expressed genes unique for Luminal B tumors from high IA ancestry group		
Genes	logFC	padj	Genes	logFC	padj
PCSK1	-3.34	3.5E-10	*GSDMB*	1.98	0.000
GABRA2	-2.27	1.7E-03	*MIEN1*	1.97	0.013
HTR1F	-2.23	1.8E-03	*ERBB2*	1.92	0.012
CNKSR3	-1.52	7.0E-03	*GRB7*	1.87	0.033
PLCB1	1.75	1.4E-02	*ONECUT2*	1.83	0.047
RNU5A-1	-1.07	4.0E-02			

https://doi.org/10.1371/journal.pone.0183179.t005

Table 6. Association between candidate genes expression and the interaction between Indigenous American ancestry and intrinsic subtype.

	ERBB2				*GRB7*				*ONECUT2*			
	B	p	IC 95%		B	p	IC 95%		B	p	IC 95%	
IA ancestry fraction	3.11	0.02	0.43	5.79	0.42	0.80	-2.82	3.65	-2.98	0.47	-11.22	5.26
Intrinsic subtype	0.84	0.27	-.065	2.34	-0.75	0.44	-2.68	1.18	-1.87	0.40	-6.38	2.63
Interaction	-1.60	0.36	-5.07	1.86	3.12	0.17	-1.37	7.61	6.60	0.18	-3.31	16.52

The association with the European ancestry fraction was not significant (Table 7).

https://doi.org/10.1371/journal.pone.0183179.t006

IA ancestry in the Luminal A group (p = 0.009, B = 3.111, CI 95% 0.821, 5.4), and the same trend was observed for the Luminal B group, in which patients with higher IA ancestry showed higher expression of *ERBB2* (Fig 4).

When we conducted the analysis by HER2 status we saw that the association between *ERBB2* expression and ancestry was also independent of immunohistochemical HER2 type (Fig 5).

All together these results suggest that *ERBB2* is a gene positively correlated with IA ancestry in Luminal breast cancer.

Discussion

Our group has previously shown that Luminal B is the most common intrinsic subtype of breast cancer in Colombian women [28]. Based on that finding we wanted to explore the underlying molecular characteristics of Luminal B tumors in Colombian-Latina women. We found 67 differentially expressed genes between Luminal B and Luminal A tumors. Genes that were up-regulated in luminal B tumors included, *CDK1*, *BUB1*, *CENPF* and *MKI67*, which participate in cell proliferation pathways consistent to what has been reported for the molecular profile in luminal B tumors in other population groups [7, 13, 14, 37–39].

Another up-regulated gene in luminal B tumors was *CYP19A1* that encodes aromatase, the enzyme that catalyzes the rate-limiting step in estrogen biosynthesis, aromatization of androstenedione and testosterone to estrone and estradiol, respectively [40, 41]. Aromatases are highly expressed in breast cancer tissue when compared to normal breast tissue [42] thus it has been suggested that the *CYP19A1* gene participates in the development and progression of breast cancer [41]. Aromatase inhibitors (AIs) that selectively inhibit aromatase activity in peripheral tissues have become a successful therapy for postmenopausal women with hormone-sensitive breast cancer [43, 44]. To the best of our knowledge, this is the first report to show differential expression of *CYP19A1* by luminal intrinsic subtype of breast cancer. The overexpression of this gene in Luminal B tumors in Colombian patients suggests that this subtype may be more sensitive to aromatase inhibitors compared to Luminal A tumors.

DNA topoisomerase IIA (TOP2A) is an isoform of TOP2 enzyme that exerts catalytic activity to induce breaks in double-stranded DNA to release torsional stress. These breaks are subsequently resealed [45]. Sparano et al. [46] suggested that in breast cancer patients with

Table 7. Association between candidate genes expression and the interaction between European ancestry and intrinsic subtype.

	ERBB2				*GRB7*				*ONECUT2*			
	B	p	IC 95%		B	p	IC 95%		B	p	IC 95%	
European ancestry fraction	1.00	0.44	-1.56	3.55	-0.08	0.96	-3.16	2.99	-5.69	0.13	-13.11	1.74
Intrinsic subtype	0.89	0.32	-0.89	2.68	1.10	0.35	-1.20	3.39	0.37	0.87	-4.09	4.84
Interaction	-1.43	0.40	-4.80	1.93	-1.10	0.62	-5.44	3.24	1.91	0.67	-6.99	10.80

https://doi.org/10.1371/journal.pone.0183179.t007

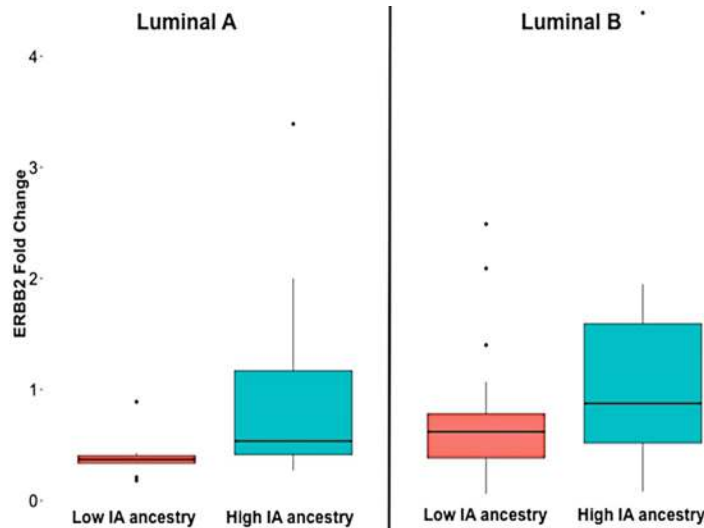


Fig 4. *ERBB2* expression according to intrinsic subtype and Indigenous American ancestry group.

https://doi.org/10.1371/journal.pone.0183179.g004

hormone receptor positive and HER2-normal expression, high levels of *TOP2A* could be associated with resistance to antracycline-based chemotherapy. This suggestion came from the finding that higher expression of TOP2A correlated with poor tumor grade and high recurrence score based on the Oncotype Dx signature. Romero et al. [45] found higher expression of TOP2A in Luminal B, HER2-enriched and basal-like when compared to Luminal A subtype, which is consistent with our finding.

As Hispanic/Latinas represent a heterogeneous population group with variation in the European, IA and African ancestry fractions [47], we explored the role of genetic ancestry as a modifier of the molecular characteristics of Luminal tumors in Colombian women. We found 5 genes potentially modulated by genetic ancestry and differentially expressed between Luminal B and Luminal A tumors (*ERBB2*, *GRB7*, *ONECUT2*, *MIEN1* and *GSDMB*). These genes were ancestry-modulated in the analysis based on the European ancestry categories as well as the IA categories.

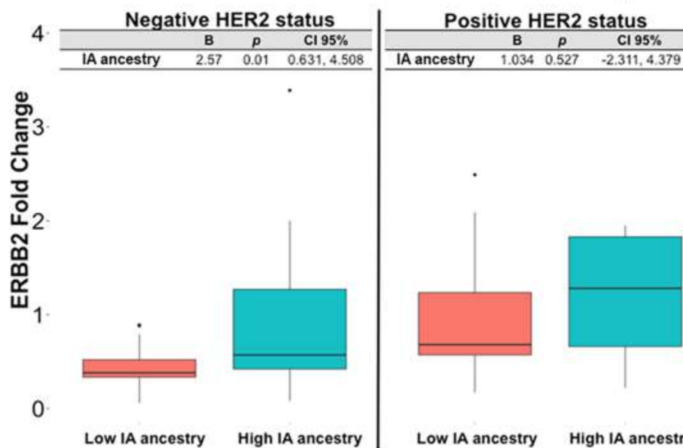


Fig 5. *ERBB2* expression according to HER2 status by immunohistochemistry and Indigenous American ancestry group.

https://doi.org/10.1371/journal.pone.0183179.g005

The *ERBB2*, *GRB7* and *MIEN1* genes, which are located on chromosome 17 in relative proximity within a region including approximately 60,000 base pairs, have been reported co-amplified and associated with poor prognosis in breast cancer [48, 49]. *ERBB2* is located in locus 17q12 and encodes a 185 KDa transmembrane glycoprotein receptor that belongs to the family of the epidermal growth factor receptor (EGFR) [50–52]. It has been reported over-expressed or amplified in 30% of breast tumors and also in ovarian, gastric, and uterine tumors [53]. *GRB7* gene is located in locus 17q11–21 and encodes for an adapter protein present in the cellular cytoplasm and interacts with multiple proteins including tyrosine kinase receptors such as EGFR and ERBB2 (HER2) through its SH2 domains [54, 55]. *GRB7* has been found to regulate migration [56–58] and recently Nadler et al. [54] found that HER2/GRB7 co-expression conferred worse prognosis than *HER2* amplification alone, and that high expression of GRB7 at the protein level is associated with shorter survival times. Finally, *MIEN1*, located in the chromosomal region 17q12-21, was recently discovered [49, 59] and has been associated with enhanced migration in several types of cancer [48, 60]. To the best of our knowledge, this report is the first to show differential expression of *MIEN1* by breast cancer intrinsic subtype and genetic ancestry and to show differential expression of *ERBB2* and *GRB7* by genetic ancestry in breast cancer patients.

After assessing the correlation between RNA-seq and RT-PCR assays, we were able to validate our observation of effect modification of Luminal B vs. Luminal A tumors differentially expressed genes by genetic ancestry for three of the five genes (*ERBB2*, *GRB7 and ONECUT2)*. In the validation set we analyzed a higher number of patients with Luminal tumors (124 additional luminal tumors) and found a significant statistical association between *ERBB2* expression levels and IA ancestry fraction, which seem to be independent of Luminal subtype and immunohistochemical HER2 characterization. This finding suggests that Hispanic/Latina women with higher IA ancestry are more likely to develop Luminal tumors with higher expression of *ERBB2* compared to women with higher European ancestry. However, the relationship between expression of *ERBB2* and HER2 IHC classification needs to be better understood, given that the association between ancestry and *ERBB2* was not paralleled by the association between genetic ancestry and Luminal subtype based on IHC classification. More patients, not only with Luminal tumors but all different subtypes, will be needed to replicate this finding and explore the prognostic value of the association and relevance for the use of trastuzumab treatment in this population. Future studies should confirm if these three genes (*ERBB2*, *GRB7* and *MIEN1*) as they have been reported co-amplified, are also ancestry-modulated together or if the modulation is independent of their co-amplification.

This is the first study, to the best of our knowledge, to explore differences in the molecular profile of an intrinsic subtype of breast cancer according to genetic ancestry in a highly admixed Latin American population. Some studies have compared the molecular profiles of breast cancer between Caucasian and African American women [61–65] in order to seek mechanistic explanations for the differences in disease biology and outcomes observed between these two populations. However, only few studies have included Hispanic/Latina women. Chavez-MacGregor et al. [66] explored differences in the transcriptome and protein expression according to race/ethnicity and intrinsic subtype in 376 women (46 African-American, 47 Hispanic/Latinas and 147 Non- Hispanic white women). They did not find any statistically significant differences in the molecular profiles or at the protein level between the groups they analyzed. However, they did not include genetic ancestry in their analyses, and therefore were unable to assess more subtle differences in expression based on the ancestral genetic architecture of the admixed genomes of African Americans and Hispanic/Latinas.

We are aware of the limitations of our study. First, the approximation of intrinsic subtypes based on IHC and not gene expression could lead to misclassification of tumors. However,

when we compared the percent expression of ER/PR/HER2/Ki67 based on IHC and gene expression, we found relatively strong correlations. We are also aware that to find differentially expressed genes by genetic ancestry it would have been more informative to analyze patients with higher variations in their European, IA and African fractions. Nevertheless, by analyzing patients from only one Colombian region (Andean), we aimed at reducing possible differences in expression due to variation in environmental exposures. One significant advantage of the present study was the fact that we analyzed gene expression differences in an admixed group of patients according to their genetic ancestry and not by their self-identification. Finally, it is important to highlight that all the RNA-seq and RT-PCR data shown in this study were obtained from FFPE samples, which confirms their value as a source of information for future work.

Conclusions

Our results suggest that the expression of *ERBB2*, a crucial gene in breast cancer tumor subtype classification associated with poor prognosis, might be associated with genetic ancestry in breast tumor samples from Colombian women. Women with higher IA ancestry express higher levels of *ERBB2*. Further analyses are necessary to further confirm this association, assess the impact that this association has on HER2 IHC classification, and explore its prognostic value.

Supporting information

S1 Fig. Pearson correlation analysis for expression levels of ER, PR, Ki67 and HER2 assessed by immunohistochemistry and RNA-seq in 42 sequenced patients. X-axis represents the measure by immunohistochemistry and Y-axis measure by RNA-seq. (**A**) Scatter plot from Progesterone receptor expression. (**B**) Scatter plot for estrogen receptor expression. (**C**) Scatter plot for Ki67 expression. (**D**) Scatter Plot for HER2 expression.
(PDF)

S2 Fig. Spearman correlation plots show a positive correlation between gene expression values obtained by RNA-seq and RT-PCR. (**A**) Scatter plot for *ERBB2*. (**B**) Scatter plot for *GRB7*. (**C**) Spearman correlation for *ONECUT2*.
(PDF)

S1 Table. Association between the expression level of CYP19A1, TOP2A and intrinsic subtype with the presence of recurrences in 42 luminal tumors from Colombian patients.
(PDF)

S2 Table. Average of different ancestry fractions when patients where stratified according to European ancestry fraction.
(PDF)

S3 Table. Average ancestry fractions when patients where stratified according to Indigenous American (IA) ancestry fraction.
(PDF)

S4 Table. Characteristics of the patients analyzed by qRT-PCR included in the validation set.
(PDF)

Acknowledgments

We are most grateful to the patients who participate in this study. We also want to thank Camilo Silva M.D. and Alejandro Romero M.D. for their support in the pathological review of

the cases included in the study and Luli Gonzalez and Yuli Castro for their technical support in the immunohistochemistry.

Author Contributions

Conceptualization: Silvia J. Serrano-Gómez, María Carolina Sanabria-Salas, Gustavo Hernández-Suarez, Juan Carlos Mejía, Oscar García, Lucio Miele, Laura Fejerman, Jovanny Zabaleta.

Data curation: Melody C. Baddoo.

Formal analysis: Silvia J. Serrano-Gómez, Laura Fejerman, Jovanny Zabaleta.

Funding acquisition: María Carolina Sanabria-Salas, Laura Fejerman, Jovanny Zabaleta.

Investigation: Silvia J. Serrano-Gómez, Jone Garay.

Methodology: Silvia J. Serrano-Gómez, María Carolina Sanabria-Salas, Laura Fejerman, Jovanny Zabaleta.

Project administration: Silvia J. Serrano-Gómez, María Carolina Sanabria-Salas, Juan Carlos Mejía, Jovanny Zabaleta.

Resources: María Carolina Sanabria-Salas, Jovanny Zabaleta.

Supervision: María Carolina Sanabria-Salas, Gustavo Hernández-Suarez, Lucio Miele, Laura Fejerman, Jovanny Zabaleta.

Validation: Silvia J. Serrano-Gómez, Jone Garay, Laura Fejerman, Jovanny Zabaleta.

Visualization: Silvia J. Serrano-Gómez, María Carolina Sanabria-Salas, Laura Fejerman, Jovanny Zabaleta.

Writing – original draft: Silvia J. Serrano-Gómez, Laura Fejerman, Jovanny Zabaleta.

Writing – review & editing: Silvia J. Serrano-Gómez, María Carolina Sanabria-Salas, Lucio Miele, Laura Fejerman, Jovanny Zabaleta.

References

1. Blows FM, Driver KE, Schmidt MK, Broeks A, Van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS medicine. 2010; 7(5): e1000279. https://doi.org/10.1371/journal.pmed.1000279 PMID: 20520800; PubMed Central PMCID: PMC2876119.
2. Eroles P, Bosch A, Perez-Fidalgo JA, Lluch A. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. Cancer treatment reviews. 2012; 38(6):698–707. https://doi.org/10.1016/j.ctrv.2011.11.005 PMID: 22178455.
3. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490(7418):61–70. https://doi.org/10.1038/nature11412 PMID: 23000897; PubMed Central PMCID: PMC3465532.
4. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. The Lancet. 2011; 378(9805):1812–23. https://doi.org/10.1016/s0140-6736(11)61539-0
5. 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(17).
6. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. Journal of the American Medical Association. 2006; 295(21).
7. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National

Academy of Sciences of the United States of America. 2001; 98(19):10869–74. https://doi.org/10.1073/pnas.191367098 PMID: 11553815; PubMed Central PMCID: PMC58566.

8. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(18):10393–8. https://doi.org/10.1073/pnas.1732912100 PMID: 12917485; PubMed Central PMCID: PMC193572.
9. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast. 2015; 24 Suppl 2:S26–35. https://doi.org/10.1016/j.breast.2015.07.008 PMID: 26253814.
10. Li ZH, Hu PH, Tu JH, Yu NS. Luminal B breast cancer: patterns of recurrence and clinical outcome. Oncotarget. 2016; 7(40):65024–33. https://doi.org/10.18632/oncotarget.11344 PMID: 27542253.
11. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC genomics. 2006; 7:96. https://doi.org/10.1186/1471-2164-7-96 PMID: 16643655; PubMed Central PMCID: PMC1468408.
12. Prat A, Cheang MC, Martin M, Parker JS, Carrasco E, Caballero R, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2013; 31(2):203–9. https://doi.org/10.1200/JCO.2012.43.4134 PMID: 23233704; PubMed Central PMCID: PMC3532392.
13. Ades F, Zardavas D, Bozovic-Spasojevic I, Pugliano L, Fumagalli D, de Azambuja E, et al. Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014; 32(25):2794–803. https://doi.org/10.1200/JCO.2013.54.1870 PMID: 25049332.
14. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. Journal of the National Cancer Institute. 2009; 101(10):736–50. https://doi.org/10.1093/jnci/djp082 PMID: 19436038; PubMed Central PMCID: PMC2684553.
15. von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2012; 30(15):1796–804. https://doi.org/10.1200/JCO.2011.38.8595 PMID: 22508812.
16. Gianni L, Zambetti M, Clark K, Baker J, Cronin M, Wu J, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2005; 23(29):7265–77. https://doi.org/10.1200/JCO.2005.02.0818 PMID: 16145055.
17. de Ronde JJ, Hannemann J, Halfwerk H, Mulder L, Straver ME, Vrancken Peeters MJ, et al. Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response. Breast cancer research and treatment. 2010; 119(1):119–26. https://doi.org/10.1007/s10549-009-0499-6 PMID: 19669409.
18. Lips EH, Mulder L, de Ronde JJ, Mandjes IA, Koolen BB, Wessels LF, et al. Breast cancer subtyping by immunohistochemistry and histological grade outperforms breast cancer intrinsic subtypes in predicting neoadjuvant chemotherapy response. Breast cancer research and treatment. 2013; 140(1):63–71. https://doi.org/10.1007/s10549-013-2620-0 PMID: 23828499; PubMed Central PMCID: PMC3706735.
19. Tishchenko I, Milioli HH, Riveros C, Moscato P. Extensive Transcriptomic and Genomic Analysis Provides New Insights about Luminal Breast Cancers. PloS one. 2016; 11(6):e0158259. https://doi.org/10.1371/journal.pone.0158259 PMID: 27341628; PubMed Central PMCID: PMCPMC4920434.
20. Staaf J, Ringner M. Making breast cancer molecular subtypes robust? Journal of the National Cancer Institute. 2015; 107(1):386. https://doi.org/10.1093/jnci/dju386 PMID: 25479803; PubMed Central PMCID: PMC4301705.
21. Lusa L, McShane LM, Reid JF, De Cecco L, Ambrogi F, Biganzoli E, et al. Challenges in projecting clustering results across gene expression-profiling datasets. Journal of the National Cancer Institute. 2007; 99(22):1715–23. https://doi.org/10.1093/jnci/djm216 PMID: 18000217.
22. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006; 10(6):529–41. https://doi.org/10.1016/j.ccr.2006.10.009 PMID: 17157792.
23. Creighton CJ, Kent Osborne C, van de Vijver MJ, Foekens JA, Klijn JG, Horlings HM, et al. Molecular profiles of progesterone receptor loss in human breast tumors. Breast cancer research and treatment. 2009; 114(2):287–99. https://doi.org/10.1007/s10549-008-0017-2 PMID: 18425577; PubMed Central PMCID: PMCPMC2635926.

24. Creighton CJ. The molecular profile of luminal B breast cancer. Biologics: targets & therapy. 2012; 6:289–97. https://doi.org/10.2147/BTT.S29923 PMID: 22956860; PubMed Central PMCID: PMC3430090.

25. Tran B, Bedard PL. Luminal-B breast cancer and novel therapeutic targets. Breast cancer research: BCR. 2011; 13(6):221. https://doi.org/10.1186/bcr2904 PMID: 22217398; PubMed Central PMCID: PMCPMC3326541.

26. Fejerman L, Hu D, Huntsman S, John EM, Stern MC, Haiman CA, et al. Genetic ancestry and risk of mortality among U.S. Latinas with breast cancer. Cancer research. 2013; 73(24):7243–53. https://doi.org/10.1158/0008-5472.CAN-13-2014 PMID: 24177181; PubMed Central PMCID: PMC3881587.

27. Rojas W, Parra MV, Campo O, Caro MA, Lopera JG, Arias W, et al. Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. Am J Phys Anthropol. 2010; 143(1).

28. Serrano-Gomez SJ, Sanabria-Salas MC, Hernandez-Suarez G, Garcia O, Silva C, Romero A, et al. High prevalence of luminal B breast cancer intrinsic subtype in Colombian women. Carcinogenesis. 2016. https://doi.org/10.1093/carcin/bgw043 PMID: 27207651.

29. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO. 2013; 24(9):2206–23. https://doi.org/10.1093/annonc/mdt303 PMID: 23917950; PubMed Central PMCID: PMC3755334.

30. Fejerman L, John EM, Huntsman S, Beckman K, Choudhry S, Perez-Stable E, et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. Cancer research. 2008; 68(23):9723–8. https://doi.org/10.1158/0008-5472.CAN-08-2039 PMID: 19047150; PubMed Central PMCID: PMC2674787.

31. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2):945–59. PubMed Central PMCID: PMC1461096. PMID: 10835412

32. Aronesty E. Command-line tools for processing biological sequencing data"; Expression Analysis, Durham, NC 2011. Available from: http://code.google.com/p/ea-utils.

33. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040; PubMed Central PMCID: PMCPMC3163565.

34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.

35. Dennis G Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology. 2003; 4(5):P3. PMID: 12734009.

36. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2013; 31(31):3997–4013. https://doi.org/10.1200/JCO.2013.50.9984 PMID: 24101045.

37. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol. 2014; 5(3):412–24. https://doi.org/10.5306/wjco.v5.i3.412 PMID: 25114856; PubMed Central PMCID: PMC4127612.

38. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(14):8418–23. https://doi.org/10.1073/pnas.0932692100 PMID: 12829800; PubMed Central PMCID: PMC166244.

39. Huang X, Dugo M, Callari M, Sandri M, De Cecco L, Valeri B, et al. Molecular portrait of breast cancer in China reveals comprehensive transcriptomic likeness to Caucasian breast cancer and low prevalence of luminal A subtype. Cancer Med. 2015. https://doi.org/10.1002/cam4.442 PMID: 25787708.

40. Artigalas O, Vanni T, Hutz MH, Ashton-Prolla P, Schwartz IV. Influence of CYP19A1 polymorphisms on the treatment of breast cancer with aromatase inhibitors: a systematic review and meta-analysis. BMC Med. 2015; 13:139. https://doi.org/10.1186/s12916-015-0373-9 PMID: 26067721; PubMed Central PMCID: PMCPMC4475294.

41. Long JR, Kataoka N, Shu XO, Wen W, Gao YT, Cai Q, et al. Genetic polymorphisms of the CYP19A1 gene and breast cancer survival. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006; 15(11):2115–22. https://doi.org/10.1158/1055-9965.EPI-06-0464 PMID: 17119036.

42. Irahara N, Miyoshi Y, Taguchi T, Tamaki Y, Noguchi S. Quantitative analysis of aromatase mRNA expression derived from various promoters (I.4, I.3, PII and I.7) and its association with expression of

TNF-alpha, IL-6 and COX-2 mRNAs in human breast cancer. International journal of cancer Journal international du cancer. 2006; 118(8):1915–21. https://doi.org/10.1002/ijc.21562 PMID: 16287071.

43. Rugo HS. The breast cancer continuum in hormone-receptor-positive breast cancer in postmenopausal women: evolving management options focusing on aromatase inhibitors. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO. 2008; 19(1):16–27. https://doi.org/10.1093/annonc/mdm282 PMID: 17693420.
44. Kristensen VN, Sorlie T, Geisler J, Yoshimura N, Linegjaerde OC, Glad I, et al. Effects of anastrozole on the intratumoral gene expression in locally advanced breast cancer. J Steroid Biochem Mol Biol. 2005; 95(1–5):105–11. https://doi.org/10.1016/j.jsbmb.2005.04.028 PMID: 16023338.
45. Romero A, Martin M, Cheang MC, Lopez Garcia-Asenjo JA, Oliva B, He X, et al. Assessment of Topoisomerase II alpha status in breast cancer by quantitative PCR, gene expression microarrays, immunohistochemistry, and fluorescence in situ hybridization. Am J Pathol. 2011; 178(4):1453–60. https://doi.org/10.1016/j.ajpath.2010.12.042 PMID: 21435434; PubMed Central PMCID: PMCPMC3078442.
46. Sparano JA, Goldstein LJ, Childs BH, Shak S, Brassard D, Badve S, et al. Relationship between Topoisomerase 2A RNA Expression and Recurrence after Adjuvant Chemotherapy for Breast Cancer. Clinical cancer research: an official journal of the American Association for Cancer Research. 2009; 15 (24):7693–700. https://doi.org/10.1158/1078-0432.CCR-09-1450 PMID: 19996222; PubMed Central PMCID: PMCPMC3396025.
47. Stern MC, Fejerman L, Das R, Setiawan VW, 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; 3:181–90. https://doi.org/10.1007/s40471-016-0083-7 PMID: 27547694; PubMed Central PMCID: PMCPMC4978756.
48. Kpetemey M, Dasgupta S, Rajendiran S, Das S, Gibbs LD, Shetty P, et al. MIEN1, a novel interactor of Annexin A2, promotes tumor cell migration by enhancing AnxA2 cell surface expression. Mol Cancer. 2015; 14:156. https://doi.org/10.1186/s12943-015-0428-8 PMID: 26272794; PubMed Central PMCID: PMCPMC4536591.
49. Evans EE, Henn AD, Jonason A, Paris MJ, Schiffhauer LM, Borrello MA, et al. C35 (C17orf37) is a novel tumor biomarker abundantly expressed in breast cancer. Molecular cancer therapeutics. 2006; 5 (11):2919–30. https://doi.org/10.1158/1535-7163.MCT-06-0389 PMID: 17121940.
50. Wu Y, Mohamed H, Chillar R, Ali I, Clayton S, Slamon D, et al. Clinical significance of Akt and HER2/neu overexpression in African-American and Latina women with breast cancer. Breast cancer research: BCR. 2008; 10(1):R3. https://doi.org/10.1186/bcr1844 PMID: 18184439; PubMed Central PMCID: PMC2374954.
51. Yu D, Hung MC. Overexpression of ErbB2 in cancer and ErbB2-targeting strategies. Oncogene. 2000; 19(53):6115–21. https://doi.org/10.1038/sj.onc.1203972 PMID: 11156524.
52. Bertucci F, Borie N, Ginestier C, Groulet A, Charafe-Jauffret E, Adelaide J, et al. Identification and validation of an ERBB2 gene expression signature in breast cancers. Oncogene. 2004; 23(14):2564–75. https://doi.org/10.1038/sj.onc.1207361 PMID: 14743203.
53. Tao Y, Shen C, Luo S, Traore W, Marchetto S, Santoni MJ, et al. Role of Erbin in ErbB2-dependent breast tumor growth. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(42):E4429–38. https://doi.org/10.1073/pnas.1407139111 PMID: 25288731; PubMed Central PMCID: PMC4210290.
54. Nadler Y, Gonzalez AM, Camp RL, Rimm DL, Kluger HM, Kluger Y. Growth factor receptor-bound protein-7 (Grb7) as a prognostic marker and therapeutic target in breast cancer. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO. 2010; 21(3):466–73. https://doi.org/10.1093/annonc/mdp346 PMID: 19717535; PubMed Central PMCID: PMCPMC2826097.
55. Stein D, Wu J, Fuqua SA, Roonprapunt C, Yajnik V, D'Eustachio P, et al. The SH2 domain protein GRB-7 is co-amplified, overexpressed and in a tight complex with HER2 in breast cancer. The EMBO journal. 1994; 13(6):1331–40. PMID: 7907978; PubMed Central PMCID: PMCPMC394949.
56. Han DC, Shen TL, Guan JL. The Grb7 family proteins: structure, interactions with other signaling molecules and potential cellular functions. Oncogene. 2001; 20(44):6315–21. https://doi.org/10.1038/sj.onc.1204775 PMID: 11607834.
57. Han DC, Guan JL. Association of focal adhesion kinase with Grb7 and its role in cell migration. The Journal of biological chemistry. 1999; 274(34):24425–30. PMID: 10446223.
58. Vinatzer U, Dampier B, Streubel B, Pacher M, Seewald MJ, Stratowa C, et al. Expression of HER2 and the coamplified genes GRB7 and MLN64 in human breast cancer: quantitative real-time reverse transcription-PCR as a diagnostic alternative to immunohistochemistry and fluorescence in situ hybridization. Clinical cancer research: an official journal of the American Association for Cancer Research. 2005; 11(23):8348–57. https://doi.org/10.1158/1078-0432.CCR-05-0841 PMID: 16322295.

59. Katz E, Dubois-Marshall S, Sims AH, Faratian D, Li J, Smith ES, et al. A gene on the HER2 amplicon, C35, is an oncogene in breast cancer whose actions are prevented by inhibition of Syk. British journal of cancer. 2010; 103(3):401–10. https://doi.org/10.1038/sj.bjc.6605763 PMID: 20628393; PubMed Central PMCID: PMCPMC2920017.

60. Kpetemey M, Chaudhary P, Van Treuren T, Vishwanatha JK. MIEN1 drives breast tumor cell migration by regulating cytoskeletal-focal adhesion dynamics. Oncotarget. 2016; 7(34):54913–24. https://doi.org/10.18632/oncotarget.10798 PMID: 27462783.

61. Field LA, Love B, Deyarmin B, Hooke JA, Shriver CD, Ellsworth RE. Identification of differentially expressed genes in breast tumors from African American compared with Caucasian women. Cancer. 2012; 118(5):1334–44. https://doi.org/10.1002/cncr.26405 PMID: 21800289.

62. Martin DN, Boersma BJ, Yi M, Reimers M, Howe TM, Yfantis HG, et al. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. PloS one. 2009; 4(2):e4531. https://doi.org/10.1371/journal.pone.0004531 PMID: 19225562; PubMed Central PMCID: PMC2638012.

63. Grunda JM, Steg AD, He Q, Steciuk MR, Byan-Parker S, Johnson MR, et al. Differential expression of breast cancer-associated genes between stage- and age-matched tumor specimens from African- and Caucasian-American Women diagnosed with breast cancer. BMC research notes. 2012; 5:248. https://doi.org/10.1186/1756-0500-5-248 PMID: 22616718; PubMed Central PMCID: PMC3476447.

64. D'Arcy M, Fleming J, Robinson WR, Kirk EL, Perou CM, Troester MA. Race-associated biological differences among Luminal A breast tumors. Breast cancer research and treatment. 2015; 152(2):437–48. https://doi.org/10.1007/s10549-015-3474-4 PMID: 26109344; PubMed Central PMCID: PMCPMC4527078.

65. Stewart PA, Luks J, Roycik MD, Sang QX, Zhang J. Differentially expressed transcripts and dysregulated signaling pathways and networks in African American breast cancer. PloS one. 2013; 8(12): e82460. https://doi.org/10.1371/journal.pone.0082460 PMID: 24324792; PubMed Central PMCID: PMC3853650.

66. Chavez-Macgregor M, Liu S, De Melo-Gagliato D, Chen H, Do KA, Pusztai L, et al. Differences in gene and protein expression and the effects of race/ethnicity on breast cancer subtypes. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2014; 23(2):316–23. https://doi.org/10.1158/1055-9965.EPI-13-0929 PMID: 24296856; PubMed Central PMCID: PMC3946290.