

## Integrated copy number and miRNA expression analysis in triple negative breast cancer of Latin American patients

Bruna M. Sugita<sup>1,2</sup>, Silma R. Pereira<sup>3</sup>, Rodrigo C. de Almeida<sup>4</sup>, Mandeep Gill<sup>5</sup>, Akanksha Mahajan<sup>5</sup>, Anju Duttargi<sup>5</sup>, Saurabh Kirolikar<sup>5</sup>, Paolo Fadda<sup>6</sup>, Rubens S. de Lima<sup>7</sup>, Cicero A. Urban<sup>7</sup>, Kepher Makambi<sup>8</sup>, Subha Madhavan<sup>5,9</sup>, Simina M. Boca<sup>5,8,9</sup>, Yuriy Gusev<sup>5,9</sup>, Iglener J. Cavalli<sup>1</sup>, Enilze M.S.F. Ribeiro<sup>1</sup>, Luciane R. Cavalli<sup>2,5</sup>

<sup>1</sup>Department of Genetics, Federal University of Paraná, Curitiba, PR, Brazil

<sup>2</sup>Faculdades Pequeno Príncipe, Instituto de Pesquisa Pelé Pequeno Príncipe, Curitiba, PR, Brazil

<sup>3</sup>Department of Biology, Federal University of Maranhão, São Luis, MA, Brazil

<sup>4</sup>Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, Netherlands

<sup>5</sup>Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington DC, USA

<sup>6</sup>Genomics Shared Resource, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA

<sup>7</sup>Breast Unit, Hospital Nossa Senhora das Graças, Curitiba, PR, Brazil

<sup>8</sup>Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center, Washington DC, USA

<sup>9</sup>Innovation Center for Biomedical Informatics (ICBI), Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington DC, USA

**Correspondence to:** Luciane R. Cavalli, **email:** luciane.cavalli@pelepequenoprincipe.org.br

**Keywords:** microRNA; triple-negative breast cancer; copy number; latinas; disparities

**Received:** July 11, 2019

**Accepted:** September 16, 2019

**Published:** October 22, 2019

**Copyright:** Sugita et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**Triple negative breast cancer (TNBC), a clinically aggressive breast cancer subtype, affects 15–35% of women from Latin America. Using an approach of direct integration of copy number and global miRNA profiling data, performed simultaneously in the same tumor specimens, we identified a panel of 17 miRNAs specifically associated with TNBC of ancestrally characterized patients from Latin America, Brazil. This panel was differentially expressed between the TNBC and non-TNBC subtypes studied ( $p \leq 0.05$ , FDR  $\leq 0.25$ ), with their expression levels concordant with the patterns of copy number alterations (CNAs), present mostly frequent at 8q21.3-q24.3, 3q24-29, 6p25.3-p12.2, 1q21.1-q44, 5q11.1-q22.1, 11p13-p11.2, 13q12.11-q14.3, 17q24.2-q25.3 and Xp22.33-p11.21. The combined 17 miRNAs presented a high power (AUC = 0.953 (0.78–0.99);95% CI) in discriminating between the TNBC and non-TNBC subtypes of the patients studied. In addition, the expression of 14 and 15 of the 17miRNAs was significantly associated with tumor subtype when adjusted for tumor stage and grade, respectively. In conclusion, the panel of miRNAs identified demonstrated the impact of CNAs in miRNA expression levels and identified miRNA target genes potentially affected by both CNAs and miRNA deregulation. These targets, involved in critical signaling pathways and biological functions associated specifically with the TNBC transcriptome of Latina patients, can provide biological insights into the observed differences in the TNBC clinical outcome among racial/ethnic groups, taking into consideration their genetic ancestry.**

## INTRODUCTION

Breast cancer is the leading cause of death in women from Latin America living in the US or in Latin American countries (referred here as Latinas). The overall incidence and mortality rates of breast cancer in this population in the US were 91.9 and 14.0 per 100,000 in 2015 [1, 2]. In Brazil, 59,700 new cases of breast cancer were expected in 2018 [3]. The highest mortality rates were observed among Puerto Ricans (19.04), Mexicans (18.78) and Cubans (17.89), and the lowest rate was observed among Central and South Americans (10.15) [4]. The breast cancer 5-year survival rate in Latin America hardly exceeds 70%, which is usually correlated with late diagnosis; approximately 30%–40% of breast cancer patients in the Latin American countries, including Brazil, are diagnosed in stages III and IV of the disease [5].

There is a well-documented disparity of breast cancer in Latinas when compared to non-Hispanic Whites (NHW); Latinas are more likely to present with non-localized disease, receive less aggressive therapy, and have a disproportionately low survival rate when compared to NHW [6–8]. Several factors can contribute to the increased mortality rate in this population group, including socio-economic barriers, which can limit their overall access to early detection and cancer prevention services [9, 10], and presence of co-morbidities, such as obesity and diabetes [11–16]. Many deaths in this population can also be attributed to after-diagnosis factors, such as inadequate access to appropriate treatment or early treatment interruption or discontinuation [17–19].

Epidemiological and molecular studies have shown that breast cancer subtypes are distributed unevenly among various racial/ethnic groups [6, 7, 20–22]. The incidence of the triple negative breast cancer (TNBC) subtype in particular, one of the most clinically aggressive breast cancer subtypes, largely varies according to ethnicity; it is present at higher frequencies in African-American (AA) (24–42%) and Hispanic/Latina (15–33%) women when compared to NHWs (11–28%) [6, 23, 24]. As in AA women, Latina patients with TNBC are more often diagnosed at an earlier age, with advanced stage, likely to experience metastasis and be refractory to treatment [25–29]. In addition to the general attributed socio-economic factors mentioned above, tumor biology and genetic background play a significant role in such disparity [11, 14, 30–33], however the individual contribution of each of these factors to their observed poor clinical outcome remains unclear. There is also lack of basic science studies and clinical trials that are conducted in Latinas [34, 35] which limits the knowledge of the biological causes that contributes to these tumors' clinical aggressiveness.

MicroRNAs (miRNAs) are a class of non-coding endogenous RNA molecules that have been identified to play a role in breast tumorigenesis [36, 37]. miRNA expression has been shown to present different expression

patterns according to the intrinsic molecular breast cancer subtype [38, 39]. In TNBC, for instance, distinct miRNAs were reported mediating cellular processes associated with aggressive tumor phenotypes, such as the ones that promote metastatic development [40–43] and treatment resistance [44, 45]. Interestingly, miRNAs were also shown to present variable expression according to race and/or ethnicity [46–50]; a number of studies have shown miRNA polymorphisms in association with the susceptibility risk of breast cancer in specific ethnic populations [49, 51–54]. Data on somatic miRNA expression levels in the breast tissue of these populations, are however, scarce [50, 55].

Although the characterization of the genomic profiles in TNBC has been extensively performed, few studies have characterized them in specific ethnic groups, such as Latinas. This translates to a deficiency in the understanding of the intrinsic characteristics of their tumors' genome, which can differentially impact their tumor phenotypes and clinical behavior. Therefore, in this study our main aim was to determine the patterns of DNA copy number and miRNA expression of TNBC of ancestrally genomic characterized patients living in Latin America, Brazil. In addition, we aimed to determine whether copy number alterations (CNAs) could impact miRNA expression levels and whether there were common miRNA target genes affected by both CNAs and miRNA deregulation.

A distinct pattern of CNAs and miRNA expression was observed between the TNBC and non-TNBC cases analysed; TNBC cases presented a higher number of CNAs when compared to the non-TNBC, affecting distinct chromosome cytobands. The integrated analysis of CNAs and miRNA expression revealed 17 miRNAs differentially expressed between these subtypes, with their expression levels concordant with the patterns of the CNAs. This 17miRNA panel presented a high combined power in discriminating the TNBC and non-TNBC subtypes and most of these miRNAs were significantly associated with these subtypes when adjusting for tumor stage and grade. miRNA target and functional enrichment analysis showed the involvement of these miRNAs in specific cancer signaling pathways, which can distinctly impact the biological phenotype and clinical outcome of the TNBC patients studied.

## RESULTS

### DNA copy number profiling

DNA copy number analysis was performed by array-CGH for 25 cases of TNBC and 16 cases of non-TNBC subtypes of the patients studied. A total number of 292 and 204 CNAs (as measured by the “number of calls” per the aberration interval base reports (Agilent-CytoGenomics v.3.0) were identified in the TNBC and non-TNBC

subtypes, with an average of  $18.25 \pm 5.98$  and  $14.57 \pm 3.09$  CNAs per case, respectively. These differences were not significant at  $p < 0.05$  (Unpaired *t* test;  $p = 0.5902$ ).

In the TNBC cases, 64.0% of the cases (16/25) showed significant CNAs; the most frequent cytobands affected were: 8q21.3-q24.3 (43.8% of the cases), 3q24-29 and 6p25.3-p12.2 (37.5% of the cases), 1q21.1-q44, 5q11.1-q22.1, 11p13-p11.2, 13q12.11-q14.3, 17q24.2-q25.3 and Xp22.33-p11.21 (31.3% of the case) and 10q23.31-q26.3, 12p13.33-p13.1, 14q21.1-q32.32, 15q26.1-q26.3, 16q11.2-q22.1 and 19p13.11-p12 (25% of the cases) (Table 1). Sixty-percent of these cytobands were affected by gains of copy number and 40% by losses. The identification of the genes mapped in these cytobands, per the aberration interval base reports, revealed a total number of 4,585 genes (ranging from 51 to 989 genes per cytoband). In the non-TNBC cases, 87.5% of the cases (14/16) showed significant CNAs, and the most frequent cytobands affected were 1q21.11-q44 (64.3% of the cases), 8q21.13-q24.3 (50.0% of the cases), 7q11.21-q36.3 (42.9% of the cases), followed by 6p22.3-p21.2, 17q21.32-q25.3 and 20q13.31-q13.33 (35.0% of the cases) and 19p13.11 (28.57% of the cases) (Table 1). In these cases, only gains of copy number were observed. A total of 3,057 genes (ranging from 324 to 1029 genes per cytoband) were found located in these affected cytobands. The combined array-CGH profiling of the TNBC and non-TNBC cases are presented in Supplementary Figure 1.

### Global miRNA expression profiling

Global miRNA expression profiling was performed using the Nanostring technology in 19 and 24 cases of the TNBC and non-TNBC subtypes, respectively. A number of 163 miRNAs were identified with significantly differential expression between these two groups ( $p \leq 0.05$ ,  $FDR \leq 0.25$ ) (Supplementary Table 1). Unsupervised (UHC) and supervised hierarchical clustering (SHC) analysis of these miRNAs showed a more concise cluster of the TNBC cases, while most of the non-TNBC were interspersed (Figure 1). Of the 163 miRNAs, 87 (53.4%) showed increased expression and 76 (46.6%) showed decreased expression in TNBC when compared to the non-TNBC cases. The miRNAs that presented with the highest changes ( $\log_2FC > 2$ ) in expression between these groups were: miR-187-3p, miR-601, miR-663a, miR-421, miR-378b and miR-1305. These six miRNAs were up-regulated in the TNBC group. In addition, 10 miRNAs (miR-720, miR-1260a, miR-4286, miR-4454, miR-200c-3p, let7b-5p, miR-199a-3p, miR-199b-3p, let-7c and let-7a-5p) were down-regulated in the TNBC group (Supplementary Table 2). To identify the function of each of the 163 identified differentially expressed miRNAs, DIANA miRPath analysis was used to perform pathway enrichment analysis (KEGG pathways). Among the top 10 pathways identified, based on the most significant

adjusted *P* value (FDR corrected), were the ones related to ECM-receptor interaction, adherens junction, mucin type-O-glycan biosynthesis, morphine addiction and proteoglycans in cancer (Supplementary Table 3).

### Integration of miRNA and copy number alterations (CNAs) analysis

#### Mapping of miRNAs in cytobands affected by CNAs

In order to identify miRNAs that could be affected by CNAs in the TNBC cases analyzed, the genomic location of the initial set of 163 miRNAs found to be differentially expressed between the TNBC and non-TNBC subtypes was verified. Forty-five of them (27.6%) were located in the cytobands mostly affected by CNAs in the same TNBC cases profiled by array-CGH as described above. From these 45 miRNAs, 17 (37.8%) presented expression levels in concordance with the observed CNAs at their respective genome *locus* (i. e. cytoband with gains/amplifications of copy number/up-regulated miRNA expression and/or losses/deletions of copy number/down-regulated miRNA expression) (Table 2, Figure 2). The analysis of each individual case per subtype, showed that in the TNBC group, 3 to 12 of the selected 17 miRNAs were observed with altered expression levels, with an average of  $6.32 \pm 0.525$  miRNAs with alterations per case. In the non-TNBC group, 0 to 10 of these miRNAs presented expression alterations, with an average of  $3.58 \pm 0.51$  miRNAs with alterations per case. This difference was statistically significant ( $p = 0.0006$ ), showing that in each of the TNBC cases there was a higher number of miRNA expression alteration levels of the selected 17 miRNA panel when compared to each of the non-TNBC cases. Unsupervised hierarchical clustering (UHC) analysis using expression levels of these miRNAs was able to distinctly cluster all the TNBC cases and most (exception of 4 cases) of the non-TNBC cases (Figure 3).

#### Identification of miRNA target genes with concomitant miRNA expression and CNAs

In addition to the mapping of the miRNAs in the cytobands affected by CNAs as described above, we then searched for the miRNA gene targets that could be potentially affected by the regulation of the 17 miRNAs mapped in these regions. Using miRNA target prediction databases, a total of 10,675 targets were identified, as predicted from more than two independent databases. MiR-608, mapped at 10q24.31, was the miRNA with the highest number of gene targets (5,268 genes) and miR-129-2-3p, mapped at 11p11.2, with the lowest (131 genes). Next, these miRNA targets were matched with the genes that were mapped in the cytobands affected by CNAs, as generated by the aberration interval base report (array-CGH gene list) (4,585 genes). The integration of these data revealed 2,098 common genes (Supplementary

**Table 1: Most common cytobands affected by CNAs and corresponding number of genes located in these regions observed in the TNBC and non-TNBC groups of patients analyzed (presented by chromosome numerical order)**

TNBC						
Chr	Cytoband	Start	Stop	CNA	Cases (%)	# Genes*
chr1	q21.1 - q44	144988715	247737874	gain	5 (31.25%)	989
chr3	q24 - q29	148802495	194903188	gain	6 (37.50%)	242
chr5	q11.1 - q22.1	49690172	111370979	loss	5 (31.25%)	245
chr6	p25.3 - p12.2	2117686	52103799	gain	6 (37.50%)	609
chr8	q21.3 - q24.3	88884192	146066584	gain	7 (43.75%)	289
chr10	q23.31 - q26.3	89507004	135372492	loss	4 (25.00%)	360
chr11	p13 - p11.2	34322106	46565735	gain	5 (31.25%)	51
chr12	p13.33 - p13.1	309062	14132896	gain	4 (25.00%)	205
chr13	q12.11 - q14.3	19703703	53876286	loss	5 (31.25%)	216
chr14	q21.1 - q32.32	38723471	103447263	loss	4 (25.00%)	485
chr15	q26.1 - q26.3	90276459	102241406	gain	4 (25.00%)	63
chr16	q11.2 - q22.1	46693731	67933130	loss	4 (25.00%)	169
chr17	q24.2 - q25.3	65989022	80993001	gain	5 (31.25%)	232
chr19	p13.11 - p12	18266482	21108358	gain	4 (25.00%)	66
chrX	p22.33 - p11.21	1314894	58051765	loss	5 (31.25%)	364
Non-TNBC						
Chr	Cytoband	Start	Stop	CNA	Cases (%)	# Genes*
chr1	q21.1 - q44	145103876	249118400	gain	9 (64.28%)	1029
chr6	p22.3 - p21.1	18093033	43409896	gain	5 (35.71%)	458
chr7	q11.21 - q36.3	62516153	158909738	gain	6 (42.86%)	687
chr8	q21.13 - q24.3	82193925	146280020	gain	7 (50.00%)	324
chr17	q21.32 - q25.3	46048958	81029941	gain	5 (35.71%)	445
chr19	p13.11	17845278	17927374	gain	4 (28.57%)	4
chr20	q13.31 - q13.33	55212094	62893189	gain	5 (35.71%)	114

\*cytobands locations, positions, size and # genes and miRNAs affected by CNAs based on the aberration interval base reports (Agilent CytoGenomics v. 5.0)

Figure 2), i. e. genes that could be potentially affected by both CNAs and miRNA expression deregulation. This integration approach reduced the initial number of the total miRNA targets by 80.3% (from 10,675 to 2,098).

#### Biological function and pathway enrichment analysis of the miRNAs and corresponding gene targets affected by miRNA expression and CNAs

In order to identify the main biological function of the 17 selected miRNAs, we performed pathway enrichment analysis (KEGG pathways) using Diana miRPath v.3.0. Two major miRNAs clusters among the 17 miRNAs regulating these pathways were found: one formed by four miRNAs (miR-539, miR-548, miR-607 and miR-944) and the other by thirteen miRNAs (miR-129-2-3p, miR-135b-5p, miR-188-5p, miR-323a-5p, miR-323b-3p, miR-323b-5p, miR-342-3p, miR-378c,

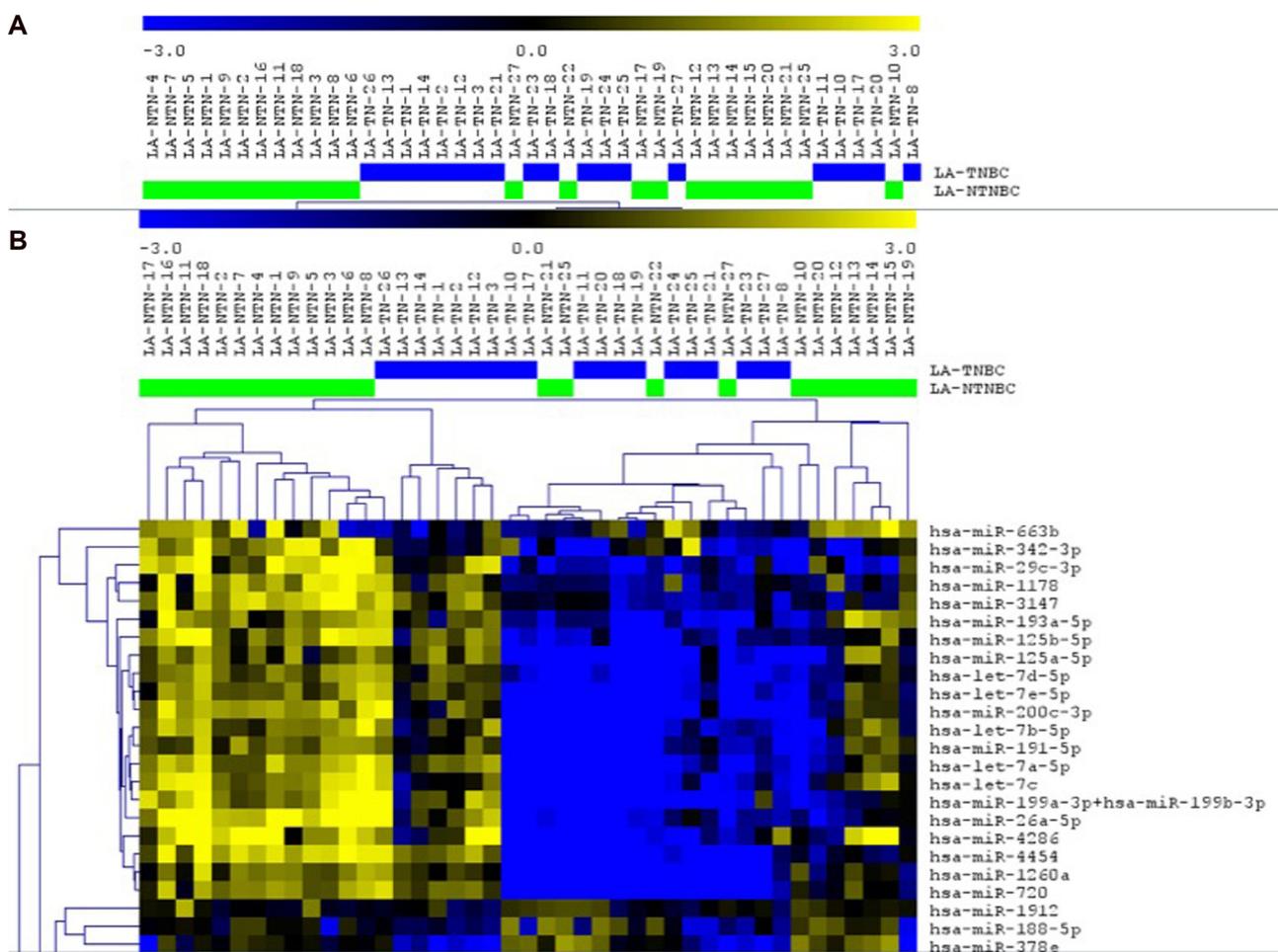
miR-608, miR-634, miR-668-3p, miR-1260a and miR-1275) (Supplementary Figure 3). The top 10 pathways identified involving these miRNAs were the axon guidance (hsa04360), glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate (hsa00532), mucin type-O-Glycan biosynthesis (hsa00512), thyroid hormone signaling pathway (hsa04919), signaling pathways regulating pluripotency of stem cells (hsa04550), proteoglycans in cancer (hsa05205), wnt signaling pathway (hsa04310), hippo signaling pathway (hsa04390), ras signaling pathway (hsa0414) and pathways in cancer (hsa05200). Remarkably, all the 17 selected miRNAs were presented in four of the top 10 pathways identified. The less representative pathways by these miRNAs were the mucin type-O-glycan biosynthesis and the glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate, with the involvement of 29.4% and 41.2% of the miRNAs, respectively (Supplementary Table 4).

Identification of KEGG pathways potentially affected by the 17 miRNAs were also conducted considering only the 2,098 putative target genes presented in the cytobands with CNAs previously selected in the integrated analysis above. As a result, three KEGG pathways were found to be affected by ten out of the 17 miRNA panel: glycosaminoglycan biosynthesis - chondroitin sulfate/dermatan sulfate (hsa00532,  $p = 0.01967$ ), biosynthesis of unsaturated fatty acids (hsa01040,  $p = 0.01967$ ) and hippo signaling pathway (hsa04390,  $p = 0.01967$ ) (Supplementary Table 5). Eight miRNAs of 17 miRNA panel were involved in the hippo signaling pathway: three upregulated (miR-944, miR-135b-5p and miR-1275) and five downregulated (miR-548p, miR-607, miR-323b-3p, miR-342-3p and miR-539-5p). Using Cytoscape and Diana miRpath data (Supplementary Figure 4), we observed that these miRNAs can act by targeting key regulators of this pathway, potentially damaging their control over several other biological functions, e. g. miR-944 targets *CRB1*, *MPP5* and *WWTR1* genes, which are involved in tight and adherens cell junction. Interestingly, three out of the

four downregulated miRNAs (miR-548p, miR-323b-3p and miR-607) were shown to target *CCND2* gene, a cell cycle regulator.

### Individual and combined discriminatory power of the 17 miRNA panel between the TNBC and non-TNBC groups of patients

Each miRNA composing the identified 17 miRNA panel was evaluated in relation to its power in discriminating the TNBC and non-TNBC subtypes of the Latina patients (Figure 4). ROC analysis showed that 70% (12/17 = 70.6%) of the miRNAs presented an Area Under the Curve (AUC) value higher than 0.7. The miR-539-5p, miR-634, miR-323a-5p, miR-944 and miR-1260a miRNAs presented the highest discriminatory power, with AUC values ranging from 0.742 (miR-1260a) to 0.780 (miR-539-5p). The combined analysis of the panel showed a AUC value of 0.953, demonstrating a robust power of the complete 17 miRNA panel in discriminating the TNBC and non-TNBC subtypes of the cases studied.



**Figure 1:** Unsupervised (A) and supervised (B) hierarchical clustering analysis applied to the TNBC (blue bars) and non-TNBC (green bars) cases analyzed. Up-regulated miRNAs (yellow) and down-regulated miRNAs (blue). Selected area of the heatmap showed. (MeV 4.9, Pearson Correlation,  $p < 0.05$ ).

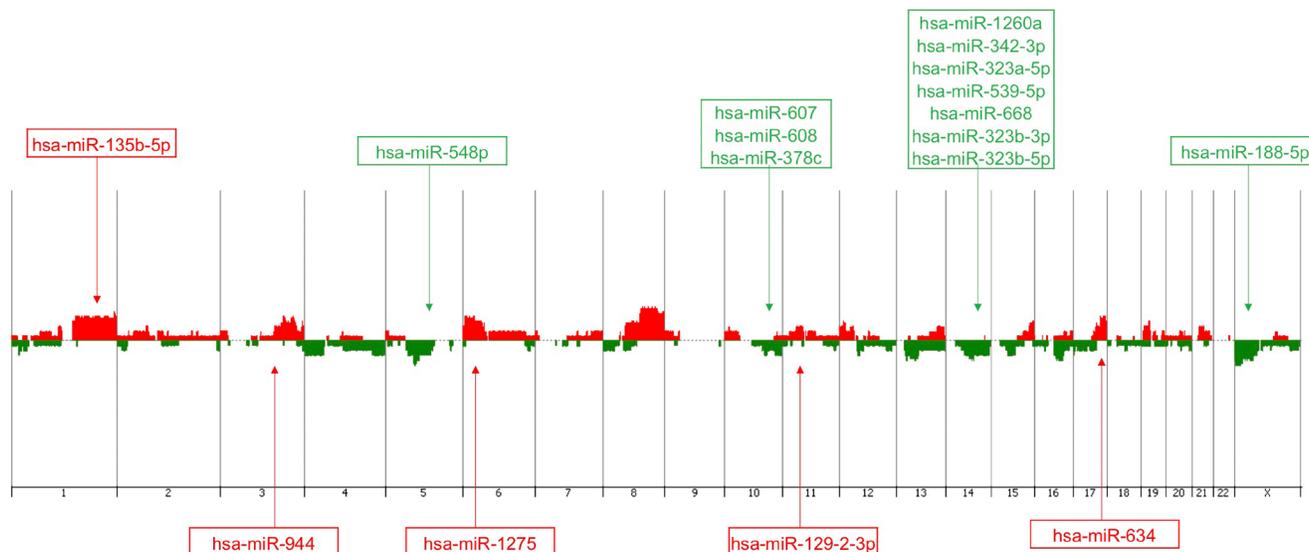
**Table 2: Chromosome location of the seventeen differentially expressed miRNAs between the TNBC and non-TNBC groups of patients, with expression levels in concordance with copy number alterations (CNAs) (presented by chromosome numerical order)**

miRNAs	Cytoband	Start	Stop	CNA	miRNA expression	Log2FC	Adj p
hsa-miR-135b-5p	1q32.1	205448302	205448398	gain	up-regulated	1.93	0.001
hsa-miR-944	3q28	189829922	189830009	gain	up-regulated	1.13	0.004
hsa-miR-548p	5q21.1	100816482	100816565	loss	down-regulated	-0.53	0.002
hsa-miR-1275	6p21.31	33999972	34000051	gain	up-regulated	0.61	0.049
hsa-miR-607	10q24.1	96828669	96828764	loss	down-regulated	-0.50	0.006
hsa-miR-608	10q24.31	100974985	100975084	loss	down-regulated	-0.52	0.04
hsa-miR-378c	10q26.3	130962588	130962668	loss	down-regulated	-0.42	0.024
hsa-miR-129-2-3p	11p11.2	43581394	43581483	gain	up-regulated	0.77	0.036
hsa-miR-1260a	14q24.3	77266218	77266290	loss	down-regulated	-3.30	0.004
hsa-miR-342-3p	14q32.2	100109655	100109753	loss	down-regulated	-1.51	0.039
hsa-miR-323a-5p	14q32.31	101025732	101025817	loss	down-regulated	-0.59	0.001
hsa-miR-539-5p	14q32.31	101047321	101047398	loss	down-regulated	-0.68	<0.05
hsa-miR-668	14q32.31	101055258	101055323	loss	down-regulated	-0.45	0.014
hsa-miR-323b-3p	14q32.31	101056219	101056300	loss	down-regulated	-0.49	0.008
hsa-miR-323b-5p	14q32.31	101056219	101056300	loss	down-regulated	-0.49	0.009
hsa-miR-634	17q24.2	66787072	66787168	gain	up-regulated	0.90	0.005
hsa-miR-188-5p	Xp11.23	50003503	50003588	loss	down-regulated	-0.93	0.02

**Association of miRNA expression with the clinical-pathological variables of the patients studied**

The association of the miRNA expression data with clinical-pathological information of the patients

was performed in 43 samples (19 TNBC and 24 non-TNBC) (Table 3). Given the retrospective nature of this study, there were substantial missing clinical data from 13 individuals, who were therefore not included in this analysis. The age and tumor size data were log-transformed in order to be more normal-like before



**Figure 2: Genomic view/penetrance plot of the array-CGH profiling of the TNBC cases ( $n = 25$ ) from the patients analyzed with the corresponding mapping of the 17 miRNAs of the identified panel. Vertical lines represent chromosome number. Red peaks indicate copy number gains and green peaks indicate copy number losses. The miRNAs expression levels, up- and down regulated, are represented in red and green color boxes, respectively (Agilent Genomic Workbench 7.0).**

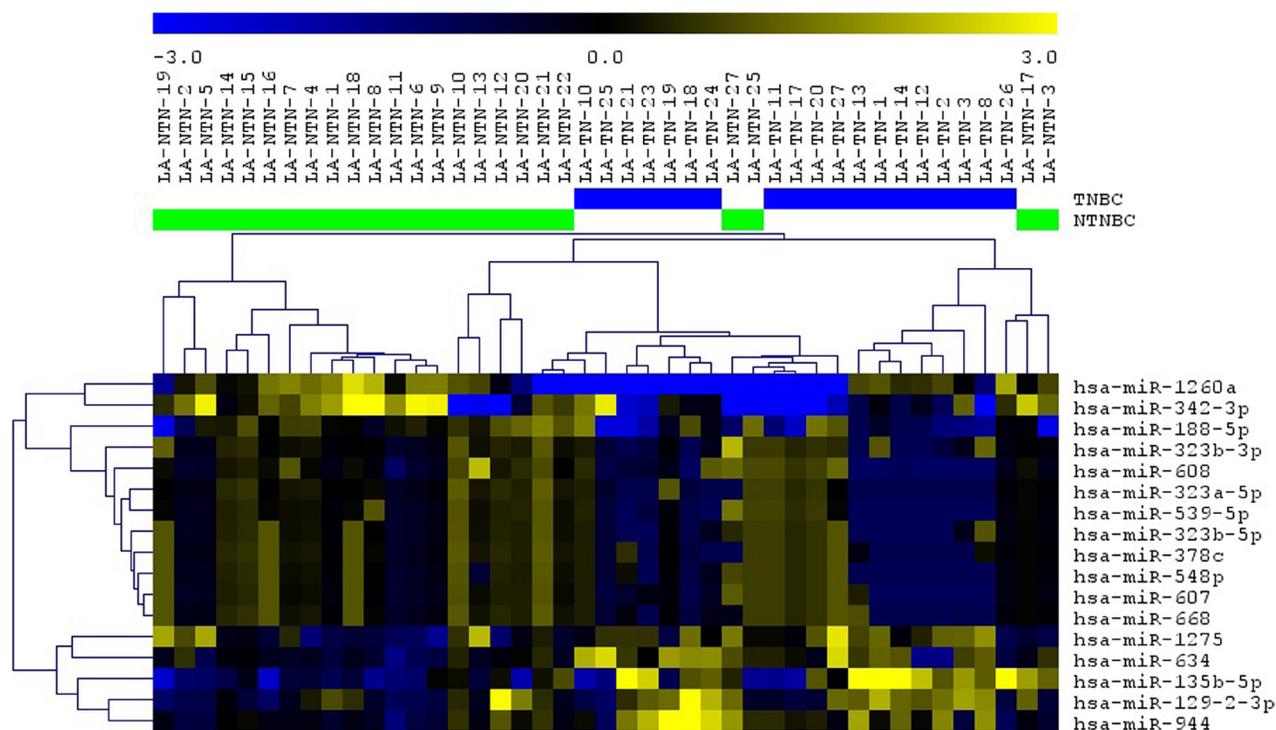
applying the Student *t* test; neither variable was associated with TNBC status ( $p = 0.46$  for age,  $p = 0.34$  for tumor size). The categorical variables were assessed by Fisher's exact test and the following results were obtained: histological type and tumor grade were not significantly different between the TNBC and non-TNBC groups ( $p = 0.12$  and  $p = 0.06$ , respectively), while lymph node status and tumor stage were significantly different ( $p = 0.002$  and  $p = 0.006$ , respectively).

Two linear regression models were considered for each of the 17 miRNAs selected for integration, with the log-transformed miRNA values as outcomes and the covariates: age, tumor size, tumor subtype, histological type, lymph node status, and either tumor stage or tumor grade for each model. Both complete case analyses and multiple imputation analyses were considered. Tumor stage and tumor grade were not included in the same model due to their high degree of overlap: of the 43 cases, 2 did not present information for both tumor stage and grade; for the remaining 41 samples, 25 were of tumor grade I or II and moderate or well differentiated tumor stage, while 14 were of tumor grade III and poorly differentiated stage. For the complete-case analyses, which included 30 cases, 15 of the 17 miRNAs presented significant associations with tumor subtype (at a FDR = 0.05) when considering tumor grade in the regression analysis and 14 miRNAs presented significant associations with tumor subtype when considering tumor stage; the non-significant miRNAs in these analysis were miR-1275

and miR-129-2-3p and miR-1275, miR-129-2-3p, and miR-378c, respectively. None of the other variables were significantly associated with the 17 miRNAs expression values. The multiple imputation results were very similar - 16 out of the 17 miRNAs had significant associations with tumor subtype (at a FDR = 0.05 using Benjamini-Hochberg) when considering either tumor grade or tumor stage in the regression analysis. In both cases, the non-significant miRNA was miR-1275 and no other variables presented significant associations with miRNA values. Furthermore, the direction of the associations always remained the same as in the univariate analyses between miRNA values and tumor type.

### Association of miRNA expression with survival using KMPlot database

The miRNA expression levels of eight out of the 17 miRNAs in our panel were previously associated with TNBC survival in the analysis of the KMPlot datasets from the TCGA and METABRIC cohorts of basal-like/TNBC patients (Supplementary Table 6). In these TNBC cohorts, higher expression levels of miR-135b-5p and miR-634, as observed in our study, were significantly associated with reduced overall survival (OS) (months) (Figure 5A). Lower expression levels of miR-323b-3p, miR-548p, miR-607, miR-608, miR-668, and miR-1260a, was also observed in our study, were also associated with lower survival (Figure 5B).



**Figure 3: Unsupervised hierarchical clustering (UHS) analysis applied to the TNBC (blue bars) and non-TNBC (green bars) cases analyzed using the selected 17 miRNAs.** Up-regulated and down-regulated miRNAs are represented in yellow and blue color, respectively (MeV 4.9, Pearson Correlation).

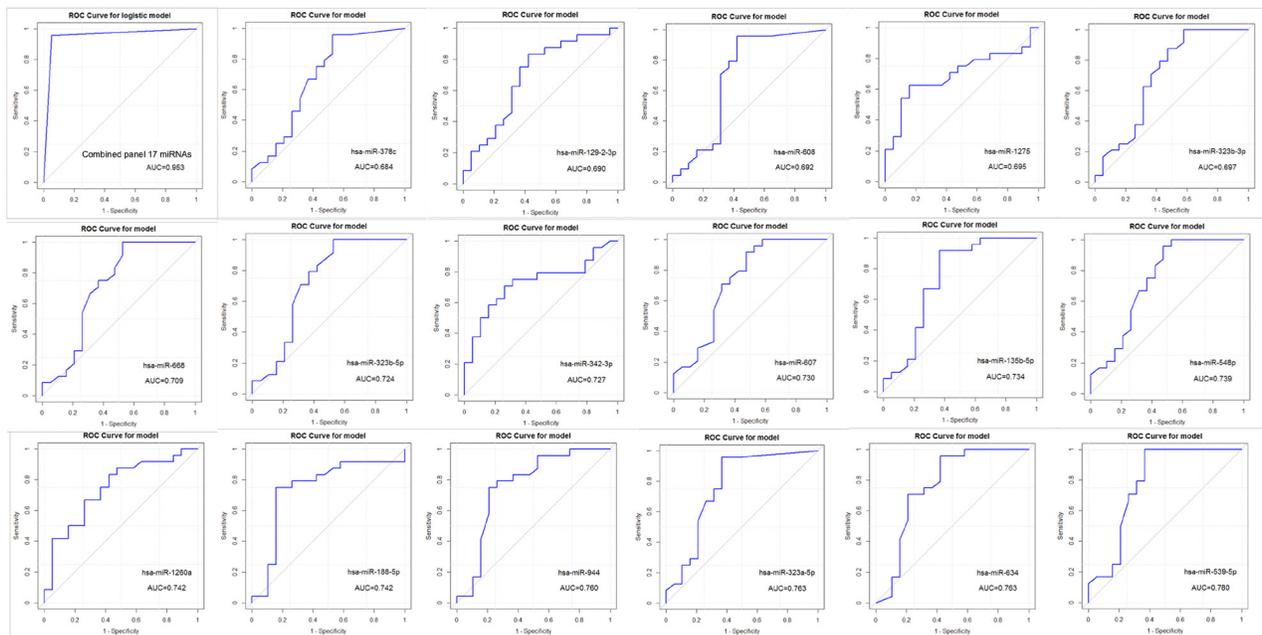
**Table 3: Clinical and histopathological information from TNBC and non-TNBC groups**

	TNBC (n = 27)	non-TNBC (n = 27)	P value
Age	54.25 ± 3.59	58.15 ± 2.34	
	16–83 (n = 20)	34–88 (n = 27)	p = 0.46
Tumor size (cm)	3.10 ± 0.47	3.48 ± 0.39	
	0.9–10.5 (n = 23)	0.7–8.0 (n = 26)	p = 0.34
Grade			
I/II	12	21	
III	11	5	p = 0.06
Tumor Stage			
Well differentiated	2	2	
Moderately differentiated	11	20	
Poorly differentiated	12	4	p = 0.002
Lymph node			
Positive	10	21	
Negative	10	3	p = 0.006

**DISCUSSION**

Population-based studies in North American patients, with breast cancer subtypes classified mostly by Immunohistochemistry (IHC) surrogate markers, have reported that Hispanic/Latina women are more likely to present with estrogen receptor negative (ER-) tumors,

compared to non-Hispanic White (NHW) women [6, 26]. This data is similar to what has been reported among the distribution of the breast cancer subtypes in African American (AA) and NHW women [9, 56, 57]. Triple negative breast cancer (TNBC), an aggressive disease that rapidly progresses [58], has been shown to be more prevalent in Latina women (both living in US and in their



**Figure 4: Receiver operating characteristic (ROC) curve plots and Area Under the Curve (AUC) values of the combined (top left) and individual 17 miRNAs differentially expressed between the TNBC and non-TNBC groups of patients.**

countries of origin) (from 15 to 35%) compared to NHW women (from 8 to 15%) [24, 59]. However, the prevalence of these tumors greatly vary in the diverse Latin American countries, including Brazil [24]. In Brazil, the frequency of TNBC is about 17%, similar to the frequencies in Costa Rica and lower than the frequencies in Mexico, Peru and Colombia (from 20 to 35%) [59–63]. These frequency variations can be related to the differences in the subtype classification methods used, biases associated with clinic/hospital and/or registry based studies, lifestyle and environmental exposures and socio-economic factors [64, 65].

Increasing evidence has demonstrated that biological factors can account not only for the prevalence but for the higher mortality rate of TNBC that is observed in minority populations [21, 25, 50, 66–68]. In AA patients, several studies [21, 66], including our own [50], have shown differences in the tumor molecular profiles, both at the DNA copy number and miRNA expression level, when compared to the NHW patients. It has been shown, that in fact, gene expression profiles might change according to the genetic ancestry of an individual's genome [67]. Characterization of the molecular profiles in each breast cancer subtype has been extensively performed, however, only a small number of studies have characterized them in specific ethnic groups, such as Latinas [21, 25, 68]. Considering the clinical implication of the molecularly defined subtypes, this lack of biological knowledge can directly impact the prognosis and treatment of this and other underrepresented populations.

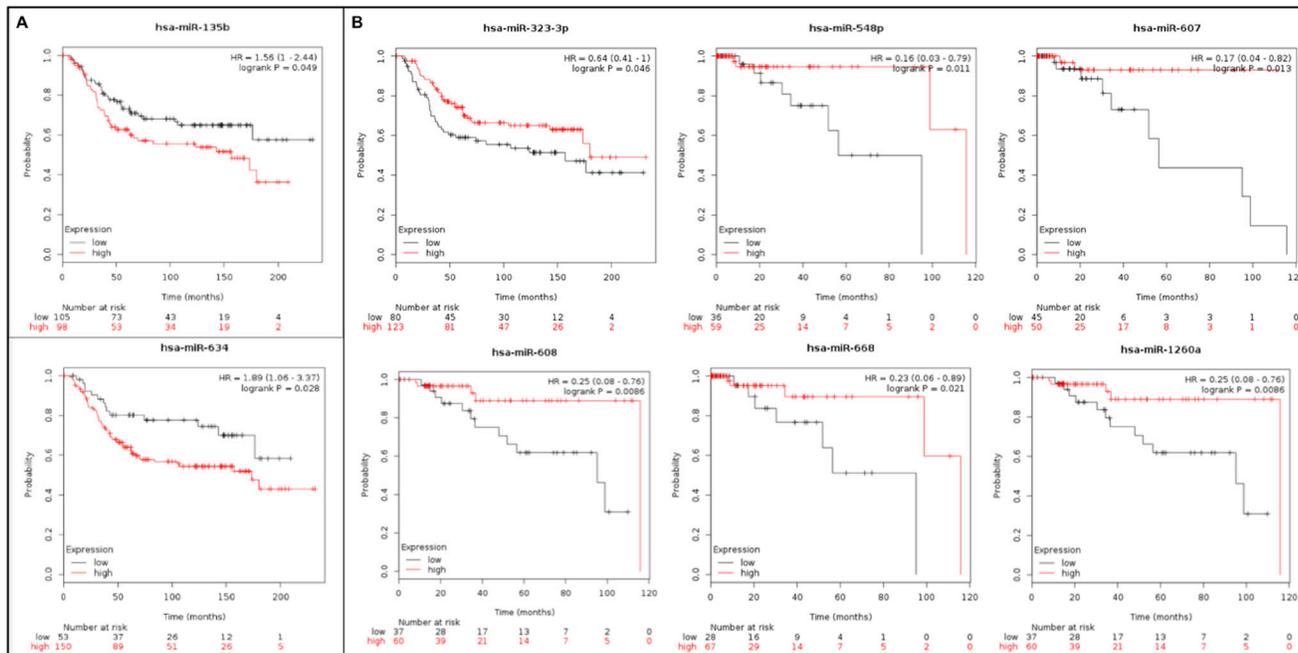
MiRNA expression patterns have been shown, like gene expression patterns, to discriminate between

the major breast cancer subtypes, in both cell lines and clinical studies [69, 70]. Race and ethnicity has also been shown to play a role in miRNA expression patterns; most of the data is however on germline miRNA alterations and their associated risk of breast cancer [71], rather than on somatic tumor cells.

In this study, using an integrated analysis of genome-wide copy number and miRNA profiling we identified a panel of 17 miRNAs differentially expressed between the TNBC and non-TNBC groups of patients that live in Latin America, specifically in the South of Brazil. Receiver Operating Characteristic (ROC) curve analysis showed a remarkable high power of this miRNA panel in discriminating the TNBC and non-TNBC subtypes of these groups.

In TNBC, five of these miRNAs were observed to be up-regulated (miR-135b-5p, miR-944, miR-1275, miR-129-2-3p and miR-634) and 12 down-regulated (miR-548p, miR-607, miR-608, miR-378c, miR-1260a, miR-342-3p, miR-323a-5p, miR-539-5p, miR-668, miR-323b-3p, miR-323b-5p and miR-188-5p) when compared to the non-TNBC group of patients. These miRNAs were selected among the significantly differentially expressed miRNAs among these subtypes and based on their localization in genomic regions affected by copy number alterations (gains/amplifications and/or losses/ deletions), as detected in the same TNBC specimens by array-CGH analysis. These regions involved mainly the 1q, 3q, 5q, 6p, 10q, 11p, 14q, 17q and Xp cytobands, some of which recurrently described as altered in TNBC [72].

All five observed up-regulated miRNAs were previously reported with altered expression in breast



**Figure 5:** Kaplan–Meier plot results of eight out of the panel of 17 miRNAs that were differentially expressed between the TNBC and non-TNBC dataset of the METABRIC and TCGA data (KMPlot - miRpower). (A) Up-regulated miRNAs (B) Down-regulated miRNAs.

cancer studies. MiR-135b was found upregulated both in breast cancer cell lines and clinical cases, and shown to confer higher proliferation, migration and invasion activity in MCF-7, MDA-MB-231 and MDA-MB-468 cells [73–75]. These activities may be the result of miR-135b regulation of Wnt pathway by targeting the *APC* gene in MDA-MB-231 and MDA-MB-468 cell lines [74]. A similar oncogenic activity, in addition to promoting resistance to cisplatin, was shown for miR-944 in the MCF7 and MDA-MB-231 cells [76]. Higher expression of miR-135b-5p was also associated with ER-negative breast cancer [77, 78] and found to regulate ER $\alpha$  protein levels by interacting with its 3'UTR regions [75]. Interestingly, miR-135b-5p was found downregulated in metastatic microinvasive breast carcinoma samples when compared to non-metastatic microinvasive breast carcinoma samples and its suppression was shown to increase proliferation, migration and invasion capacity by targeting syndecan binding protein (*SDCBP*) gene [79]. These might suggest that miR-135b-5p may play different roles depending on tumor stage by targeting different set of genes: in initial stages, high levels of miR-135b-5p may be important for promoting the formation of tumor while in late stages its lower expression can induce invasion and metastatic activity. Differential expression of miR-135b may also contribute to TNBC molecular heterogeneity as higher levels of this miRNA were found in basal-like cases (EGFR and CK5/6 positivity) when compared to non-basal like cases. Expression levels of miR-135b were also significantly associate with Ki67 ( $\beta = 0.94, p < 0.05$ ) and AR expression ( $\beta = -25.9, p < 0.05$ ): overexpression of miR-135b was positively correlated with high Ki67 expression ( $\rho = 0.434, p < 0.05$ ) and low levels of miR-135b showed a negative correlation with AR expression ( $\rho = -0.276, p < 0.05$ ) [80]. Other studies have also reported higher levels of miR-135b in basal-like tumor subtypes [81, 82] and suggests a correlation between its overexpression and poor survival and early metastasis relapse [83]. Upregulation of miR-1275 levels was particularly found in breast cancer of young women (<35 years) when compared to older women [84] and was also found hypomethylated in samples of healthy individuals who developed breast cancer when compared to individuals who remained healthy [85]. Although this combined evidence warrants further validation, it suggests that this miRNA might play a role in the initiation of breast cancer and could be used as a predictive cancer biomarker. MiR-129-2, that targets the Progesterone Receptor (PR) gene, has been described as upregulated in patients with low PR expression levels (PR-) [86]. This target suppression, is compatible with our findings, showing the up-regulation of miR-129-2 in the TNBC cases when compared to the non-TNBC cases. Finally, miR-634 was shown to regulate HER2 signaling by inducing apoptosis and inhibiting levels of HER2, p-AKT and p-ERK [87]. As TNBC have no expression of HER2, miR-634 may present a different role in these tumors by targeting other genes.

Among the 12 down-regulated miRNAs in our TNBC cases, miR-548p, miR-539-5p, miR-342-3p and miR-668 were previously reported with altered expression in breast cancer, with possible function as tumor suppressors [88–95]. The miR-548p's anti-oncogenic activity was observed in two different studies in breast cancer; in the study of Shi *et al* (2015) its up-regulation was found to inhibit cell proliferation and induce apoptosis by targeting Enoyl Coenzyme A Hydratase short chain 1 (*ECHS1*) [88] and in the study of Ke *et al* (2016), it was found to perform the same functions by targeting the Nuclear Paraspeckle Assembly Transcript (*NEAT*) gene [89]. MiR-539-5p was also found to interfere with cell proliferation activity, in addition to migration, by regulating EGFR expression in MCF-7 and MDA-MB-231 cell lines [90] and lamin subunit alpha 4 (*LAMA4*) gene in BT549 cell line [91]. Low levels of miR-342-3p was reported in TNBC and is significantly associated with poor prognosis. Its tumor suppressor activity was already shown in several TNBC cell lines (BT549, SUM149, SUM159, MDA-MB-157 and MDA-MB-468) as miR-342-3p can suppress cell growth, viability and migration activity [92]. Expression of miR-342-3p is positively correlated with ER $\alpha$  [93] and its downexpression has been associated with tamoxifen resistant breast tumors [93, 94]. In ER-positive cases, high expression of miR-342 has been associated with better survival but not in ER-negative or TNBC cases, highlighting its role in tamoxifen response [95].

Expression levels of miR-1206a were already inversely correlated with HER-2 expression levels in MCF-7 and BT474 cell lines [96]. Up-regulation of miR-1206a was observed in HER-2 overexpressed MCF-7 cells when compared to regular MCF-7 cells, and downregulation of miR-1206a was observed in HER-2 intervened BT474 cells when compared to regular BT474 cells. These results corroborates our findings that miR-1260a was found down-regulated in TNBC cases. Interestingly, altered expression of miR-668 and miR-1206a may present an importance for non-TNBC patients as their overexpression presented clinical relevance in different studies. Overexpression of miR-668 was found to confer resistance to previously radiosensitive MCF-7 and T-47D cells by targeting *Ikb $\alpha$*  [97], and high levels of circulating miR-1260a in serum samples of metastatic breast cancer patients was related to poor prognosis [98]. Overexpression of miR-1260a was also found to be associated with poor prognosis in studies with different tumor types: melanome [99], neuroblastoma [100], prostate [101]. These findings suggest that miR-668 and miR-1260a present potential as prognostic biomarkers for non-TNBC patients.

Identification of the main pathways and networks potentially affected by the 17 miRNAs and putative gene targets resulted in a total of 46 KEGG pathways, among them Axon guidance, Glycosamin biosynthesis – chondroitin sulfate/ dermatan sulfate and Mucin type O-Glycan biosynthesis. Interestingly, all of the 17

miRNAs of the identified panel were found to regulate gene targets associated with cancer signaling pathways, such as the Wnt, Ras, ErbB and Rap1. It is of note that 15 out of the 17 of the miRNAs were found to affect the Hippo signaling pathway.

A second integration analysis of the genes that were potentially affected by the CNAs (identified by the copy number profiling) and the corresponding gene targets of the 17 miRNAs located at the same affected cytobands, revealed a total of 2,098 genes mapped in these regions, suggesting that they may be commonly affected by both of these mechanisms. Among them were included genes that are critical to the TNBC tumorigenesis (and miRNA biogenesis), such as *CDKN1A*, *DICER1*, *ETV6*, *IGF1R*, *MYC* and *PIK3CA*. Enrichment functional analysis of these genes revealed three signaling pathways preferentially involved including the Hippo signaling pathway, observed in the first data integration approach presented above. Eight out of the 17 miRNAs (miR-1275, miR-135b-5p, miR-323b-3p, miR-342-3p, miR-5395p, miR-548p, miR-607 and miR-944) were found related to this pathway, potentially targeting 18 genes: *DVL3*, *PPP2RD2*, *LATS2*, *TCF7L2*, *CCND2*, *FZD6*, *WNT8B*, *BTRC*, *CSNK1D*, *FRMD6*, *STK3*, *YWHAZ*, *MPP5*, *SAV1*, *SOX2*, *CRB1*, *GDF6* and *WWTR1*. A search on miRTarBase database however, showed that few of these interactions were experimentally validated [102]. Considering the critical role of the Hippo signaling pathway in regulating cell proliferation and apoptosis, and other tumorigenic processes, it is relevant to pursue downstream functional studies to confirm these interactions and determine their biological significance to the TNBC phenotype.

Finally, the association of the expression of the 17 miRNA panel with the clinical-histopathological parameters from the patients showed, except for three miRNAs, miR-1275, miR-129-2-3p, and miR-378c, association with tumor grade and tumor stage. Interestingly, eight out of the 17 miRNAs were previously associated with TNBC survival in the analysis of the KMPlot datasets from the TCGA and METABRIC cohorts of basal-like/TNBC patients. In particular, the up-regulation of miR-135b-5p and miR-634 and down regulation of miR-323b-3p, miR-548p, miR-607, miR-608, miR-669 and miR-1260a levels, were associated with reduced overall survival.

In conclusion, the integrated analysis of DNA copy number alterations and miRNAs expression levels, performed in this study, led to the identification of a robust 17 miRNA panel, with a high power in discriminating between the TNBC and non-TNBC subtypes of Latina patients. The clinical validation of this panel in a novel and independent ancestrally characterized Latina population, can reveal whether this panel, or a subset of its composing miRNAs, can represent the intrinsic biology of their TNBC transcriptomes, that can differentially impact their tumor phenotypes and clinical behavior.

## MATERIALS AND METHODS

### General study design

Genome-wide array-CGH and miRNA profiling were performed in TNBC and non-TNBC subtypes of patients from Latin America, Brazil, to detect the patterns of DNA copy number alterations (CNAs) and changes in miRNA expression, respectively. The differentially expressed miRNAs between the subtypes were integrated with copy number profiling data performed in the same TNBC tissue specimens, to select miRNAs that were mapped in regions affected by CNAs and gene targets potentially affected by both miRNA deregulation and CNAs. Combinatorial target prediction algorithms in conjunction with functional and pathway annotation enrichment systems were then applied to the selected miRNAs to identify the most relevant miRNAs and their corresponding targets associated with TNBC. Receiver Operating Characteristic (ROC) curve analysis was performed to determine the individual and combined power of the differentially expressed miRNAs in discriminating the TNBC and non-TNBC subtypes of the patients. Finally, the molecular data was associated with clinical-pathological information from the patients and external survival data to determine their potential prognostic relevance.

All experiments of this study were performed in accordance with relevant guidelines and regulations.

### Patient accrual and sample collection

Fifty-four formalin-fixed paraffin-embedded (FFPE) tissue samples of non-treated primary breast tumors were collected from the pathology tumor bank at the Hospital Nossa Senhora das Graças (HNSG), Paraná, Brazil. All samples were transferred to Georgetown University with no patient identifiers, under patient informed consent and through the IRB approval of Georgetown University, HNSG and the National Review Board of Ethics in Research (CONEP-Brazil). The TNBC and non-TNBC subtypes were determined by ER, PR and HER2 receptors status by immunohistochemistry (IHC) analysis performed at the time of diagnosis, following international guidelines [103, 104]. Briefly, the Monoclonal Mouse Anti-Human Estrogen Receptor  $\alpha$  and Polyclonal Rabbit Anti-Human Progesterone Receptor were used for ER and PR analysis, respectively. ER and PR positivity were considered using a cut-off of 1%. The HercepTest (Dako North America Inc, Carpinteria, CA, USA) was used for HER2/Neu+ status. Using these criteria, 27 patients presented tumors of the TNBC subtype and 27 of the non-TNBC subtype (10 ER+/PR+/HER2-, 3 ER+/PR+/HER2+, 3 ER-/PR+/HER2-, 2 ER+/PR-/HER2- and 1 ER+/PR-/HER2+). In the remaining eight non-TNBC cases, 6 did not present information for PR expression (5 were ER+/HER2- and

1 ER+/HER2+), one for HER2 status (ER+/PR+) and one for ER and PR status (HER2+).

The clinical and histopathological information from the patients was retrieved in a de-codified manner from the pathology reports, and included age, tumor size, tumor stage and tumor grade and lymph node status (Table 3).

### DNA and RNA isolation

Prior to DNA and RNA isolation, the FFPE specimens were evaluated by the pathologist for the presence of at least 80% of tumor cells. The selected tumor areas were microdissected from unstained 10 $\mu$ m FFPE tissue sections and used for the subsequent molecular analysis. Consecutive tissue sections from the same tissue blocks were used to isolate DNA and RNA ensuring a direct correlation of DNA copy number and miRNA expression profiling data, as previously performed by our group [50].

DNA isolation was performed using phenol-chloroform protocol optimized for FFPE material [105] and RNA isolation was performed using MasterPure™ Complete DNA and RNA Purification kit (Epicentre Biotechnologies) following manufacture protocol. DNA and RNA quantity and quality were assessed using NanoDrop™ Spectrophotometer (Thermo Scientific Inc.) and the Bioanalyzer (Agilent Technologies Inc.), respectively.

### Ancestral analysis

The Latina population is highly heterogeneous and comprises individuals of several different genetic ancestries [106–109]. To obtain genomic based information of the population studied in relation to ethnicity, a subset of the patients (15 patients) was genotyped using the SNP chip Illumina Infinium QC Array (Illumina Inc., CA, USA), which contains 15,949 markers, including approximately 3,000 ancestral informative markers (AIMs). The genotype calling was performed as we previously described [50], using the GenomeStudio Software v. 2011.1. SNPs with MAF  $\leq$  0.01 were excluded from analysis. The data obtained was subsequently merged with the 1000 Genomes Project phase 1 ( $n = 1,902$  samples) dataset [67], which present an overlap of 14,718 variants with the one from our study. Finally, Principal Components Analysis (PCA) was performed using PLINK 1.9 [110], which uses the EIGENSTRAT method [111] to calculate model ancestry differences between different samples. Based on the results of PC1 and PC2 (Supplementary Figure 5) the patients of this study clustered with the European (EUR) defined group from the 1000 Genome Project as well as with the Admixed Americans (AMR) main group, mainly composed of Colombians and Mexicans. This data was not surprising considering the markedly ancestral heterogeneity of the Brazilian population [112, 113].

### Array-CGH analysis

Genome-wide copy number profiling was performed by array-CGH using the SurePrint G3 Human CGH Microarray (Agilent, Santa Clara, CA, USA) according to our previous protocol for FFPE samples [105]. DNA isolated from peripheral blood from multiple normal individuals was used as control (reference) DNA. Control and case samples were directly labeled using the Bioprimer a-CGH Genomic Labeling kit and hybridized to the arrays for 40 hours. The arrays were scanned using Scanner Agilent G2565CA, and the data extracted using Feature Extraction (FE) software v10.10 (Agilent Tech. Inc.). The Agilent Cytogenomics v.3.0 software (Agilent Technologies Inc., Santa Clara, CA, USA), was used to analyze the data, using the algorithm ADM-2, threshold of 6.0 and an aberration filter with a minimum of 3 probes. Gene amplifications and deletions were defined as minimum average absolute log<sub>2</sub> ratio (intensity of the Cy5 dye (reference DNA)/intensity of the Cy3 dye (test DNA) value of  $>0.25$  and  $<-0.25$ , respectively, as per Agilent Cytogenomics guidelines. The number of “calls” (total significant number of CNAs) and the specifically affected cytobands were obtained from the generated aberration interval base reports (Agilent Cytogenomics v.3.0).

### MiRNA expression analysis

Global miRNA expression profiling was performed using NanoString nCounter technology Human v2 miRNA Expression Assay, as we previously described [50]. This specific assay contains 800 endogenous miRNAs, six positive miRNA assay controls, six negative miRNA assay controls, and five housekeeping transcripts (*ACTB*, *B2M*, *GAPDH*, *RPL19*, *RPLP0*). Raw miRNA expression data was pre-processed and normalized using NanoString's nCounter RCC collector and nSolver v2 software respectively. Unsupervised (UHC) and supervised hierarchical cluster (SHC) analysis were performed on significantly differentially expressed miRNAs among the patients' subtypes, using Pearson's correlation coefficient, average linkage and Benjamini-Hochberg multiple testing correction on the Multiexperiment Viewer software (MeV 4.9.0) ( $t$ -test,  $p < 0.05$ , FDR  $\leq 0.25$ ). Fold changes, represented on the log<sub>2</sub> scale (log<sub>2</sub>FC), were calculated for all differentially expressed miRNAs. Adjusted  $p$ -values were used to rank miRNAs of interest.

### Integrated analysis of array-CGH and miRNA data

Direct integration of the most differentially expressed miRNAs associated with the TNBC subtype with CNAs identified in the genome of the same tissue samples was performed using two distinct approaches, as previously described [50]: 1. Mapping of the miRNAs

at the cytobands with high levels of CNAs and further selection based on their concordance level (i. e., cytoband with gains/amplifications of copy number/ up-regulated miRNA expression and/or losses/deletions of copy number/ down-regulated miRNA expression). Only the significant DNA segments affected by CNAs that were present in more than 25% of the cases (to assure that the CNAs were non-random and recurrent and were representative of most of the cases analysed), as identified in the aberration interval base reports (Agilent Cytogenomics v.3.0) were considered in this analysis. The location of each miRNA was determined using miRBase (<http://www.mirbase.org>); 2. Identification of common gene targets of the selected miRNAs above, that may be affected by both CNAs and miRNA expression alterations. In this second approach, for the previously selected miRNAs, gene targets were queried using the available miRNA target databases (Diana micro-T-CDS v.5.0 (diana. imis. athena-innovation/gr/DianaTools/index. phpr = microT\_CDS/index) [114], miRDB (<http://www.mirdb.org/miRDB/>) [115] and TargetScan Release 7.1 ([http://www.targetscan.org/vert\\_71/](http://www.targetscan.org/vert_71/)) [116]; only miRNA target genes that were present in two out of the three miRNA databases were selected.

### Biological function and pathway analysis

In order to assess the potential impact of the deregulated miRNAs identified above in cancer associated biological processes and pathways, Diana miRPath v.3.0 was used (<http://snf-515788.vm.okeanos.grnet.gr/>) based on adjusted *p*-values (FDR correction) [117]. Enrichment analysis of multiple miRNA gene targets comparing each set of miRNA targets to all known KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways was obtained and selected by significant *p*-value ( $p < 0.05$ ) and cancer-associated biological functions. CyKEGG Parser [118], GENEMANIA [119] and CyTargetLinker [120] (Cytoscape 3.5.1 software Applications [121]) were used to build KEGG pathways including miRNAs and their respective gene targets.

### Receiver operating characteristic (ROC) curve analysis

Receiver operating characteristic (ROC) curves and their area under the curve (AUC), was used to identify the power of the selected miRNAs in discriminating the TNBC and non-TNBC subtypes of the Latina patients. Sensitivity was plotted against 1-specificity for the binary classifier (TNBC and non-TNBC). An AUC of 100% denotes perfect discrimination by the miRNA, whereas an AUC of 50% denotes complete lack of discrimination by the miRNA. AUCs and 95% corresponding confidence intervals were calculated for each miRNA and for the combined panel of 17 miRNAs.

### Association analysis of clinical-pathological variables and the array-CGH and miRNA data

The Student *t* test with the unequal variance assumption was used to assess the differences in the mean age at diagnosis of the patients and mean tumor size in the TNBC and non-TNBC groups of patients. Fisher's exact test was used to compare tumor stage, tumor grade, lymph node status and histology between the TNBC and non-TNBC groups. Discrete categories were grouped prior to the analysis as follows: For grade, category "I or II" included 26 samples and category "III" included 15 samples; for stage, category "moderately or well differentiated" included 26 samples and category "poorly differentiated" included 15 samples. A significance level of 0.05 was used for all these binary comparisons.

For the miRNAs that were selected from the integration with the array-CGH data, linear regression models were considered having the log-transformed miRNA values as the outcomes and tumor subtype, age, pathology, lymph node status, tumor grade or stage, and tumor size. Given the presence of cases with missing clinical-pathological data, both complete case analyses and multiple imputations to impute missing variables (with 10 imputed datasets) were performed using the `aregImpute` function in the `Hmisc` package version 4.1-1 in R [122, 123]. A significance level of FDR < 0.05 using the Benjamini and Hochberg FDR control method [124] was considered.

### Kaplan–Meier plot analysis

The KM Plotter Tool (<http://kmpplot.com/analysis/>) was used to calculate hazard ratios, confidence intervals, and log-rank *P* values for each of the selected 17 miRNAs in relation to survival in the aggregated breast cancer clinical studies extracted from The Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) databases (selected specifically for the TNBC subtype).

### Author contributions

B.M.S performed all the miRNA profiling analysis, data integration and wrote the manuscript; S.R.P., M.G. A.M, A.D., S.K. processed the samples and conducted the array-CGH assays and analysis; P.F. conducted the miRNA profiling assays; R.C.A. conducted the ancestral marker analysis; R.S.L, C.A.U. acquired and retrieved the clinical and pathological data; K.M., S.M.B. conducted the biostatistics analysis; S.M., Y.G. assisted in the miRNA data analysis and performed the deposit of the molecular data; I.J.C. Intellectual input and revision of the final version of the manuscript; E.M.S.F.R. Assisted in the overall data analysis and in the writing of the manuscript; L.R.C. conceived the study, coordinated the overall

experimental and data analysis and assisted in the writing of the manuscript.

## ACKNOWLEDGMENTS

We thank Dr. Deborah Berry and the staff of the Histopathology and Tissue Shared Resources (HTSR) of LCCC for assisting with the tissue sections of the cases and Dr. Ira Shoulson and Erin Wilhelm, principal investigator and executive director, respectively, of the Georgetown University Center of Excellence in Regulatory Science and Innovation (CERSI) program.

## CONFLICTS OF INTEREST

All the authors declare no conflicts of interest.

## FUNDING

This research was supported by the Georgetown University Center of Excellence in Regulatory Science and Innovation (CERSI; U01FD004319), a collaborative effort between the University and the U.S. Food and Drug Administration to promote regulatory science through innovative research and education. This research does not necessarily reflect the views of the FDA. We acknowledge the support of the Histopathology and Tissue Shared Resource (HTSR) and the Genomics and Epigenomics Shared Resources (GESR) at LCCC, partially supported by the NIH/NCI grant P30-CA051008. B.M.S. and S.R.P. received scholarships from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

## REFERENCES

1. Siegel RL, Fedewa SA, Miller KD, Goding-Sauer A, Pinheiro PS, Martinez-Tyson D, Jemal A. Cancer statistics for Hispanics/Latinos, 2015. *CA Cancer J Clin.* 2015; 65:457–480. <https://doi.org/10.3322/caac.21314>. [PubMed]
2. Cancer Facts AC. Figures for Hispanic/Latinos 2015–2017. In: Society AC, editor. Atlanta; 2015.
3. INCA/MS. (2018). Estimativa 2018 - Incidência de Câncer no Brasil. Rio de Janeiro.
4. Hunt BR. Breast Cancer Prevalence and Mortality among Hispanic Subgroups in the United States, 2009–2013. *J Cancer Epidemiol.* 2016; 2016:8784040. <https://doi.org/10.1155/2016/8784040>. [PubMed]
5. Justo N, Wilking N, Jonsson B, Luciani S, Cazap E. A review of breast cancer care and outcomes in Latin America. *Oncologist.* 2013; 18:248–256. <https://doi.org/10.1634/theoncologist.2012-0373>. [PubMed]
6. Schinkel JK, Zahm SH, Jatoi I, McGlynn KA, Gallagher C, Schairer C, Shriver CD, Zhu K. Racial/ethnic differences in breast cancer survival by inflammatory status and hormonal receptor status: an analysis of the Surveillance, Epidemiology, and End Results data. *Cancer Causes Control.* 2014; 25:959–968. <https://doi.org/10.1007/s10552-014-0395-1>. [PubMed]
7. Kohler BA, Sherman RL, Howlader N, Jemal A, Ryerson AB, Henry KA, Boscoe FP, Cronin KA, Lake A, Noone AM, Henley SJ, Ehemann CR, Anderson RN, Penberthy L. Annual Report to the Nation on the Status of Cancer, 1975–2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State. *J Natl Cancer Inst.* 2015; 107:djv048. <https://doi.org/10.1093/jnci/djv048>. [PubMed]
8. Noone AM, Cronin KA, Altekruse SF, Howlader N, Lewis DR, Petkov VI, Penberthy L. Cancer Incidence and Survival Trends by Subtype Using Data from the Surveillance Epidemiology and End Results Program, 1992–2013. *Cancer Epidemiol Biomarkers Prev.* 2017; 26:632–641. <https://doi.org/10.1158/1055-9965.EPI-16-0520>. [PubMed]
9. Brandzel S, Chang E, Tuzzio L, Campbell C, Coronado N, Bowles EJ, Bradford SC, Buist DS. Latina and Black/African American Women's Perspectives on Cancer Screening and Cancer Screening Reminders. *J Racial Ethn Health Disparities.* 2016; 4:1000–1008. <https://doi.org/10.1007/s40615-016-0304-2>. [PubMed]
10. Lynce F, Graves KD, Jandorf L, Ricker C, Castro E, Moreno L, Augusto B, Fejerman L, Vadaparampil ST. Genomic Disparities in Breast Cancer Among Latinas. *Cancer Control.* 2016; 23:359–372. <https://doi.org/10.1177/107327481602300407>. [PubMed]
11. Connor AE, Visvanathan K, Baumgartner KB, Baumgartner RN, Boone SD, Hines LM, Wolff RK, John EM, Slattery ML. Ethnic differences in the relationships between diabetes, early age adiposity and mortality among breast cancer survivors: the Breast Cancer Health Disparities Study. *Breast Cancer Res Treat.* 2016; 157:167–178. <https://doi.org/10.1007/s10549-016-3810-3>. [PubMed]
12. Arce-Salinas C, Aguilar-Ponce JL, Villarreal-Garza C, Lara-Medina FU, Olvera-Caraza D, Alvarado Miranda A, Flores-Diaz D, Mohar A. Overweight and obesity as poor prognostic factors in locally advanced breast cancer patients. *Breast Cancer Res Treat.* 2014; 146:183–188. <https://doi.org/10.1007/s10549-014-2977-8>. [PubMed]
13. Bandera EV, Chandran U, Hong CC, Troester MA, Bethea TN, Adams-Campbell LL, Haiman CA, Park SY, Olshan AF, Ambrosone CB, Palmer JR, Rosenberg L. Obesity, body fat distribution, and risk of breast cancer subtypes in African American women participating in the AMBER Consortium. *Breast Cancer Res Treat.* 2015; 150:655–666. <https://doi.org/10.1007/s10549-015-3353-z>. [PubMed]
14. Fejerman L, Stern MC, John EM, Torres-Mejia G, Hines LM, Wolff RK, Baumgartner KB, Giuliano AR, Ziv E, Perez-Stable EJ, Slattery ML. Interaction between common breast cancer susceptibility variants, genetic ancestry, and nongenetic risk factors in Hispanic women. *Cancer Epidemiol Biomarkers Prev.* 2015; 24:1731–1738. <https://doi.org/10.1158/1055-9965.EPI-15-0392>. [PubMed]

15. Connor AE, Baumgartner RN, Pinkston CM, Boone SD, Baumgartner KB. Obesity, ethnicity, and quality of life among breast cancer survivors and women without breast cancer: the long-term quality of life follow-up study. *Cancer Causes Control*. 2016; 27:115–124. <https://doi.org/10.1007/s10552-015-0688-z>. [PubMed]
16. Connor AE, Visvanathan K, Baumgartner KB, Baumgartner RN, Boone SD, Hines LM, Wolff RK, John EM, Slattery ML. Pre-diagnostic breastfeeding, adiposity, and mortality among parous Hispanic and non-Hispanic white women with invasive breast cancer: the Breast Cancer Health Disparities Study. *Breast Cancer Res Treat*. 2017; 161:321–331. <https://doi.org/10.1007/s10549-016-4048-9>. [PubMed]
17. Tejada S, Gallardo RI, Ferrans CE, Rauscher GH. Breast cancer delay in Latinas: the role of cultural beliefs and acculturation. *J Behav Med*. 2017; 40:343–351. <https://doi.org/10.1007/s10865-016-9789-8>. [PubMed]
18. Molina Y, Thompson B, Espinoza N, Ceballos R. Breast cancer interventions serving US-based Latinas: current approaches and directions. *Womens Health (Lond)*. 2013; 9:335–48; quiz 349–50. <https://doi.org/10.2217/whe.13.30>. [PubMed]
19. Friese CR, Martinez KA, Abrahamse P, Hamilton AS, Graff JJ, Jaggi R, Griggs JJ, Hawley ST, Katz SJ. Providers of follow-up care in a population-based sample of breast cancer survivors. *Breast Cancer Res Treat*. 2014; 144:179–184. <https://doi.org/10.1007/s10549-014-2851-8>. [PubMed]
20. O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, Dressler LG, Geradts J, Millikan RC. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res*. 2010; 16:6100–6110. <https://doi.org/10.1158/1078-0432.CCR-10-1533>. [PubMed]
21. Chavez-Macgregor M, Liu S, De Melo-Gagliato D, Chen H, Do KA, Pusztai L, Fraser Symmans W, Nair L, Hortobagyi GN, Mills GB, Meric-Bernstam F, Gonzalez-Angulo AM. Differences in gene and protein expression and the effects of race/ethnicity on breast cancer subtypes. *Cancer Epidemiol Biomarkers Prev*. 2014; 23:316–323. <https://doi.org/10.1158/1055-9965.EPI-13-0929>. [PubMed]
22. Kroenke CH, Sweeney C, Kwan ML, Quesenberry CP, Weltzien EK, Habel LA, Castillo A, Bernard PS, Factor RE, Kushi LH, Caan BJ. Race and breast cancer survival by intrinsic subtype based on PAM50 gene expression. *Breast Cancer Res Treat*. 2014; 144:689–699. <https://doi.org/10.1007/s10549-014-2899-5>. [PubMed]
23. Plasilova ML, Hayse B, Killelea BK, Horowitz NR, Chagpar AB, Lannin DR. Features of triple-negative breast cancer: Analysis of 38,813 cases from the national cancer database. *Medicine (Baltimore)*. 2016; 95:e4614. <https://doi.org/10.1097/MD.0000000000004614>. [PubMed]
24. Serrano-Gomez SJ, Fejerman L, Zabaleta J. Breast Cancer in Latinas: A Focus on Intrinsic Subtypes Distribution. *Cancer Epidemiol Biomarkers Prev*. 2018; 27:3–10. <https://doi.org/10.1158/1055-9965.EPI-17-0420>. [PubMed]
25. Patel TA, Colon-Otero G, Bueno Hume C, Copland JA 3rd, Perez EA. Breast cancer in Latinas: gene expression, differential response to treatments, and differential toxicities in Latinas compared with other population groups. *Oncologist*. 2010; 15:466–475. <https://doi.org/10.1634/theoncologist.2010-0004>. [PubMed]
26. Lara-Medina F, Perez-Sanchez V, Saavedra-Perez D, Blake-Cerda M, Arce C, Motola-Kuba D, Villarreal-Garza C, Gonzalez-Angulo AM, Bargallo E, Aguilar JL, Mohar A, Arrieta O. Triple-negative breast cancer in Hispanic patients: high prevalence, poor prognosis, and association with menopausal status, body mass index, and parity. *Cancer*. 2011; 117:3658–3669. <https://doi.org/10.1002/ncr.25961>. [PubMed]
27. Killelea BK, Yang VQ, Wang SY, Hayse B, Mougalian S, Horowitz NR, Chagpar AB, Pusztai L, Lannin DR. Racial Differences in the Use and Outcome of Neoadjuvant Chemotherapy for Breast Cancer: Results From the National Cancer Data Base. *J Clin Oncol*. 2015; 33:4267–4276. <https://doi.org/10.1200/JCO.2015.63.7801>. [PubMed]
28. Sheppard VB, Hicks J, Makambi K, Hurtado-de-Mendoza A, Demark-Wahnefried W, Adams-Campbell L. The feasibility and acceptability of a diet and exercise trial in overweight and obese black breast cancer survivors: The Stepping STONE study. *Contemp Clin Trials*. 2016; 46:106–113. <https://doi.org/10.1016/j.cct.2015.12.005>. [PubMed]
29. Warner ET, Ballman KV, Strand C, Boughey JC, Buzdar AU, Carey LA, Sikov WM, Partridge AH. Impact of race, ethnicity, and BMI on achievement of pathologic complete response following neoadjuvant chemotherapy for breast cancer: a pooled analysis of four prospective Alliance clinical trials (A151426). *Breast Cancer Res Treat*. 2016; 159:109–118. <https://doi.org/10.1007/s10549-016-3918-5>. [PubMed]
30. Villarreal-Garza C, Soto-Perez-de-Celis E. Ethnicity, adverse events and outcomes in breast cancer. *Breast*. 2015; 24:539–540. <https://doi.org/10.1016/j.breast.2015.05.005>. [PubMed]
31. Wu AH, Gomez SL, Vigen C, Kwan ML, Keegan TH, Lu Y, Shariff-Marco S, Monroe KR, Kurian AW, Cheng I, Caan BJ, Lee VS, Roh JM, et al. The California Breast Cancer Survivorship Consortium (CBCSC): prognostic factors associated with racial/ethnic differences in breast cancer survival. *Cancer Causes Control*. 2013; 24:1821–1836. <https://doi.org/10.1007/s10552-013-0260-7>. [PubMed]
32. Fejerman L, Ahmadiyah N, Hu D, Huntsman S, Beckman KB, Caswell JL, Tsung K, John EM, Torres-Mejia G, Carvajal-Carmona L, Echeverry MM, Tuazon AM, Ramirez C, et al, and COLUMBUS Consortium. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. *Nat Commun*. 2014; 5:5260. <https://doi.org/10.1038/ncomms6260>. [PubMed]
33. Kalinsky K, Lim EA, Andreopoulou E, Desai AM, Jin Z, Tu Y, Hibshoosh H, Wang A, Greenlee H, Crew KD, Maurer

- M, Sparano JA, Hershman DL. Increased expression of tumor proliferation genes in Hispanic women with early-stage breast cancer. *Cancer Invest.* 2014; 32:439–444. <https://doi.org/10.3109/07357907.2014.958232>. [PubMed]
34. Sly J, Jandorf L, Erwin DO. Who's Missing? Predictors of Attrition Following Participation in Culturally Targeted Educational Breast and Cervical Cancer Outreach Programs for Latinas. *J Health Commun.* 2015; 20:851–858. <https://doi.org/10.1080/108010730.2015.1018596>. [PubMed]
  35. Hamel LM, Penner LA, Albrecht TL, Heath E, Gwede CK, Eggly S. Barriers to Clinical Trial Enrollment in Racial and Ethnic Minority Patients With Cancer. *Cancer Control.* 2016; 23:327–337. <https://doi.org/10.1177/107327481602300404>. [PubMed]
  36. Zhang W, Liu J, Wang G. The role of microRNAs in human breast cancer progression. *Tumour Biol.* 2014; 35:6235–6244. <https://doi.org/10.1007/s13277-014-2202-8>. [PubMed]
  37. Takahashi RU, Miyazaki H, Ochiya T. The Roles of MicroRNAs in Breast Cancer. *Cancers (Basel).* 2015; 7:598–616. <https://doi.org/10.3390/cancers7020598>. [PubMed]
  38. Li JY, Jia S, Zhang WH, Zhang Y, Kang Y, Li PS. Differential distribution of microRNAs in breast cancer grouped by clinicopathological subtypes. *Asian Pac J Cancer Prev.* 2013; 14:3197–3203. <https://doi.org/10.7314/APJCP.2013.14.5.3197>. [PubMed]
  39. Riaz M, van Jaarsveld MT, Hollestelle A, Prager-van der Smissen WJ, Heine AA, Boersma AW, Liu J, Helmijr J, Ozturk B, Smid M, Wiemer EA, Foekens JA, Martens JW. miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. *Breast Cancer Res.* 2013; 15:R33. <https://doi.org/10.1186/bcr3415>. [PubMed]
  40. D'Ippolito E, Iorio MV. MicroRNAs and triple negative breast cancer. *Int J Mol Sci.* 2013; 14:22202–22220. <https://doi.org/10.3390/ijms141122202>. [PubMed]
  41. Gyparaki MT, Basdra EK, Papavassiliou AG. MicroRNAs as regulatory elements in triple negative breast cancer. *Cancer Lett.* 2014; 354:1–4. <https://doi.org/10.1016/j.canlet.2014.07.036>. [PubMed]
  42. Yang F, Zhang W, Shen Y, Guan X. Identification of dysregulated microRNAs in triple-negative breast cancer (review). *Int J Oncol.* 2015; 46:927–932. <https://doi.org/10.3892/ijo.2015.2821>. [PubMed]
  43. Zhou L, Liu F, Wang X, Ouyang G. The roles of microRNAs in the regulation of tumor metastasis. *Cell Biosci.* 2015; 5:32. <https://doi.org/10.1186/s13578-015-0028-8>. [PubMed]
  44. Tian W, Chen J, He H, Deng Y. MicroRNAs and drug resistance of breast cancer: basic evidence and clinical applications. *Clin Transl Oncol.* 2013; 15:335–342. <https://doi.org/10.1007/s12094-012-0929-5>. [PubMed]
  45. Wang J, Yang M, Li Y, Han B. The Role of MicroRNAs in the Chemoresistance of Breast Cancer. *Drug Dev Res.* 2015; 76:368–374. <https://doi.org/10.1002/ddr.21275>. [PubMed]
  46. Rawlings-Goss RA, Campbell MC, Tishkoff SA. Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. *BMC Med Genomics.* 2014; 7:53. <https://doi.org/10.1186/1755-8794-7-53>. [PubMed]
  47. Evans-Knowell A, LaRue AC, Findlay VJ. MicroRNAs and Their Impact on Breast Cancer, the Tumor Microenvironment, and Disparities. *Adv Cancer Res.* 2017; 133:51–76. <https://doi.org/10.1016/bs.acr.2016.08.003>. [PubMed]
  48. Bushel PR, McGovern R, Liu L, Hofmann O, Huda A, Lu J, Hide W, Lin X. Population differences in transcript-regulator expression quantitative trait loci. *PLoS One.* 2012; 7:e34286. <https://doi.org/10.1371/journal.pone.0034286>. [PubMed]
  49. Yao S, Graham K, Shen J, Campbell LE, Singh P, Zirpoli G, Roberts M, Ciupak G, Davis W, Hwang H, Khoury T, Bovbjerg DH, Jandorf L, et al. Genetic variants in microRNAs and breast cancer risk in African American and European American women. *Breast Cancer Res Treat.* 2013; 141:447–459. <https://doi.org/10.1007/s10549-013-2698-4>. [PubMed]
  50. Sugita B, Gill M, Mahajan A, Duttargi A, Kirolikar S, Almeida R, Regis K, Oluwasanmi OL, Marchi F, Marian C, Makambi K, Kallakury B, Sheahan L, et al. Differentially expressed miRNAs in triple negative breast cancer between African-American and non-Hispanic white women. *Oncotarget.* 2016; 7:79274–79291. <https://doi.org/10.18632/oncotarget.13024>. [PubMed]
  51. Wang PY, Gao ZH, Jiang ZH, Li XX, Jiang BF, Xie SY. The associations of single nucleotide polymorphisms in miR-146a, miR-196a and miR-499 with breast cancer susceptibility. *PLoS One.* 2013; 8:e70656. <https://doi.org/10.1371/journal.pone.0070656>. [PubMed]
  52. Khan S, Greco D, Michailidou K, Milne RL, Muranen TA, Heikkinen T, Aaltonen K, Dennis J, Bolla MK, Liu J, Hall P, Irwanto A, Humphreys K, et al. MicroRNA related polymorphisms and breast cancer risk. *PLoS One.* 2014; 9:e109973. <https://doi.org/10.1371/journal.pone.0109973>. [PubMed]
  53. Bensen JT, Xu Z, Smith GJ, Mohler JL, Fontham ET, Taylor JA. Genetic polymorphism and prostate cancer aggressiveness: a case-only study of 1,536 GWAS and candidate SNPs in African-Americans and European-Americans. *Prostate.* 2013; 73:11–22. <https://doi.org/10.1002/pros.22532>. [PubMed]
  54. Wang W, Ingles SA, Torres-Mejia G, Stern MC, Stanczyk FZ, Schwartz GG, Nelson DO, Fejerman L, Wolff RK, Slattery ML, John EM. Genetic variants and non-genetic factors predict circulating vitamin D levels in Hispanic and non-Hispanic White women: the Breast Cancer Health Disparities Study. *Int J Mol Epidemiol Genet.* 2014; 5:31–46. [PubMed]

55. Toyama T, Kondo N, Endo Y, Sugiura H, Yoshimoto N, Iwasa M, Takahashi S, Fujii Y, Yamashita H. High expression of microRNA-210 is an independent factor indicating a poor prognosis in Japanese triple-negative breast cancer patients. *Jpn J Clin Oncol*. 2012; 42:256–263. <https://doi.org/10.1093/jjco/hys001>. [PubMed]
56. Arpino G, Milano M, De Placido S. Features of aggressive breast cancer. *Breast*. 2015; 24:594–600. <https://doi.org/10.1016/j.breast.2015.06.001>. [PubMed]
57. Pierga JY, Hajage D, Bachelot T, Delaloge S, Brain E, Campone M, Dieras V, Rolland E, Mignot L, Mathiot C, Bidard FC. High independent prognostic and predictive value of circulating tumor cells compared with serum tumor markers in a large prospective trial in first-line chemotherapy for metastatic breast cancer patients. *Ann Oncol*. 2012; 23:618–624. <https://doi.org/10.1093/annonc/mdr263>. [PubMed]
58. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010; 363:1938–1948. <https://doi.org/10.1056/NEJMra1001389>. [PubMed]
59. Villarreal-Garza C, Mohar A, Bargallo-Rocha JE, Lasagunsebatt F, Reynoso-Noverson N, Matus-Santos J, Cabrera P, Arce-Salinas C, Lara-Medina F, Alvarado-Miranda A, Ramirez-Ugalde MT, Soto-Perez-de-Celis E. Molecular Subtypes and Prognosis in Young Mexican Women With Breast Cancer. *Clin Breast Cancer*. 2017; 17:e95–e102. <https://doi.org/10.1016/j.clbc.2016.11.007>. [PubMed]
60. Martinez ME, Wertheim BC, Natarajan L, Schwab R, Bondy M, Daneri-Navarro A, Meza-Montenegro MM, Gutierrez-Millan LE, Brewster A, Komenaka IK, Thompson PA. Reproductive factors, heterogeneity, and breast tumor subtypes in women of Mexican descent. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:1853–1861. <https://doi.org/10.1158/1055-9965.EPI-13-0560>. [PubMed]
61. Srur-Rivero N, Cartin-Brenes M. Breast cancer characteristics and survival in a Hispanic population of Costa Rica. *Breast Cancer (Auckl)*. 2014; 8:103–108. <https://doi.org/10.4137/BCBCR.S15854>. [PubMed]
62. Carvalho FM, Bacchi LM, Pincerato KM, Van de Rijn M, Bacchi CE. Geographic differences in the distribution of molecular subtypes of breast cancer in Brazil. *BMC Womens Health*. 2014; 14:102. <https://doi.org/10.1186/1472-6874-14-102>. [PubMed]
63. de Macedo Andrade AC, Ferreira Júnior CA, Dantas Guimaraes B, Pessoa Barros AW, Sarmiento de Almeida G, Weller M. Molecular breast cancer subtypes and therapies in a public hospital of northeastern Brazil. *BMC Womens Health*. 2014; 14:110. <https://doi.org/10.1186/1472-6874-14-110>. [PubMed]
64. Jeronimo AFA, Weller M. Differential Association of the Lifestyle-Related Risk Factors Smoking and Obesity with Triple Negative Breast Cancer in a Brazilian Population. *Asian Pac J Cancer Prev*. 2017; 18:1585–1593. <https://doi.org/10.22034/APJCP.2017.18.6.1585>. [PubMed]
65. Martinez ME, Gomez SL, Tao L, Cress R, Rodriguez D, Unkart J, Schwab R, Nodora JN, Cook L, Komenaka I, Li C. Contribution of clinical and socioeconomic factors to differences in breast cancer subtype and mortality between Hispanic and non-Hispanic white women. *Breast Cancer Res Treat*. 2017; 166:185–193. <https://doi.org/10.1007/s10549-017-4389-z>. [PubMed]
66. 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:e82460. <https://doi.org/10.1371/journal.pone.0082460>. [PubMed]
67. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA, and 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. <https://doi.org/10.1038/nature11632>. [PubMed]
68. Santos SC, Cavalli IJ, Ribeiro EM, Urban CA, Lima RS, Bleggi-Torres LF, Rone JD, Haddad BR, Cavalli LR. Patterns of DNA copy number changes in sentinel lymph node breast cancer metastases. *Cytogenet Genome Res*. 2008; 122:16–21. <https://doi.org/10.1159/000151311>. [PubMed]
69. Dai X, Chen A, Bai Z. Integrative investigation on breast cancer in ER, PR and HER2-defined subgroups using mRNA and miRNA expression profiling. *Sci Rep*. 2014; 4:6566. <https://doi.org/10.1038/srep06566>. [PubMed]
70. Bhattacharyya M, Nath J, Bandyopadhyay S. MicroRNA signatures highlight new breast cancer subtypes. *Gene*. 2015; 556:192–198. <https://doi.org/10.1016/j.gene.2014.11.053>. [PubMed]
71. Iuliano R, Vismara MF, Dattilo V, Trapasso F, Baudi F, Perrotti N. The role of microRNAs in cancer susceptibility. *Biomed Res Int*. 2013; 2013:591931. <https://doi.org/10.1155/2013/591931>. [PubMed]
72. Han W, Jung EM, Cho J, Lee JW, Hwang KT, Yang SJ, Kang JJ, Bae JY, Jeon YK, Park IA, Nicolau M, Jeffrey SS, Noh DY. DNA copy number alterations and expression of relevant genes in triple-negative breast cancer. *Genes Chromosomes Cancer*. 2008; 47:490–499. <https://doi.org/10.1002/gcc.20550>. [PubMed]
73. Hua K, Jin J, Zhao J, Song J, Song H, Li D, Maskey N, Zhao B, Wu C, Xu H, Fang L. miR-135b, upregulated in breast cancer, promotes cell growth and disrupts the cell cycle by regulating LATS2. *Int J Oncol*. 2016; 48:1997–2006. <https://doi.org/10.3892/ijo.2016.3405>. [PubMed]
74. Lv ZD, Xin HN, Yang ZC, Wang WJ, Dong JJ, Jin LY, Li FN. miR-135b promotes proliferation and metastasis by targeting APC in triple-negative breast cancer. *J Cell Physiol*. 2019; 234:10819–10826. <https://doi.org/10.1002/jcp.27906>. [PubMed]
75. Aakula A, Leivonen SK, Hintsanen P, Aittokallio T, Ceder Y, Borresen-Dale AL, Perala M, Ostling P, Kallioniemi

- O. MicroRNA-135b regulates ERalpha, AR and HIF1AN and affects breast and prostate cancer cell growth. *Mol Oncol.* 2015; 9:1287–1300. <https://doi.org/10.1016/j.molonc.2015.03.001>. [PubMed]
76. He H, Tian W, Chen H, Jiang K. MiR-944 functions as a novel oncogene and regulates the chemoresistance in breast cancer. *Tumour Biol.* 2016; 37:1599–1607. <https://doi.org/10.1007/s13277-015-3844-x>. [PubMed]
  77. Paszek S, Gablo N, Barnas E, Szybka M, Morawiec J, Kolacinska A, Zawlik I. Dysregulation of microRNAs in triple-negative breast cancer. *Ginekol Pol.* 2017; 88:530–536. <https://doi.org/10.5603/GPa2017.0097>. [PubMed]
  78. Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C, Benes V, Schmidt S, Blake J, Ball G, Kerin MJ. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res.* 2009; 11:R27. <https://doi.org/10.1186/bcr2257>. [PubMed]
  79. Pu T, Shen M, Li S, Yang L, Gao H, Xiao L, Zhong X, Zheng H, Liu Y, Ye F, Bu H. Repression of miR-135b-5p promotes metastasis of early-stage breast cancer by regulating downstream target SDCBP. *Lab Invest.* 2019; 99:1296–1308. <https://doi.org/10.1038/s41374-019-0258-1>. [PubMed]
  80. Uva P, Cossu-Rocca P, Loi F, Pira G, Murgia L, Orru S, Floris M, Muroli MR, Sanges F, Carru C, Angius A, De Miglio MR. miRNA-135b Contributes to Triple Negative Breast Cancer Molecular Heterogeneity: Different Expression Profile in Basal-like Versus non-Basal-like Phenotypes. *Int J Med Sci.* 2018; 15:536–548. <https://doi.org/10.7150/ijms.23402>. [PubMed]
  81. Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG, Ronneberg JA, Johnsen H, Navon R, Rodland E, Makela R, Naume B, Perala M, et al. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One.* 2011; 6:e16915. <https://doi.org/10.1371/journal.pone.0016915>. [PubMed]
  82. Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, Tavare S, Caldas C, Miska EA. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol.* 2007; 8:R214. <https://doi.org/10.1186/gb-2007-8-10-r214>. [PubMed]
  83. Arigoni M, Barutello G, Riccardo F, Ercole E, Cantarella D, Orso F, Conti L, Lanzardo S, Taverna D, Merighi I, Calogero RA, Cavallo F, Quaglino E. miR-135b coordinates progression of ErbB2-driven mammary carcinomas through suppression of MID1 and MTCH2. *Am J Pathol.* 2013; 182:2058–2070. <https://doi.org/10.1016/j.ajpath.2013.02.046>. [PubMed]
  84. Pena-Chilet M, Martinez MT, Perez-Fidalgo JA, Peiro-Chova L, Oltra SS, Tormo E, Alonso-Yuste E, Martinez-Delgado B, Eroles P, Climent J, Burgues O, Ferrer-Lozano J, Bosch A, et al. MicroRNA profile in very young women with breast cancer. *BMC Cancer.* 2014; 14:529. <https://doi.org/10.1186/1471-2407-14-529>. [PubMed]
  85. Cordero F, Ferrero G, Polidoro S, Fiorito G, Campanella G, Sacerdote C, Mattiello A, Masala G, Agnoli C, Frasca G, Panico S, Palli D, Krogh V, et al. Differentially methylated microRNAs in prediagnostic samples of subjects who developed breast cancer in the European Prospective Investigation into Nutrition and Cancer (EPIC-Italy) cohort. *Carcinogenesis.* 2015; 36:1144–1153. <https://doi.org/10.1093/carcin/bgv102>. [PubMed]
  86. Godbole M, Chandrani P, Gardi N, Dhamne H, Patel K, Yadav N, Gupta S, Badwe R, Dutt A. miR-129-2 mediates down-regulation of progesterone receptor in response to progesterone in breast cancer cells. *Cancer Biol Ther.* 2017; 18:801–805. <https://doi.org/10.1080/15384047.2017.1373216>. [PubMed]
  87. Leivonen SK, Sahlberg KK, Makela R, Due EU, Kallioniemi O, Borresen-Dale AL, Perala M. High-throughput screens identify microRNAs essential for HER2 positive breast cancer cell growth. *Mol Oncol.* 2014; 8:93–104. <https://doi.org/10.1016/j.molonc.2013.10.001>. [PubMed]
  88. Shi Y, Qiu M, Wu Y, Hai L. MiR-548-3p functions as an anti-oncogenic regulator in breast cancer. *Biomed Pharmacother.* 2015; 75:111–116. <https://doi.org/10.1016/j.biopha.2015.07.027>. [PubMed]
  89. Ke H, Zhao L, Feng X, Xu H, Zou L, Yang Q, Su X, Peng L, Jiao B. NEAT1 is Required for Survival of Breast Cancer Cells Through FUS and miR-548. *Gene Regul Syst Bio.* 2016; 10:11–17. <https://doi.org/10.4137/GRSB.S29414>. [PubMed]
  90. Guo J, Gong G, Zhang B. miR-539 acts as a tumor suppressor by targeting epidermal growth factor receptor in breast cancer. *Sci Rep.* 2018; 8:2073. <https://doi.org/10.1038/s41598-018-20431-z>. [PubMed]
  91. Yang ZX, Zhang B, Wei J, Jiang GQ, Wu YL, Leng BJ, Xing CG. MiR-539 inhibits proliferation and migration of triple-negative breast cancer cells by down-regulating LAMA4 expression. *Cancer Cell Int.* 2018; 18:16. <https://doi.org/10.1186/s12935-018-0512-4>. [PubMed]
  92. Romero-Cordoba SL, Rodriguez-Cuevas S, Bautista-Pina V, Maffuz-Aziz A, D'Ippolito E, Cosentino G, Baroni S, Iorio MV, Hidalgo-Miranda A. Loss of function of miR-342-3p results in MCT1 over-expression and contributes to oncogenic metabolic reprogramming in triple negative breast cancer. *Sci Rep.* 2018; 8:12252. <https://doi.org/10.1038/s41598-018-29708-9>. [PubMed]
  93. He YJ, Wu JZ, Ji MH, Ma T, Qiao EQ, Ma R, Tang JH. miR-342 is associated with estrogen receptor-alpha expression and response to tamoxifen in breast cancer. *Exp Ther Med.* 2013; 5:813–818. <https://doi.org/10.3892/etm.2013.915>. [PubMed]
  94. Cittelly DM, Das PM, Spoelstra NS, Edgerton SM, Richer JK, Thor AD, Jones FE. Downregulation of miR-342 is associated with tamoxifen resistant breast tumors. *Mol Cancer.* 2010; 9:317. <https://doi.org/10.1186/1476-4598-9-317>. [PubMed]

95. Young J, Kawaguchi T, Yan L, Qi Q, Liu S, Takabe K. Tamoxifen sensitivity-related microRNA-342 is a useful biomarker for breast cancer survival. *Oncotarget*. 2017; 8:99978–99989. <https://doi.org/10.18632/oncotarget.21577>. [PubMed]
96. Wang X, Jiang D, Xu C, Zhu G, Wu Z, Wu Q. Differential expression profile analysis of miRNAs with HER-2 overexpression and intervention in breast cancer cells. *Int J Clin Exp Pathol*. 2017; 10:5039–5062.
97. Luo M, Ding L, Li Q, Yao H. miR-668 enhances the radioresistance of human breast cancer cell by targeting IκBα. *Breast Cancer*. 2017; 24:673–82. <https://doi.org/10.1007/s12282-017-0756-1>. [PubMed]
98. Madhavan D, Peng C, Wallwiener M, Zucknick M, Nees J, Schott S, Rudolph A, Riethdorf S, Trumpp A, Pantel K, Sohn C, Chang-Claude J, Schneeweiss A, Burwinkel B. Circulating miRNAs with prognostic value in metastatic breast cancer and for early detection of metastasis. *Carcinogenesis*. 2016; 37:461–470. <https://doi.org/10.1093/carcin/bgw008>. [PubMed]
99. Sand M, Skrygan M, Sand D, Georgas D, Gambichler T, Hahn SA, Altmeyer P, Bechara FG. Comparative microarray analysis of microRNA expression profiles in primary cutaneous malignant melanoma, cutaneous malignant melanoma metastases, and benign melanocytic nevi. *Cell Tissue Res*. 2013; 351:85–98. <https://doi.org/10.1007/s00441-012-1514-5>. [PubMed]
100. Khan FH, Pandian V, Ramraj S, Aravindan S, Herman TS, Aravindan N. Reorganization of metastamiRs in the evolution of metastatic aggressive neuroblastoma cells. *BMC Genomics*. 2015; 16:501. <https://doi.org/10.1186/s12864-015-1642-x>. [PubMed]
101. Said R, Garcia-Mayea Y, Trabelsi N, Setti Boubaker N, Mir C, Blél A, Ati N, Paciucci R, Hernández-Losa J, Rammeh S, Derouiche A, Chebil M, LLeonart ME, Ouerhani S. Expression patterns and bioinformatic analysis of miR-1260a and miR-1274a in Prostate Cancer Tunisian patients. *Mol Biol Rep*. 2018; 45:2345–2358. <https://doi.org/10.1007/s11033-018-4399-x>. [PubMed]
102. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013; 495:333–338. <https://doi.org/10.1038/nature11928>. [PubMed]
103. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, Hicks DG, Lester S, Love R, 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; 28:2784–2795. <https://doi.org/10.1200/JCO.2009.25.6529>. [PubMed]
104. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, Bilous M, Fitzgibbons P, Hanna W, Jenkins RB, Mangu PB, et al, and American Society of Clinical Oncology, and College of American Pathologists. 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. *Arch Pathol Lab Med*. 2014; 138:241–256. <https://doi.org/10.5858/arpa.2013-0953-SA>. [PubMed]
105. Torresan C, Oliveira MM, Pereira SR, Ribeiro EM, Marian C, Gusev Y, Lima RS, Urban CA, Berg PE, Haddad BR, Cavalli IJ, Cavalli LR. Increased copy number of the DLX4 homeobox gene in breast axillary lymph node metastasis. *Cancer Genet*. 2014; 207:177–187. <https://doi.org/10.1016/j.cancergen.2014.04.007>. [PubMed]
106. Ruiz-Linares A, Adhikari K, Acuna-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F, Gomez-Valdes J, Leon-Mimila P, Hunemeier T, et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet*. 2014; 10:e1004572. <https://doi.org/10.1371/journal.pgen.1004572>. [PubMed]
107. Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, Pons-Estel BA, Acevedo-Vasquez E, Miranda P, Langefeld CD, Gravel S, Alarcon-Riquelme ME, Bustamante CD. Genomic Insights into the Ancestry and Demographic History of South America. *PLoS Genet*. 2015; 11:e1005602. <https://doi.org/10.1371/journal.pgen.1005602>. [PubMed]
108. Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, Sofer T, Fernandez-Rhodes L, Justice AE, Graff M, Young KL, Seyerle AA, Avery CL, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet*. 2016; 98:165–184. <https://doi.org/10.1016/j.ajhg.2015.12.001>. [PubMed]
109. Healy ME, Hill D, Berwick M, Edgar H, Gross J, Hunley K. Social-group identity and population substructure in admixed populations in New Mexico and Latin America. *PLoS One*. 2017; 12:e0185503. <https://doi.org/10.1371/journal.pone.0185503>. [PubMed]
110. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015; 4:7. <https://doi.org/10.1186/s13742-015-0047-8>. [PubMed]
111. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2:e190. <https://doi.org/10.1371/journal.pgen.0020190>. [PubMed]
112. Moura RR, Coelho AV, Balbino Vde Q, Crovella S, Brandao LA. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am J Hum Biol*. 2015; 27:674–680. <https://doi.org/10.1002/ajhb.22714>. [PubMed]

113. Salzano FM, Sans M. Interethnic admixture and the evolution of Latin American populations. *Genet Mol Biol.* 2014; 37:151–170. <https://doi.org/10.1590/S1415-47572014000200003>. [PubMed]
114. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* 2013; 41:W169–73. <https://doi.org/10.1093/nar/gkt393>. [PubMed]
115. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* 2015; 43:D146–D152. <https://doi.org/10.1093/nar/gku1104>. [PubMed]
116. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* 2015; 4:e05005. <https://doi.org/10.7554/eLife.05005>. [PubMed]
117. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.* 2015; 43:W460–6. <https://doi.org/10.1093/nar/gkv403>. [PubMed]
118. Nersisyan L, Samsouyan R, Arakelyan A. CyKEGGParser: tailoring KEGG pathways to fit into systems biology analysis workflows. *F1000Res.* 2014; 3:145. <https://doi.org/10.12688/f1000research.4410.2>. [PubMed]
119. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics.* 2010; 26:2927–2928. <https://doi.org/10.1093/bioinformatics/btq562>. [PubMed]
120. Kutmon M, Kelder T, Mandaviya P, Evelo CT, Coort SL. CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. *PLoS One.* 2013; 8:e82160. <https://doi.org/10.1371/journal.pone.0082160>. [PubMed]
121. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–2504. <https://doi.org/10.1101/gr.1239303>. [PubMed]
122. Team RDC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2011.
123. Harrell FE Jr. (2016). Package “Hmisc”.
124. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995; 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.