

Prevalence of Breast Cancer Subtypes Among Different Ethnicities and Bangladeshi Women: Demographic, Clinicopathological, and Integrated Cancer Informatics Analysis

Cancer Informatics
Volume 22: 1–14
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351221148584



Diganta Islam¹, Md. Shihabul Islam², Sanjida Islam Dorin¹
and Jesmin¹

¹Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka, Bangladesh.

²Department of Genetic Engineering & Biotechnology, University of Rajshahi, Rajshahi, Bangladesh.

ABSTRACT

BACKGROUND: The molecular subtyping of breast cancer is related to estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). The present study aimed to systematically analyze the expression, function, and prognostic value of ER, PR, HER2, and their prevalence in different ethnic groups and among Bangladeshi breast cancer (BC) patients.

METHOD: This study included 25 BC patients and 25 healthy controls, aged between 25 and 70 years. The study characteristics were compared using the ANOVA and Chi-square tests. Also, the multi-Omics dataset of 775 BC patients from TCGA was analyzed for ER, PR, and HER2 in breast cancer subtypes and compared among different ethnicities.

RESULTS: For most BD breast cancer cases, the age at diagnosis was ≥ 40 years, had only a histopathological diagnosis (P -value .004), and no history of mammography or other pathological tests. For treatment, had only chemotherapy (P -value .004) and no hormone therapy (P -value $< .001$). The majority of patients ($>60\%$) were of stage-II cancer and TNBC (40%) subtype. The BC ethnicity-stratified data of ER, PR, and HER2 indicated a strong correlation across all ethnicities (P -value $4.99e-35$; P -value $3.79e-18$). The subtypes stratified data indicated a higher percentage of Luminal A (58.3%) in Caucasians whereas Luminal B (24.3%) and HER2 (25.2%) subtypes were found higher in Asians and TNBC (36.0%) were found in Africans. However, a significantly higher frequency of TNBC (52.2%) compared to Asians (14.8%) was found in BD patients (P -value $< .001$). The overall survival analysis of BC subtypes demonstrated that Luminal B (P -value .005) and HER2 enriched (P -value .015) were significantly more aggressive and were dominant in the Asian population.

CONCLUSION: A significant association was found between BC subtypes with different ethnicities and Bangladeshi women and these findings might aid in the prevention, management, and raising of awareness against risk factors in the near future.

KEYWORDS: Breast cancer, subtypes, luminal A, luminal B, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), triple-negative breast cancer (TNBC), cBioPortal, TCGA, survival analysis

RECEIVED: October 27, 2022. ACCEPTED: December 13, 2022.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Ministry of Science & Technology (MOST), Government of the People's Republic of Bangladesh for supporting DI (NST Fellowship 39.012.002.01.03.021.2014-09/260) to conduct this research work.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Jesmin, Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh. Email: jesmin@du.ac.bd

Background

Breast cancer is the most prevalent cancer in women and is a leading cause of death, worldwide. In 2020, there were 2.3 million women diagnosed with breast cancer and 685,000 deaths occurred globally.^{1,2} Almost half of the incidence (45.4%) and mortality (50.6%) cases were observed in Asia.^{3,4} Breast cancer is not a transmissible disease. The female gender is the strongest risk factor in association with breast cancer. In almost half of breast cancer incidences, women have no other potential risk factor other than gender (female) and age (over 40 years). Some other factors like obesity, alcohol consumption, family history of breast cancer, radiation exposure, reproductive history, such as age that menstrual periods began and age at first pregnancy, tobacco use, and postmenopausal hormone therapy increase the risk of breast cancer.⁵⁻¹¹

Most breast cancer arises in the epithelium cells of the ducts (85%) or lobules (15%) tissues in the breast. If the cancerous growth remains confined to the duct or lobule causes no symptoms and has minimal potential for spreading (metastasis). However, if, over time, the cancers spread and invade the surrounding breast tissue (invasive breast cancer), and then progress to the lymph nodes (regional metastasis) or other organs in the body (distant metastasis), could be life-threatening. Most of the women who die from breast cancer are because of widespread metastasis.^{2,6,7} Breast cancer treatment can be highly effective, with survival probabilities of 90% or higher, especially when the disease is identified early. Treatment generally consists of a combination of surgery, radiation therapy, and anti-cancer medication (endocrine hormonal therapy, chemotherapy, and/or targeted biologic therapy, antibodies) to



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

control, treat and/or reduce the risk of cancer spreading (metastasis).^{2,5,6}

Medical treatments for breast cancer are based on the molecular subtyping of the cancers: depending on the expression (+) or not (-) of the 3 essential receptor proteins: estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2).^{2,9} Cancer that expresses the ER and/or PR, is likely to respond to hormone therapies. Breast cancers may independently overexpress the HER2 oncogene. These HER2-enriched cancers are amenable to treatment with targeted biological agents. The major subtypes based on hormonal receptors and progressions are Luminal A, Luminal B, HER2 positive, and Basal type or Triple-Negative Breast Cancer (TNBC).^{2,12-15} As the treatment process relies heavily on these molecular subtypes of breast cancer, the prevalence of the hormonal receptors in various ethnic groups across the globe may indicate or reveal which subtypes of breast cancer are more prevalent in a specific ethnic community. Such studies had been conducted before for different races.¹⁵⁻¹⁸ The hormone receptors are used clinically for diagnosis and treatment and the prevalence may vary for every race.¹⁸⁻²¹

By 2040, the incidence and mortality of breast cancer patients will increase globally, dominantly in Asia and Bangladesh.²² The molecular understanding of breast cancer requires characterizing its genetic alterations like mutations, copy number variations, and structural changes in tumor DNA that might cause changes in transcription, protein activity, and epigenetic regulations. The present study aimed to systematically analyze the expression, function, and prognostic value of ER, PR, and HER2 in different breast cancer subtypes. The main objective was to assess the prevalence of the hormone receptors (ER, PR), and epidermal growth factor receptor (HER2) in Bangladeshi BC patients by following through with treatment, diagnosis, demographic, and clinical data analysis. To compare the breast cancer subtypes in different races (Caucasian, African, and Asian) multi-omics data were retrieved from TCGA (The Cancer Genome Atlas) and cBioPortal^{19-21,23-26} and evaluated through bioinformatics analysis using various tools and databases like cBioPortal, GEPIA2, GISTIC, HPA, PDB, etc. Also, conducting an in-detailed evaluation of the hormone receptors ER, PR, and epidermal growth factor receptor HER2 (levels of mRNA, protein, mRNA vs protein expression, and mutual exclusivity analysis) for the Asian population. The goal was to interrelate genomic subtype data with clinical features to understand how these might affect clinically relevant phenotypes such as survival. In particular, assessing the link between survival and expression of the ER, PR, and HER2 receptors in breast cancer to validate their prognostic relevance in overall survival in different ethnicity. This study is also an attempt to facilitate better management, diagnosis and treatment for Bangladeshi women.

Methods

Patients and healthy controls: This study included 25 random breast cancer patients and 25 healthy controls between 25 and 70 years of age by random sampling. The breast cancer patient samples were collected from the National Institute of Cancer Research and Hospital (NICRH), Mohakhali, Dhaka, during the period from December 2021 to February 2022. Breast cancer was confirmed by the histopathological evaluation. The inclusion criteria for this study were: females between 25 and 70 years of age, confirmed breast cancer patients (based on medical reports), not pregnant, had no major comorbidities and willingly agreed to participate. The clinicopathological characteristics of breast cancer patients including the molecular subtyping, histopathological type, grade, stage of the disease, and the expression of reproductive hormones ER, PR, and the epidermal growth receptor HER2 were obtained from the patient's medical records. The study was approved by the Ethics Committee of the National Institute of Cancer Research and Hospital (NICRH) and written informed consent was obtained from all study participants.

Data collection: A semi-structured questionnaire was used to collect data from the study participants. Information was collected through interviews and data about the disease, diagnosis and treatment of breast cancer patients were taken from their medical records. Sociodemographic data and participants' personal information like menstrual history, menarche age, food habits, lifestyle, family history of the disease, and medical history of the study participants were collected. Breast cancer stage, type, and hormonal status were also collected from their medical records. All collected data were stored both digitally and manually in different record files. Anthropometric data of the study participants such as weight were measured using a digital balance, and height, the circumference of the hip, and the waist were measured using a standard measuring tape. Then, body mass index (BMI), and waist-to-hip ratio (WHR) were calculated through standard formula. BMI ≥ 30 was considered obese and WHR ≥ 0.85 was considered above the normal range.²⁷ The study participants' blood pressure was measured using a digital sphygmomanometer (Normal range for systolic: ≤ 120 mmHg, for diastolic: ≤ 80 mmHg). The study was conducted fully, ensuring the participants' privacy and anonymity and following the Declaration of Helsinki guidelines throughout the process.

Clinicopathological tests: The study participants' blood samples were collected into BD Vacutainer® tubes by trained authorized personnel. All biochemical analysis of the blood samples was conducted at the Department of Biochemistry and Molecular Biology, Bangabandhu Sheikh Mujib Medical University (BSMMU), and their reference values were used to evaluate the study characteristics.

Ethical Approval: The protocol for this study was approved and ethical clearance was provided by the Institutional review board (IRB) of the National Institute of Cancer Research and

Hospital (NICRH) (Reference: NICRH/Ethics/2021/323). All participants were properly informed about the procedures of the study and their verbal and written consents were taken before inclusion in the study.

Statistical analysis: All data were analyzed using SPSS Statistics 20 (Version 27.0. Armonk, NY: IBM Corp.2020), and Microsoft Excel (2013). The number of participants was presented as a whole number, percentage, and average \pm standard deviation. Chi-square and ANOVA tests were performed and a *P*-value less than .05 was considered statistically significant.

Cancer informatics analysis

The cBio Cancer Genomics Portal (cBioPortal), an online platform for cancer genomics data (<http://www.cbioportal.org>) was used to retrieve and analyze multi-omics data from The Cancer Genome Atlas (TCGA, Pan-Cancer).^{20,21,23-26,28,29} A total of 24 datasets of Breast cancer (BC) from cBioPortal were screened for inclusion based on the following criteria: female, presence of race/ethnicity data, information about cancer subtypes based on estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 status (HER2). After screening, the cBioPortal breast invasive carcinoma dataset (TCGA, Pan-Cancer Atlas) containing 1084 samples was selected for further investigation.³⁰⁻³⁶

Data source: The *z*-score-transformed expression data of 1084 samples (*n* = 1084; publicly available cases) of breast invasive carcinoma (TCGA, Pan-Cancer Atlas) were downloaded for further bioinformatics analysis from the cBioPortal (<https://www.cbioportal.org>). In this study, the 1084 dataset was screened for up-regulation, with a minimum of 2-fold higher expression values of the cancer subtypes based on estrogen (ER, *ESR1* gene), progesterone (PR, *PGR* gene), and human epidermal growth factor receptor 2 status (HER2, *ERBB2* gene). The genomic data of ER, PR, and HER2 containing mutations, putative copy-number alterations, with an mRNA expression *z*-scores (RNASeq V2 RSEM) threshold ± 2.0 were retrieved from GISTIC (Genomic Identification of Significant Targets in Cancer). After screening for BC subtypes, RNAseq data of 149 TNBC (Triple Negative or Basal BC) samples, 561 Luminal BC (Luminal A or Luminal B breast cancer) samples, and 65 HER2 positive BC samples and their corresponding clinical information were downloaded from the National Cancer Institute Genome Data Sharing web portal TCGA for further analysis (Supplemental File S1). All gene expression data were normalized by log₂ transformation. Comparison between normal tissue and cancer tissue used 2 sets of *t*-tests; *P*-value <.05 indicated the statistical significance. The cBioPortal for Cancer Genomics database was used to analyze the frequency of gene alterations (including mutations, deletions, copy number gains, and amplifications) in breast cancer, and the "Plots tab" was used to perform

correlation analysis between genomics data (copy number alteration and gene expression), and proteomics (RPPA, and mass spectrometry-based) data, where mutation and copy number data were overlaid onto the correlation plots. The Kaplan-Meier curve, log-rank test, and Cox proportional hazard regression model were used for all survival analyses in this study. The correlation analysis between the 2 variables used Spearman's or Pearson's test; *P*-value <.05 were considered significant. All analyses were conducted according to the cBioPortal guidelines.

Patient population analysis: For the BC patient population analysis, information on patients' age, sex, tumor stage, breast cancer subtype, and race were evaluated using the OncoPrint in cBioPortal. Co-expression of ER, PR, and HER2 to BC was conducted using the GTEx and TCGA data in GEPIA2.³⁷ In cBioPortal, co-expression analysis for ER, PR, and HER2 genes, Pearson's and Spearman's correlation coefficients were calculated against all other genes in the selected gene expression profile, and the *P*-value <.05 indicated the statistical significance. Also, in cBioPortal mutual exclusivity analyses were conducted for ER, PR, and HER2, where Fisher's exact test was used to analyze whether alterations are significantly mutually exclusive or co-occurring between every pair of query genes, and a *P*-value <.05 was considered significant.

Survival analysis: In this study, survival analysis was conducted using the Kaplan-Meier estimators for overall survival (OS), disease-specific survival (DSS), progression-free survival (PFS), and disease-free survival (DFS) on the Breast Invasive Carcinoma dataset (TCGA, Pan-Cancer Atlas) from cBioPortal.³⁰⁻³⁶ The dataset was filtered by the immune histochemistry (IHC) status of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), and survival analyses were compared between BC subtypes. The statistical significance between the survival curves was analyzed using the log-rank test, and a *P*-value <.05 was considered statistically significant.

Results and Discussion

Anthropometric and clinicopathological data analysis of Bangladeshi patients: The results of ANOVA analysis data of Bangladeshi (BD) breast cancer patients indicated that the risk factors like age (≥ 45 years; *P*-value <.001), height (≤ 1.5 m; *P*-value <.001), and blood pressure (systolic >120 mmHg, *P*-value .030; diastolic >80 mmHg; *P*-value .006) were significantly associated with breast cancer (Table 1). Among the included study participants a significantly higher number of breast cancer cases were with ≥ 45 years of age (60%), ≤ 1.5 m in height (60%), systolic blood pressure >120 mmHg (44%), and diastolic blood pressure >80 mmHg (44%) compared to their respective controls (Table 1). No significant associations were found between breast cancer progression and weight, BMI, waist circumference, WHR, RBS, HbA1c, and creatinine levels (Table 1).

Table 1. Anthropometric and clinicopathological characteristics data of the Bangladeshi study participants.

VARIABLES	REFERENCE VALUE	CASE N (%)	CONTROL N (%)	P-VALUE
Age (y)		46.12 ± 10.22	36.49 ± 11.11	<.001
	≥45	15 (60)	06 (24)	.01
	<45	10 (40)	19 (76)	
Height (m)		01.50 ± 00.08	01.57 ± 0.07	<.001
	>1.5	10 (40)	20 (80)	
	≤1.5	15 (60)	05 (20)	.004
Weight (kg)		57.55 ± 12.34	61.84 ± 08.87	.076
	>59	11 (44)	15 (60)	.258
	≤59	14 (56)	10 (40)	
Body mass index (BMI) (kg/m ²)		25.45 ± 04.97	25.19 ± 03.14	.766
	≥30	04 (16)	01 (04)	.157
	<30	21 (84)	24 (96)	
Waist-to-hip ratio (WHR)		00.94 ± 00.08	00.90 ± 00.05	.259
	≥0.85	22 (88)	23 (92)	.637
	<0.85	03 (12)	02 (08)	
Waist circumference (cm)		84.66 ± 09.87	87.25 ± 12.57	.361
	≥80	18 (72)	19 (76)	.747
	<80	07 (28)	06 (24)	
Blood pressure (Systolic) (mmHg)		119.16 ± 21.97	113.45 ± 09.89	.030
	>120	11 (44)	03 (12)	.012
	≤120	14 (56)	22 (88)	
Blood pressure (Diastolic) (mmHg)		81.32 ± 15.97	74.05 ± 07.64	.006
	>80	11 (44)	02 (08)	.004
	≤80	14 (56)	23 (92)	
Random blood glucose (RBS) (mmol/dL)		05.87 ± 01.21	05.32 ± 01.49	.108
	≥7.8	01 (04)	02 (08)	.552
	<7.8	24 (96)	23 (92)	
HbA1c (%)		05.90 ± 00.67	05.81 ± 00.87	.638
	>6.3	07 (28)	04 (16)	.306
	≤6.3	18 (72)	21 (84)	
Creatinine (mg/dL)		00.69 ± 00.15	00.75 ± 00.16	.096
	<0.6 & >1.2	06 (24)	04 (16)	.479
	0.6-1.2	19 (76)	21 (84)	

Data are expressed as number (%), mean ± standard deviation (SD). *P*-value was determined by the Chi-square test and ANOVA where appropriate. *P*-value <.05 is considered significant and marked in bold.

Table 2. Study characteristics of Bangladeshi breast cancer patients' data.

VARIABLE	CATEGORIES	CASE N (%)	P-VALUE
Age at diagnosis		45.25 ± 08.66	
	<40	06 (25)	
	≥40	18 (75)	.014
Cancer stage	II	11 (61)	.346
	III	07 (39)	
Cancer type	TNBC	09 (41)	.216
	Invasive ductal cell carcinoma	06 (27)	
	Ductal cell carcinoma	07 (32)	
Diagnosis			
Physical	Yes	11 (46)	.683
	No	13 (54)	
Mammography	Yes	05 (21)	.004
	No	19 (79)	
FNAC	Yes	11 (46)	.683
	No	13 (54)	
Pathology	Yes	07 (29)	.041
	No	17 (71)	
Histopathology	Yes	19 (79)	.004
	No	05 (21)	
Treatment			
Oral	Yes	16 (67)	.102
	No	08 (33)	
Hormone	Yes	00 (00)	<.001
	No	24 (100)	
Radiotherapy	Yes	08 (33)	.102
	No	16 (67)	
Chemotherapy	Yes	19 (79)	.004
	No	05 (21)	
Surgery	Yes	05 (20)	.003
	No	20 (80)	

Data are expressed as number (%), mean ± standard deviation (SD). *P*-value was determined by the Chi-square test. *P*-value <.05 is considered significant and marked in bold. The zero was adjusted.

Diagnosis, treatment, of BD breast cancer subtype: The Chi-square test results on the collected data indicated that for most of the BD breast cancer (BC) cases the age at diagnosis was ≥40 years (*P*-value .014), had a histopathological diagnosis (*P*-value .004) but had no records/history of mammography (*P*-value .004) or other pathological tests (*P*-value .041)

(Table 2). For treatment, most of the BC patients had chemotherapy (*P*-value .004), and had not undergone surgery (*P*-value .003) or had hormone therapy (*P*-value <.001) (Table 2). Moreover, the analysis of Bangladeshi breast cancer patients' data demonstrated that more than 60% of study participants were of stage II cancer and about 40% were of TNBC. The



Figure 1. Flow chart of the data screening, selection, and retrieval of breast invasive carcinoma from TCGA, Pan-Cancer Atlas.

frequency of cases who were diagnosed by physical examination, mammography, FNAC, pathological and histopathological tests were as 46%, 21%, 46%, 29%, and 79% respectively. The percentage of patients who took medication orally, radiotherapy, chemotherapy, and surgery were as 67%, 33%, 79%, and 20% respectively. None of the BC study participants had any history of hormone therapy (Table 2).

Cancer informatics analysis: For cancer bioinformatics, multi-omics data analysis was performed on the TCGA Pan-Cancer dataset after screening and filtering various datasets from TCGA, and cBioPortal (Figure 1). Finally, a dataset comprising 775 samples of breast invasive carcinoma patients based on subtypes, and race ($n=775$; TCGA, Pan-Cancer Atlas) was included for further analysis (Figure 1). The included dataset comprised of 3 different ethnicities/race categories: Caucasians ($n=585$); Africans ($n=137$) and Asians ($n=53$) (Figure 1). The Caucasian BC subtypes included samples Luminal A ($n=343$), Luminal B ($n=112$), HER2 ($n=35$), and Basal/Triple Negative (TNBC) ($n=95$). Similarly for the African BC

subtype samples of Luminal A ($n=50$), Luminal B ($n=24$), HER2 ($n=15$), TNBC ($n=48$); and the Asian BC subtype samples of Luminal A ($n=18$), Luminal B ($n=14$), HER2 ($n=15$) and TNBC ($n=6$) included for further analysis (Supplemental File S1).

Fraction genome alteration and mutation count analysis of the BC subtypes in different races

The mRNA expression data from cBioPortal of Breast Invasive Carcinoma ($n=775$), stratified by different ethnicity/race categories, were analyzed for frequency of gene alterations (including mutations, deletions, copy number gains, and amplifications) in the hormone receptor genes of ER (*ESR1*), PR (*PGR*), and the epidermal growth factor receptor gene HER2 (*ERBB2*). The scattered plot of the Spearman's and Pearson's correlation analyzed data indicated significant fractions of genomes alteration for the ER, PR, and HER2 genes (Spearman's P -value $4.99e-35$; Pearson's P -value $3.79e-18$)

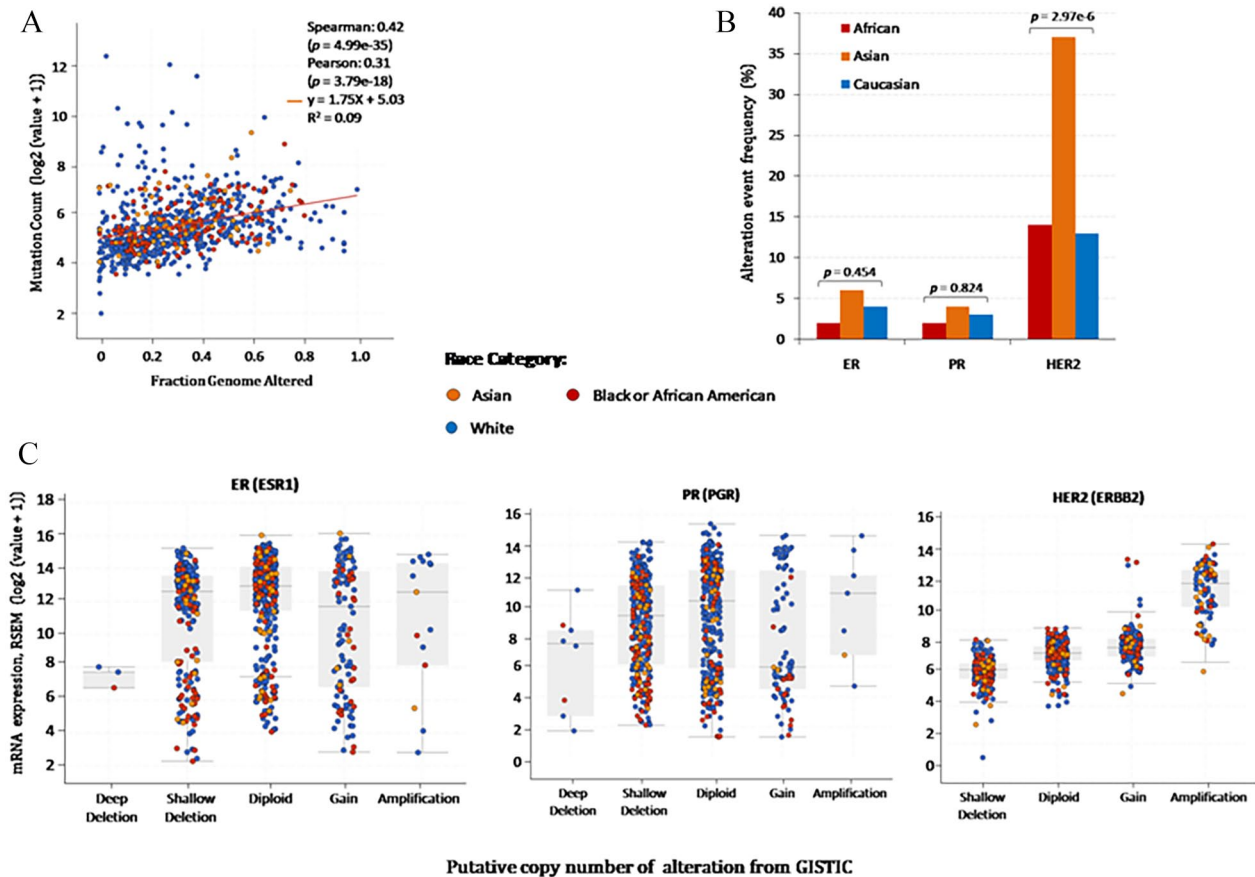


Figure 2. ER, PR, and HER2 mRNA expression data stratified by different ethnicity/race categories. (A) Fraction genome altered versus mutation count correlation plot. (B) Bar plot of alteration event frequency (%) for the genetic subtypes of ER, PR, and HER2 receptors in different races. (C) Putative copy number alterations from GISTIC.

across all ethnicity (Figure 2A). The Chi-squared analysis of the sample-level enriched data of ER, PR, and HER2 receptors suggested a strong correlation between mutation count and fraction genome alteration in the race category. Moreover, the bar plot of alteration event frequency in hormone receptors for different races and results of alteration event frequency for genetic subtype data of ER, PR, and HER2—receptors showed that the alteration event frequency (%) of HER2 was significantly higher in BC across all races than in ER and PR (Figure 2B). Comparing across different races, the HER2 data indicated significantly higher alteration frequency in Asians (P -value $2.97e-6$) than in Caucasians or Africans (Figure 2B). Alterations in ER and PR were also higher in Asian than in other races but was not significant (Figure 2B). Also, the box plot of copy number alterations (CNAs), versus mRNA expression for hormone receptors among different races analyzed data indicated amplification/mutation events in these genes. The CNAs for hormone receptors from GISTIC indicated higher amplification of HER2 in BC across all races than ER and PR (Figure 2C). For these 3 receptors, diploid and shallow deletion were most common for ER (*ESR1*), and PR (*PGR*) whereas HER2 (*ERBB2*) had more gain or amplification in BC samples (Figure 2C).

Breast cancer molecular subtype analysis: The stacked bar plot analysis of fraction genome altered data stratified by breast cancer subtypes suggested a higher percentage of Luminal A (58.3%) in the Caucasian population whereas Luminal B (24.3%) and HER2 (25.2%) subtypes were more in the Asian population and a higher frequency of TNBC (36.0%) was found in the African population (Figure 3A) (Table 3A). Also, the stacked bar plot analyzed data indicated that the highest number of fraction genome alterations was found in basal/TNBC subtype among all race groups, and predominantly higher in Caucasian and African ethnicity based on the median value. Similarly, the lowest fraction of genome alterations was detected in the Luminal A breast cancer subtype patients of all races and mostly in Caucasian ethnicity (Figure 3A) (Supplemental File S2). However, the included Vietnamese population analyzed data indicated a higher frequency of Luminal A (32.5%) subtype and the Bangladeshi BC participants' data in this present study showed a significantly higher frequency of TNBC (52.2%) subtype (P -value $<.001$) (Table 3A). When the BD population data were compared with the Asian population, the results indicated significantly higher Luminal A frequency in the Asians (P -value $.001$) than in the Bangladeshi population. Further, the expression data of

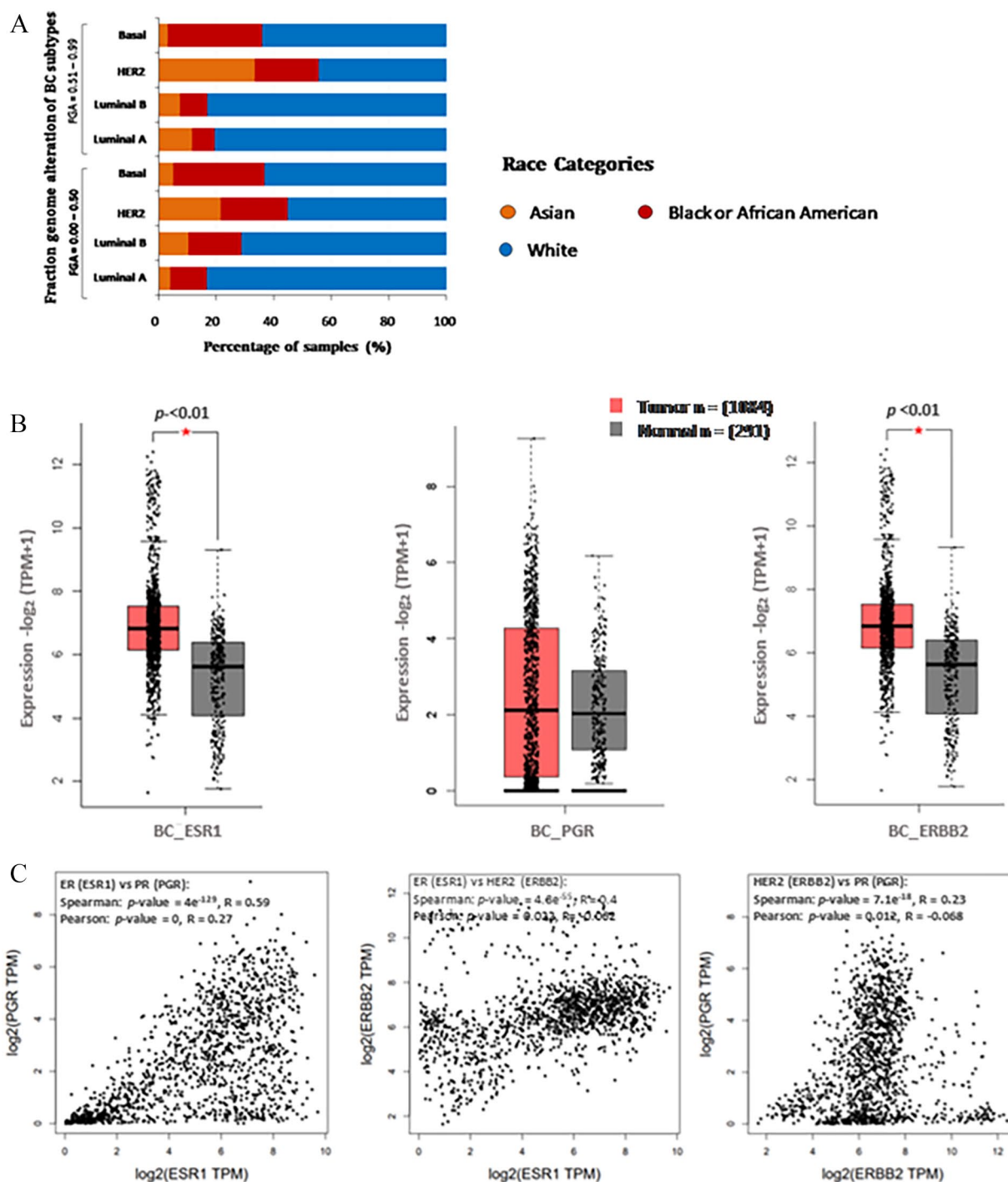


Figure 3. ER, PR, and HER2 expression analysis in TCGA Tumor/Normal GTEx expression datasets. (A) Fraction genome altered in different races by BC subtypes, (B) compared the expression status of ER, PR, and HER2 in the TCGA breast cancer (BC) dataset with the corresponding normal tissues of the GTEx datasets, and (C) correlation analysis plot of ER, PR, and HER2 expression in TCGA. A P-value $< .05$ was considered significant.

BC subtypes of the Caucasian, African, and Asian races were analyzed for mutual exclusivity based on hormone receptors ER (*ESR1*), PR (*PGR*), and epidermal growth factor receptor HER2 (*ERBB2*). The summary statistics on mutual exclusivity and co-occurrence of each pair of these receptors suggested that ER had a greater tendency toward co-occurrence with both PR and HER2 (*ESR1/PGR*: P-value .049 and *ESR1/ERBB2*: P-value .052), while PR and HER2 were mutually

exclusive (Table 3B). This study's breast cancer subtype analysis results align line with the observations of previous similar studies in different populations.³⁸⁻⁴¹

To get further insights into the up-regulated and amplified genes, we compared and evaluated the expression data of ER, PR, and HER2 between the cases and controls. The box plot of tumor/normal expression analyses of the TCGA GTEx dataset of the hormonal regulator genes ER (*ESR1*), PR (*PGR*),

Table 3. (A) Breast cancer subtypes in different ethnicity and Bangladeshi women. (B) Mutual exclusivity data between ER, PR, and HER2 receptors.

A. BREAST CANCER SUBTYPES		CBIOPORTAL DATASET ^{7,9}					CLINICAL DATA										
		CAUCASIAN (CC) N (%) ^{8,9}	AFRICAN (AF) N (%) ^{8,9}	ASIAN (AS) N (%) ^{7,9}	VIETNAM (VTN) (%) ⁷	BANGLADESH (BD) N (%)	P-VALUE (AS VS BD)										
Luminal A (ER+ IPR+ IHER2-)		420 (58.3)	61 (37.9)	41 (35.7)	13 (32.5)	03 (13.0)	0.001										
Luminal B (ER+ IPR- IHER2+/-)		139 (19.2)	26 (16.2)	28 (24.3)	09 (22.5)	04 (17.4)	0.285										
HER2 enriched (ER- IPR- IHER2+)		44 (06.1)	16 (09.9)	29 (25.2)	10 (25)	04 (17.4)	0.232										
TNBC/Basal (ER- IPR- IHER2-)		118 (16.4)	58 (36.0)	17 (14.8)	08 (20)	12 (52.2)	<0.001										
B. HORMONE RECEPTORS OF ER, PR, AND HER2																	
HORMONE RECEPTORS		NEITHER		A NOT		B NOT		BOTH		LOG2 ODDS RATIO		P-VALUE		Q-VALUE		TENDENCY	
A	B	A OR B		A	B	A	B	A	B								
ESR1	PGR	726		27	19	03	03	2.086		.049	0.078						Co-occurrence
ESR1	ERBB2	642		22	103	08	08	1.180		.052	0.078						Co-occurrence
PGR	ERBB2	645		19	108	03	03	-0.085		.611	0.611						Mutual exclusivity

The P-value was calculated using the chi-square test. P-value < .05 was considered significant. Log2 Odds ratio (OR) was calculated as, $OR = (Neither \times Both) / (A \text{ Not } B \times B \text{ Not } A)$ where a positive value indicated co-occurrence and a negative value indicated mutual exclusivity. The P-value was calculated by the one-sided Fisher's exact test and the q-value was calculated using Benjamini-Hochberg FDR (False discovery rate) correction procedure. A P-value < .05 and a q-value < .05 were considered significant.

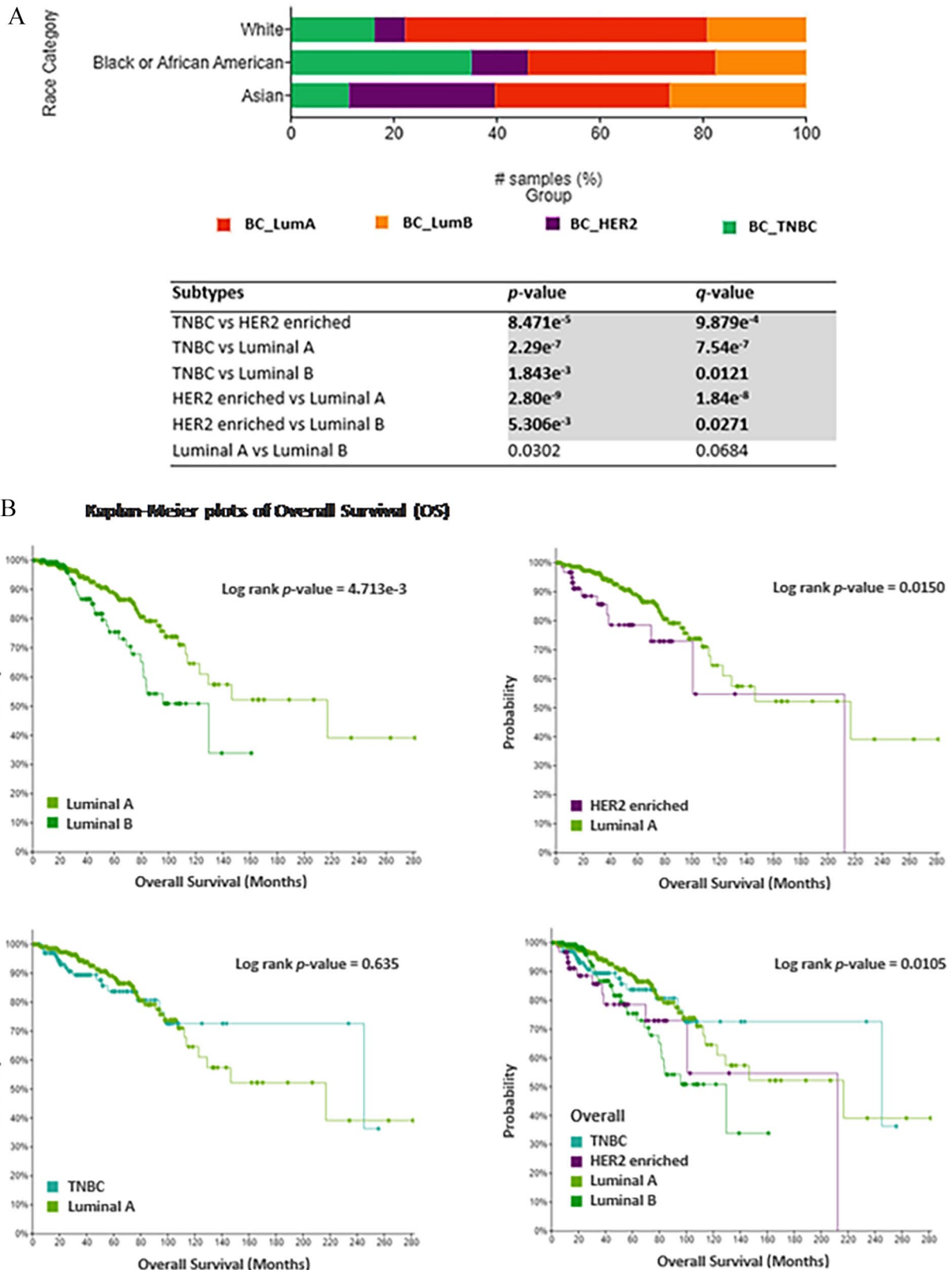


Figure 4. Survival analysis outcome in patients by breast cancer (BC) subtypes. (A) The stack bar plot of breast cancer (BC) subtypes expression data in different races with *P* and *q* values. (B) Kaplan-Meier plots of overall survival (OS) difference in patients by cBioPortal. Note that the different survival characteristics were of different datasets.

and the epidermal growth factor receptor HER2 (*ERBB2*) between normal healthy breast tissue samples (controls, *n* = 291) and breast invasive carcinoma samples (cases, *n* = 1084)

analyzed data indicated significant up-regulation of ER (*ESR1*, *P*-value < .01) and HER2 (*ERBB2*, *P*-value < .01) in breast cancer patients compared to the healthy controls (Figure 3B).

Table 4. Survival analysis data of breast cancer subtypes by cBioPortal.

SUBTYPES	TOTAL NUMBER OF CASES, N (%)	NUMBER OF EVENTS N (%)	MEDIAN MONTHS OVERALL (95% CI)
TNBC/Basal	149 (19.2)	18 (12.1)	245.09 (245.09-NA)
HER2 enriched	65 (08.4)	12 (18.5)	212.25 (100.70-NA)
Luminal A	411 (53.0)	42 (10.2)	216.75 (122.83-NA)
Luminal B	150 (19.4)	27 (18.0)	129.57 (81.63-NA)

In addition, the correlation analysis of these receptors' ER, PR, and HER2 expression was performed. The analyzed data indicated a moderate positive correlation between ER and PR (Spearman's: $R .59$, P -value $4e-129$; Pearson's: $R .27$, P -value 0). However a moderate to low negative correlation was observed between ER and HER2 (Spearman's: $R 0$, P -value $4.8e-55$; Pearson's: $R -.062$, P -value $.022$) and between HER2 and PR (Spearman's: $R .23$, P -value $7.1e-18$; Pearson's: $R -.068$, P -value $.012$) (Figure 3C).

Prevalence of BC subtypes in different races and survival analysis: The breast cancer subtype expression data of the TCGA, Pan-Cancer Atlas dataset were analyzed and the stacked bar plots of BC subtypes protein expression in different races with the percentage of samples between subtypes data also indicated that the Luminal A subtype (ER+, PR+, and HER2-) was more prevalent in the Caucasian, whereas the Luminal B (ER+, PR- and HER2+/-) and HER2 enriched (ER-, PR-, and HER2+) subtypes were more prevalent in the Asians (Table 3A) (Figure 4A). The TNBC (ER-, PR-, and HER2-) subtype was found more prevalent in Africans (Table 3A) (Figure 4A). In addition, the comparison between breast cancer subtypes data indicated that most of the pair comparisons were significant P -values except for the pair (Luminal A vs Luminal B) (Figure 4A).

Finally, to evaluate the clinical outcome with patient prognosis, the Kaplan-Meier plots of survival analyses of overall survival (OS) differences in patients of BC subtypes compared with Luminal A indicated the lower median months (95% CI) for patients with Luminal B [95% CI: 129.57 (81.63-NA), log-rank P -value $4.713e-3$] and HER2 enriched [95% CI: 212.25 (100.70-NA), log-rank P -value $.015$] BC subtypes and overall lower survival rate against Luminal A in log-rank test (Figure 4B) (Table 4). Although the KM plot of TNBC did not show significant change from Luminal A, the number of events with median months (95% CI) indicated Luminal A had a lower number of events (10.2%) than TNBC (12.1%), Luminal B (18%), and HER2 (18.5%) (Table 4). Results indicated that Luminal A with overall survival duration (months) was significantly higher in comparison to other subtypes. The median months of overall survival analysis were lowest for Luminal B and then for HER2 which suggested that these subtypes were more aggressive and dominant in the Asian population. These results support and are in line with the findings

of some previous studies.⁴²⁻⁴⁴ KM plot of univariate analysis for OS showed significant differences in survival between the BC molecular subtypes (log-rank P -value $.0105$) (Figure 4B).

Evaluation of ER, PR, and HER2 receptor biomarkers in Asian ethnicity: The vital information of each patient's clinical data was extracted from TCGA and cBioPortal. Immunohistochemistry (IHC)-based receptor status, from Cancer Digital Slide Archive⁴⁵ for BC patients indicated differentiated expression pattern staining of Luminal A, Luminal B, HER2 enriched, and TNBC (Figure 5A) (Supplemental File 3). Further, the Asian patients' data on hormone receptors ER, PR, and the epidermal growth receptor HER2 were analyzed. The correlation analysis between mRNA versus protein expression levels of ER, PR, and HER2 demonstrated a strong correlation determined by the Spearman's and Pearson's correlation coefficients [ER (Spearman's: P -value $4.76e-22$, $R .92$ Pearson's: P -value $2.36e-21$, $R .92$); PR (Spearman's: P -value $4.98e-11$, $R .76$; Pearson's: P -value $4.41e-11$, $R .76$); and HER2 (Spearman's: P -value $4.81e-16$, $R .86$; Pearson's: P -value $2.66e-17$, $R .87$)] in breast cancer subtypes (Figure 5B). Also, results revealed that protein expression and receptor status were significantly correlated with PR (P -value $1.587e-6$) and HER2 (P -value $5.885e-3$) (Supplemental File 4).

Finally, we assessed the copy number alteration versus mRNA expression for the ER, PR, and HER2 receptors, their 3D structure, and mutation/variation data (SNPs) in Asian population samples from cBioPortal. The results of CNA versus mRNA expression of the BC subtype in the Asian population indicated that the mRNA expression of ER and PR was increased and showed shallow deletion while HER2 receptor expression showed gain and amplification (Figure 6A). The 3D structure of ER, PR, and HER2 with mutations and their single nucleotide polymorphisms (SNPs) were evaluated to assess amino acid changes in particular positions for hormone receptors in BC patients (Figure 6B) (Supplemental File 5).

The overall implication of the results: Breast cancer is one of the most occurring cancer types among women all over the world. In this study, we evaluated that age, height, and blood pressure were significant risk factors for breast cancer development in Bangladeshi (BD) women (Table 1) and, most of the BD breast cancer patients only had chemotherapy treatment (Table 2). The BD breast cancer patients' data demonstrated that more than 60% of the study participants were of stage II

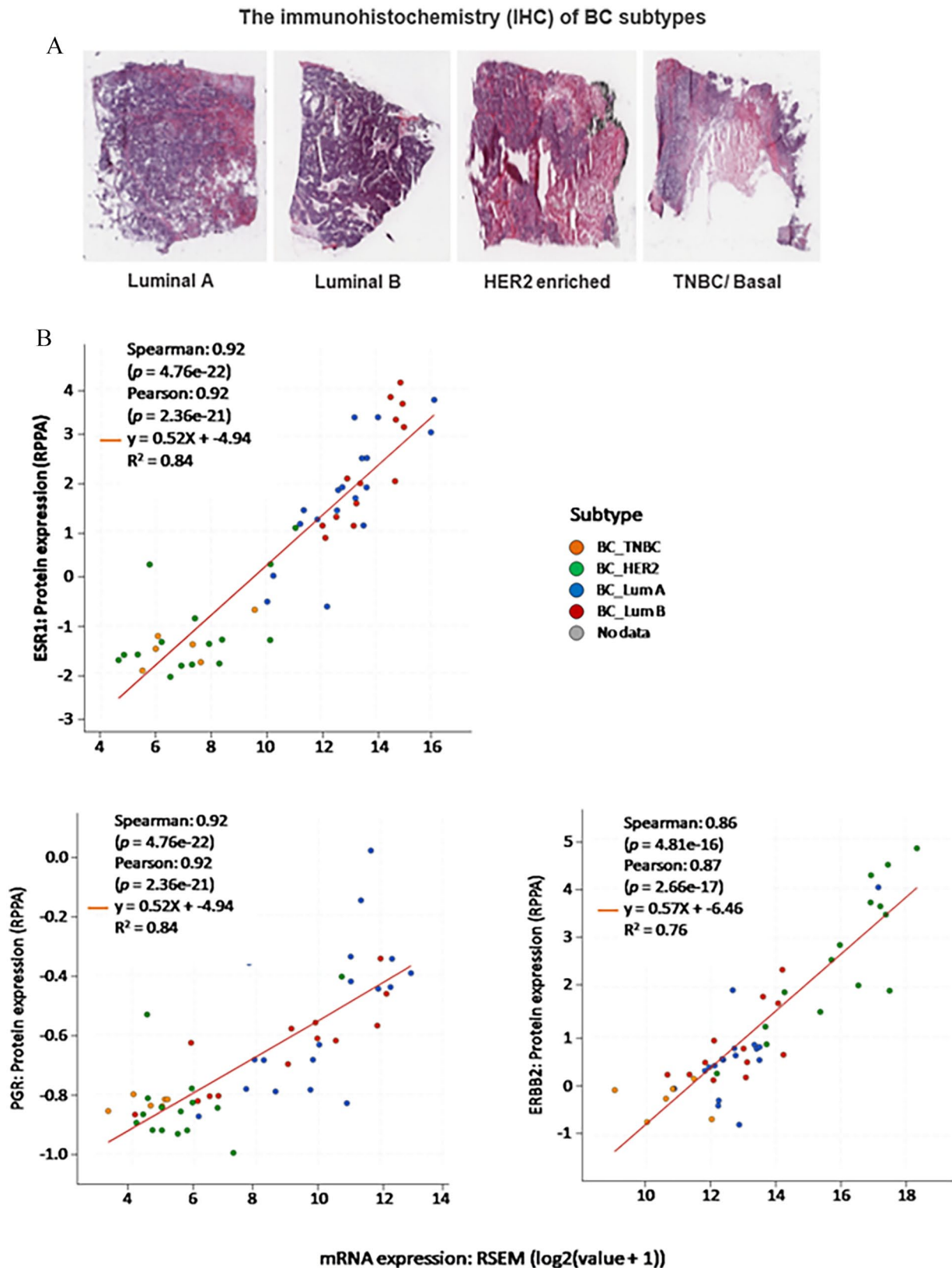


Figure 5. (A) Tissue image differences in patients with different subtypes by immunohistochemistry (IHC). (B) Correlation analysis between mRNA expression, RSEM and protein expression, RPPA data of ER, PR, and HER2 receptors in Asian ethnicity.

cancer and 40% were of TNBC. To provide a broader view of the same among different ethnicities breast cancer datasets from cBioportal were analyzed for BC subtypes. The fraction genome alteration of ER (*ESR1*), PR (*PGR*), and HER2

(*ERBB2*) suggested a strong correlation between mutation count and fraction genome alteration in different ethnicities and significantly higher alteration frequency was observed in the Asian population with HER2 subtype (Figure 2). The

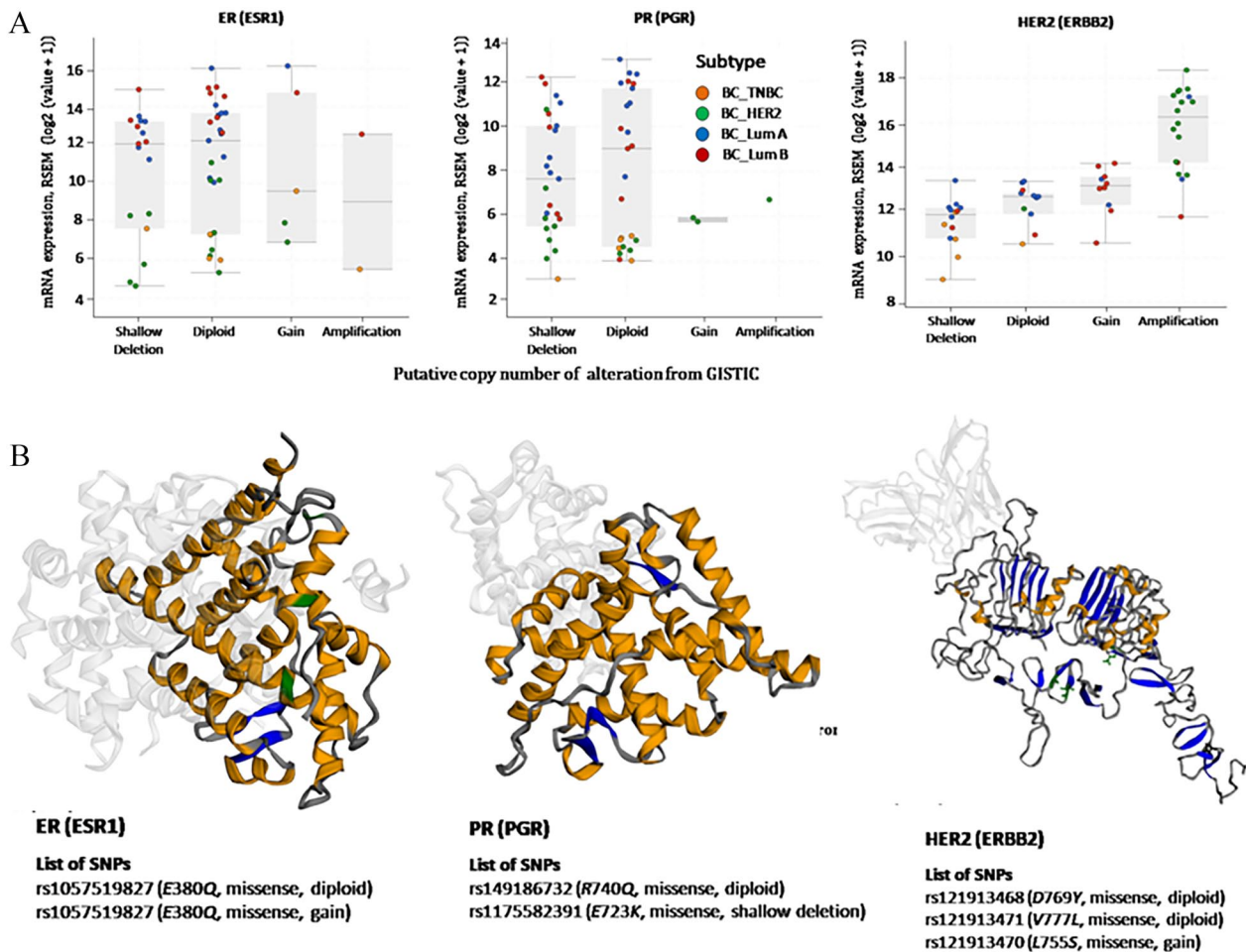


Figure 6. (A) Box plot for putative copy number alteration (CNA) versus mRNA expression of ER, PR, and HER2 in Asian ethnicity from GISTIC. (B) 3D structure of ER, PR, and HER2 with mutations (Green: missense; Blue: truncated; Brown: inframe; Orange: spliced; and Pink: fusion) and some single nucleotide polymorphisms (SNPs).

TCGA expression data confirmed significant up-regulation of ER, and HER2 in breast cancer patients in comparison to their respective control (Figure 3B) and a significant positive correlation was observed between (ER and PR). However, a significant negative correlation was observed between (ER and HER2), and (PR and HER2) (Figure 3C). The correlation analysis between mRNA and Protein expression presented that ER, PR, and HER2 receptors had a strong correlation with different breast cancer subtypes (Figure 5B). The CNA versus mRNA expression analysis suggested that the HER2 receptor gained or amplified markedly in the Asian population whereas the ER and PR hormone receptors showed shallow deletion (Figure 6A). Results from various bioinformatics and statistical analysis revealed that the highest percentage of Luminal A, was prevalent in Caucasians whereas Luminal B, and HER2 enriched subtypes were more prevalent in Asians and the TNBC/basal subtype was found most prevalent in Africans. However, in Bangladeshi women, the TNBC subtype was found more prevalent (Table 3A). The KM plots of the overall survival of different BC subtypes indicated that the Luminal A subtype was significantly higher in comparison to other subtypes (Figure 4B). The median months of overall survival were lowest for Luminal B and HER2, the 2 most dominant

subtypes in Asians indicating more aggressive and dominant subtypes. Thus, results from this study depicted the molecular signature and functional importance of breast cancer subtypes in different ethnicities and their implications for treatment and diagnosis purposes.

Conclusion

In this study, we used a comprehensive computational approach, to evaluate genomic alterations in tumor samples and assess their functional importance in breast cancer subtypes. Results from this study indicated a significant association between BC subtypes and ethnicity. Taken together, all data from literature mining, statistical, and cancer informatics analysis indicated that Luminal B and HER2 enriched, the 2 more aggressive BC subtypes were most prevalent in Asians compared to Caucasians. However, the Bangladeshi BC patients' data indicated a significantly higher frequency of TNBC and lower Luminal A subtypes in comparison to the Asian population. This could be due to the small sample size and further investigations with a larger dataset are crucial to validate breast cancer subtype association. Nevertheless, the initial findings from this study suggested a difference in the prevalence of BC subtypes among Bangladeshi women. In the future, conducting more

population-based studies of larger sample sizes would provide further insights into genetic risk factors and evidence regarding BC subtypes in Asian and Bangladeshi women. Major global improvements in breast cancer can result from implementing what we already know will work. And early detection and treatment have proven successful in high-income countries and should be applied in countries with limited resources like Bangladesh. Thus, by understanding the importance of early detection and treatment, more women would be encouraged to consult medical practitioners when breast cancer is first suspected, and before any further cancer progression. We believe results from this study might aid in the prevention, management, and raising of awareness against the specific risk factors among Bangladeshi women in near future.

Authors' Contributions

Jesmin conceptually designed the research; MSI, SID, and DI collected data and screened data; MSI and DI analyzed the data; Jesmin, MSI, and DI validate the interpretation and drafted and revised the manuscript. All authors have read and approved the final manuscript.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- World Health Organization (WHO). Global cancer observatory, cancer today, graph: pie chart of estimated incidences in 2020 for cancer sites. Accessed 19 July 2022, Available at: <https://gco.iarc.fr/today/online-analysis-pie>
- World Health Organization (WHO). Fact sheets, details about breast cancer. Accessed 21 July 2022, Available at: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>
- World Health Organization (WHO). Global cancer observatory, cancer today, table: estimated number of cases in 2020 worldwide. Accessed 20 July 2022, Available at: <https://gco.iarc.fr/today/online-analysis-table?>
- World Health Organization (WHO). Global cancer observatory, cancer today, graph: estimated number of cases in 2020. Accessed 21 July 2022, Available at: <https://gco.iarc.fr/today/online-analysis-multi-bars?>
- WHO.int. Breast cancer. 2021. Accessed 22 July 2022, Available at: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>
- Ginsburg O, Yip C, Brooks A, et al. Breast cancer early detection: a phased approach to implementation. *Cancer*. 2020;126:2379-2393.
- Mutebi M, Anderson BO, Duggan C, et al. Breast cancer treatment: a phased approach to implementation. *Cancer*. 2020;126:2365-2378.
- Chowdhury FA, Islam MF, Prova MT, et al. Association of hyperlipidemia with breast cancer in Bangladeshi women. *Lipids Health Dis*. 2021;20:52.
- Kamińska M, Ciszewski T, Łopacka-Szatan K, Miotła P, Starosławska E. Breast cancer risk factors. *Prz Menopauzalny*. 2015;3:196-202.
- Yuan JM, Yu MC, Ross RK, Gao YT, Henderson BE. Risk factors for breast cancer in Chinese women in Shanghai. *Cancer Res*. 1988;48:1949-1953.
- Hadjisavvas A, Loizidou MA, Middleton N, et al. An investigation of breast cancer risk factors in Cyprus: a case control study. *BMC Cancer*. 2010;10:447.
- Tran B, Bedard PL. Luminal-B breast cancer and novel therapeutic targets. *Breast Cancer Res*. 2011;13:221.
- Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast*. 2015;24:S26-S35.
- Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160-1167.
- Li ZH, Hu PH, Tu JH, Yu NS. Luminal B breast cancer: patterns of recurrence and clinical outcome. *Oncotarget*. 2016;7:65024-65033.
- Parise CA, Bauer KR, Caggiano V. Variation in breast cancer subtypes with age and race/ethnicity. *Crit Rev Oncol Hematol*. 2010;76:44-52.
- Telli ML, Chang ET, Kurian AW, et al. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. *Breast Cancer Res Treat*. 2011;127:471-478.
- Satagopan JM, Stroup A, Kinney AY, Dharamdasani T, Ganesan S, Bandera EV. Breast cancer among Asian Indian and Pakistani Americans: a surveillance, epidemiology and end results-based study. *Int J Cancer*. 2021;148:1598-1607.
- cBioPortal For Cancer Genomics. cBioPortal invasive breast cancer dataset containing information about race and subtypes of breast cancer (Broad, Nature 2012). Accessed 23 July 2022, Available at: <https://www.cbioportal.org/study/summary?>
- Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401-404.
- Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioportal. *Sci Signal*. 2013;6:1.
- World Health Organization (WHO). Global cancer observatory, cancer tomorrow, graphs: bar chart of breast cancer incidence and mortality for females in continents, in Asian countries, and Bangladesh. Accessed 24 July 2022, Available at: <https://gco.iarc.fr/tomorrow/en/dataviz/bars?>
- cBioPortal. User guide, how to extract data from query using gene, "Query By Gene". Accessed 25 July 2022, Available at: <https://raw.githubusercontent.com/cBioPortal/cbioportal/master/docs/tutorials/cBioPortal%20Tutorial%202%20Single%20Study%20Query.pdf>
- cBioPortal. User guide, how to compare between groups of samples or patients, "Group comparison". Accessed 26 July 2022, Available at: <https://www.cbioportal.org/tutorials#group-comparison>
- cBioPortal. . User guide, data extraction about co-expression of genes from cBioportal. Accessed 27 July 2022, Available at: <https://raw.githubusercontent.com/cBioPortal/cbioportal/master/docs/tutorials/cBioPortal%20Tutorial%202%20Single%20Study%20Query.pdf>
- The Cancer Genome Atlas Program. TCGA Research Network, 2018 Accessed 28 July 2022, Available at: <https://www.cancer.gov/tcga>
- James PT. Obesity: the worldwide epidemic. *Clin Dermatol*. 2004;22:276-280.
- cBioPortal, multi-omics data of The Cancer Genome Atlas (TCGA, Pan-Cancer), 2018. Accessed May 13, 2021 and February 21, 2022, Available at: <https://www.cbioportal.org>.
- Weinstein JN, Collisson EA, Mills GB, et al.; Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45:1113-1120.
- cBioPortal Invasive Breast Cancer. Dataset containing information of race and subtypes of breast cancer (TCGA, PanCancer Atlas), 2018. Accessed 30 July 2022, Available at: https://www.cbioportal.org/study/summary?id=brca_tcga_pan_can_atlas_2018.
- Hoadley KA, Yau C, Hinoue T, et al. Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*. 2018;173:291-304.e6.
- Gao Q, Liang WW, Foltz SM, et al. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep*. 2018;23:227-238.e3.
- Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in the Cancer Genome Atlas. *Cell*. 2018;173:321-337.e10.
- Liu J, Lichtenberg T, Hoadley KA, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018;173:400-416.e11.
- Taylor AM, Shih J, Ha G, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*. 2018;33:676-689.e3.
- Ellrott K, Bailey MH, Saksena G, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst*. 2018;6:271-281.e7.
- Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*. 2019;47:W556-W560.
- Li CI, Malone KE, Daling JR. Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. *Cancer Epidemiol Biomarkers Prev*. 2002;11:601-607.
- Troester MA, Sun X, Allott EH, et al. Racial differences in PAM50 subtypes in the Carolina Breast Cancer Study. *J Natl Cancer Inst*. 2018;110:176-182.
- Holowatyj AN, Ruterbusch JJ, Ratnam M, Gorski DH, Cote ML. HER2 status and disparities in luminal breast cancers. *Cancer Med*. 2016;5:2109-2116.
- Hausauer AK, Keegan TH, Chang ET, Clarke CA. Recent breast cancer trends among Asian/Pacific Islander, Hispanic, and African-American women in the US: changes by tumor subtype. *Breast Cancer Res*. 2007;9:R90.
- Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295:2492-2502.
- O'Brien KM, Cole SR, Tse CK, et al. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res*. 2010;16:6100-6110.
- Kurian AW, Fish K, Shema SJ, Clarke CA. Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups. *Breast Cancer Res*. 2010;12:R99.
- Cancer Digital Slide Archive. Retrieving immunohistochemistry stain of breast cancer subtypes. Accessed 31 July 2022, Available at: <https://cancer.digitalslidearchive.org/>