Am J Transl Res 2023;15(10):6058-6070
www.ajtr.org /ISSN:1943-8141/AJTR0142850

Original Article

Comprehensive multi-level expression profiling of key biomarkers in breast cancer patients

Tayyaba Zia[1], Liu Bangfan[2], Amun Nadeem[3], Abid Hussain[4], Mostafa A Abdel-Maksoud[5], Adel M Zakri[6], Muhammad Kazim Bashir[7], Mubarak Ali[8], Norina Jabeen[9], Muhammad Jamil[10], Wahidah H Al-Qahtani[11], Taghreed N Almanaa[5]

[1]School of Public Management, Yanshan University, Qinhuangdao 066000, Hebei, China; [2]School of Public Administration, Yanshan University, Qinhuangdao 066000, Hebei, China; [3]Department of Pathology, Gujranwala Medical College Teaching Hospital and DHQ Teaching Hospital, Gujranwala 52250, Punjab, Pakistan; [4]APMO Nishter Hospital, Multan 60000, Punjab, Pakistan; [5]Department of Botany and Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; [6]Department of Plant Production, College of Food and Agricultural Sciences, King Saud University, Saudi Arabia; [7]Shaukat Khanum Memorial Cancer Hospital and Research Centre, Lahore 13014, Punjab, Pakistan; [8]Animal Science Institute, National Agricultural Research Center, Islamabad 54000, Pakistan; [9]Department of Rural Sociology, University of Agriculture Faisalabad, Faisalabad 38000, Punjab, Pakistan; [10]PARC Arid Zone Research Centre, Dera Ismail Khan 29050, Pakistan; [11]Department of Food Sciences & Nutrition, College of Food and Agricultural Sciences, King Saud University, Saudi Arabia

Received March 13, 2022; Accepted October 11, 2023; Epub October 15, 2023; Published October 30, 2023

Abstract: Objectives: In this comprehensive breast cancer (BC) study, we aimed to identify, validate, and characterize key biomarkers with significant implications in BC diagnosis, prognosis, and as therapeutic targets. Methods: Our research strategy involved a multi-level methodology, combining bioinformatic analysis with experimental validation. Results: Initially, we conducted an extensive literature search to identify BC biomarkers, selecting those with reported accuracies exceeding 20% in specificity and sensitivity. This yielded nine candidate biomarkers, which we subsequently analyzed using Cytoscape to identify a few key biomarkers. Based on the degree method, we denoted four key biomarkers, including progesterone receptor (PGR), epidermal growth factor receptor (EGFR), estrogen receptor 1 (ESR1), and Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2). Expression analysis using The Cancer Genome Atlas (TCGA) dataset revealed that PGR and EGFR exhibited significant (p-value < 0.05) down-regulation in BC samples when compared to controls, while ESR1 and ERBB2 showed up-regulation. To strengthen our findings, we collected clinical BC tissue samples from Pakistani patients and performed expression verification using real-time quantitative polymerase chain reaction (RT-qPCR). The results aligned with our initial TCGA dataset analysis, further validating the differential expression of these key biomarkers in BC. Furthermore, we utilized receiver operating characteristic (ROC) curves to demonstrate the diagnostic use of these biomarkers. Our analysis underscored their accuracy and sensitivity as diagnostic markers for BC. Survival analysis using the Kaplan-Meier Plotter tool revealed a prognostic significance of PGR, ESR1, EGFR, and ERBB2. Their expression levels were associated with poor overall survival (OS) of BC patients, shedding light on their roles as prognostic indicators in BC. Lastly, we explored DrugBank to identify drugs that may reverse the expression patterns , and estradiol, decitabine, and carbamazepine were singled out. Conclusion: Our study gives valuable insight into BC biomarkers, for diagnosis and prognosis. These findings have implications for BC management using personalized and targeted therapeutic approaches for BC patients.

Keywords: Breast cancer, diagnosis, prognosis, biomarker

Introduction

Breast cancer (BC), a complex and heterogeneous disease, continues to be a significant global health concern [1-3]. BC prevalence has steadily increased, with millions of new cases diagnosed each year [4]. Multiple factors contribute to this rising trend, including genetic predisposition, hormonal influences, lifestyle choices, and environmental exposures [5-7]. Additionally, hormonal factors, like early menstruation, late menopause, and hormone

replacement therapy, can play a role. Lifestyle elements encompass diet, physical activity, and alcohol consumption [8-12].

Early and precise diagnosis is crucial for effective treatment and improved patient outcome [13-16]. The identification of diagnostic and prognostic biomarkers has played a pivotal role in enhancing BC detection and prognosis [17-20]. Among these biomarkers, the genes associated with hereditary BC, particularly BRCA1 and BRCA2, have been extensively studied and acknowledged as significant contributors to BC susceptibility [21-23]. Moreover, beyond BRCA1 and BRCA2 genes, various other biomarkers have also found their place in clinical practice.

Our present study aims to identify all the already-reported BC biomarkers from the literature and make a shortlist and validate a few more important biomarkers by a multi-level methodology. We started by systematically extracting widely recognized BC diagnostic biomarkers from the extensive body of scientific literature. Recognizing the intricate interplay of these diagnostic biomarkers within the biological landscape, we employed Protein-Protein Interaction (PPI) network analysis to construct an informative network. This network illuminates the molecular relationships and interactions among these biomarkers, shedding light on their collective significance for BC diagnosis.

In our pursuit of precision and selectivity, we employed the degree method, a robust network analysis technique. This method allows us to distill the complex PPI network and spotlight the four most pivotal diagnostic biomarkers for BC. For final validation of the selective four biomarkers, our research turned its focus towards The Cancer Genome Atlas (TCGA) and clinical samples from BC patients. Through these concerted efforts, we aspire to pave the way for more precise and personalized BC management strategies, ultimately improving the outlook for individuals at risk of BC.

Methods

Literature search

Information on diagnostic and prognostic biomarkers was retrieved from SciVerse Scopus® (Elsevier Properties, SA, USA), Web of Science® (Thomson Reuters, USA), and PubMed. The keywords used for the search included “Breast cancer”, “Diagnosis”, and “Prognosis”. Only those expression-based biomarkers having more than 20% accuracy in sensitivity and specificity were included in the present study.

Protein-protein interaction network construction and identification of key biomarkers

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) is a vital resource in the field of molecular biology and bioinformatics [24]. It serves as a comprehensive platform for the exploration of protein-protein interactions (PPIs). STRING compiles and integrates vast datasets, including experimental evidence, computational predictions, and curated knowledge, to provide a holistic view of protein interactions within various organisms. Researchers worldwide rely on STRING to unravel complex molecular networks, identify functional associations among proteins, and gain insight into biological processes. In the present study, this database was used to develop a PPI of the extracted biomarker genes.

Cytoscape is a powerful and widely-used software tool in the field of network biology and bioinformatics [25]. It enables researchers to visualize, analyze, and model complex biological networks, particularly focusing on molecular interaction networks such as protein-protein interactions, gene regulatory networks, and metabolic pathways. In this study, we employed this tool to conduct an analysis of PPIs and to pinpoint four key biomarkers, employing the degree method as our selection criterion.

Expression analysis of key biomarkers in The Cancer Genome Atlas datasets

UALCAN (https://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl) and GEPIA (http://gepia.cancer-pku.cn/) are invaluable resources in the field of bioinformatics, providing researchers with powerful tools to explore and analyze gene expression data in health and disease [26, 27]. These databases offer an intuitive web interface that grants researchers access to TCGA data, facilitating the exploration of gene expression profiles across various cancer types. These user-friendly platforms allow for the comparison of gene expression between tumor and normal tissues, making it a valuable asset for investigating candidate biomarkers or therapeutic targets. In the present study, these data-

bases were used to analyze the expression profiles of key biomarkers across a BC sample group and control group.

Kaplan Meier plotter

The Kaplan Meier (KM) Plotter tool (https://kmplot.com/analysis/) is a valuable resource in cancer research, specifically designed for survival analysis [28]. It harnesses extensive gene expression data from a variety of sources to provide insight into the impact of gene expression on patient survival. Researchers can easily explore the relationship between gene expression levels and survival outcomes across multiple cancer types. This tool generates Kaplan-Meier survival plots, helping scientists identify genes that may serve as prognostic indicators or therapeutic targets. In the present study, KM plotter tool was used to analyze the prognostic value of key biomarkers in BC.

Ribonucleic acid (RNA) extraction and real-time quantitative polymerase chain reaction (RT-qPCR)

In the study, total RNA extraction and qRT-PCR were carried out as follows: RNA extraction was performed using TRIzol® reagent (Ambion, USA) in accordance with the manufacturer's instructions. Subsequently, cDNA synthesis was accomplished using the PrimeScript RT reagent kit (Takara, China). The RT-qPCR analysis was conducted on an ABI 7500 RT PCR system employing the SYBR Premix Ex Taq II Kit (Takara, China). To ensure accurate quantification, all measurements were normalized to the expression level of glyceraldehyde phosphate dehydrogenase (GAPDH) within the reaction. The comparative threshold cycle (C_T) method, which involves comparing C_T values between a common reference RNA and the target gene RNA, was employed to determine the relative fold changes in gene expression. Gene expression levels were calculated using the $2^{-\Delta\Delta CT}$ method. Each experiment was replicated three times. The following primers were used to assess key biomarker genes.

GAPDH-F 5'-ACCCACTCCTCCACCTTTGAC-3', GAPDH-R 5'-CTGTTGCTGTAGCCAAATTCG-3'; PGR-F 5'-GTCGCCTTAGAAAGTGCTGTCAG-3', PGR-R 5'-GCTTGGCTTTCATTTGGAACGCC-3'; ESR1-F 5'-GCTTACTGACCAACCTGGCAGA-3', ESR1-R 5'-GGATCTCTAGCCAGGCACATTC-3'; EGFR-F 5'-AACACCCTGGTCTGGAAGTACG-3', EGFR-R 5'-TCGTTGGACAGCCTTCAAGACC-3'; ERBB2-F 5'-GGAAGTACACGATGCGGAGACT-3', ERBB2-R 5'-ACCTTCCTCAGCTCCGTCTCTT-3'.

Receiver operating characteristic (ROC) curve

The ROC curve is a comprehensive metric that amalgamates the continuous variables of sensitivity and specificity. It provides a holistic assessment of a test's diagnostic performance. Traditionally, an area under the ROC curve (AUC) surpassing 0.7 signifies the accuracy of a diagnostic test. In this study, ROC curve analysis was conducted using RT-qPCR expression data for key biomarkers.

Gene enrichment analysis

Metascape (https://metascape.org/gp/index.html) is a valuable resource for conducting functional enrichment analysis [29]. It simplifies the interpretation of extensive omics data by detecting enrichments in biological processes, pathways, and molecular function. Metascape's intuitive interface and extensive gene annotation capabilities aid researchers in uncovering the biologic relevance of their datasets. This resource proves instrumental in advancing discoveries across a spectrum of research areas, spanning genomics to proteomics. For our study, we used Metascape to perform gene enrichment analysis on the pivotal biomarker genes.

DrugBank database

The DrugBank database (https://go.drugbank.com/) is a pivotal resource in the realm of pharmacology and drug research [30]. It is a comprehensive repository of detailed information on drugs, their molecular targets, pharmacologic actions, and associated pathways. Researchers and healthcare professionals rely on DrugBank to access a vast wealth of data, including drug structures, interactions, and side effects, which aid in drug development, prescription, and patient care. In this study, we used DrugBank database to retrieve key biomarker-associated drugs.

Statistics

For enrichment analysis, we used Fisher's Exact test for computing a statistical difference. Correlational analyses were carried out using Pearson method. For comparisons, a

Table 1. Compilation of BC biomarkers extracted from the literature, meeting the criteria of specificity and sensitivity exceeding 20%

Sr. No	Biomarker	Accuracy in terms of specificity and sensitivity	Reference
1	HRE2/ERBB2	95%	[42]
2	ER/ESR1	90%	[43]
3	PGR	90%	[44]
4	EGFR	30-50%	[45]
5	CA27-29/MUC1	30-50%	[46]
6	CA-125/MUC16	30-50%	[47]
7	CEA/CEACAM5	20-50%	[48]
8	MYC	20-50%	[49]
9	CCND1	20-50%	[50]

HRE2 = Human epidermal growth factor receptor 2, ERBB2 = Erb-B2 receptor tyrosine kinase 2, ER = Estrogen receptor, ESR1 = Estrogen Receptor 1, PGR = Progesterone receptor, EGFR = Epidermal growth factor receptor, CA27-29 = Cancer antigen 27-29, MUC1 = Mucin1, CA-125 = Cancer antigen 125, MUC16 = Mucin16, CEA = Carcinoembryonic antigen, CEACAM5 = Carcinoembryonic antigen-related cell adhesion molecule 5, MYC = Myelocytomatosis, CCND1 = Cyclin D1.

Student t-test was adopted. All the analyses were carried out in R version 3.6.3 software.

Results

Retrieval of expression-based diagnostic and prognostic biomarkers from the literature

After conducting an exhaustive literature search, we extracted a total of nine expression-based biomarkers, including HRE2/ERBB2, ER/ESR1, PGR, EGFR, CA27-29/MUC1, CA27-29/MUC1, CA-125/MUC16, CEA/CEACAM5, MYC, and CCND1 that met our stringent criteria of having an accuracy exceeding 20% in both specificity and sensitivity. These biomarkers, along with their respective details, are presented in **Table 1**. Our meticulous screening process aimed to ensure that only the most reliable and clinically relevant diagnostic and prognostic biomarkers for BC were included in our study.

Protein-protein interaction (PPI) network construction and identification of key biomarkers

First, the PPIs of 9 extracted biomarkers were constructed with the help of STRING database. This comprehensive PPI network encompassed nine nodes and featured a total of 35 interactions (**Figure 1A**, **1B**). Subsequently, the meticulously constructed PPI network was imported into Cytoscape, where we conducted a detailed analysis employing the degree method. This analytical approach was employed to identify four key biomarkers, with the most crucial roles in BC. After analyzing PPIs, four biomarkers emerged as pivotal players in BC: PGR (Progesterone Receptor), ESR1 (Estrogen Receptor 1), EGFR (Epidermal Growth Factor Receptor), and Erb-B2 Receptor Tyrosine Kinase 2 (ERBB2) (**Figure 1B**). These biomarkers were identified based on their significant connectivity within the PPI network, signifying their central roles in the intricate landscape of BC.

Expression analysis of key biomarkers in The Cancer Genome Atlas datasets

We used the UALCAN and GEPIA databases to scrutinize the expression profiles of PGR, ESR1, EGFR, and ERBB2 in the BC sample group and the control group. The results of expression analysis by UALCAN showed significant alterations in gene expression levels between these two groups. Specifically, PGR and EGFR were downregulated (p-value < 0.05) in BC samples when compared to controls (**Figure 1C**). In contrast, ESR1 and ERBB2 displayed up-regulation (p-value < 0.05) in the BC samples (**Figure 1C**).

Continuing our investigation using the GEPIA database, we delved deeper into the expression patterns of PGR, ESR1, EGFR, and ERBB2, in BC patients with different stages of cancer. This analysis provided insight into how the expression of these biomarkers changed according to stage. Results showed consistent trends. Specifically, PGR and EGFR were consistently down-regulated, with a reduction in their expression as BC stage increased (**Figure 2**). On the other hand, ESR1 and ERBB2 exhibited consistent up-regulation, with an increase of expression as BC stage increased (**Figure 2**).

The findings highlight the dynamic nature of these biomarkers in the context of BC progression. The differential expression of PGR, ESR1, EGFR, and ERBB2 at different stages of the disease may have implications for understanding disease progression, prognosis, and the use of

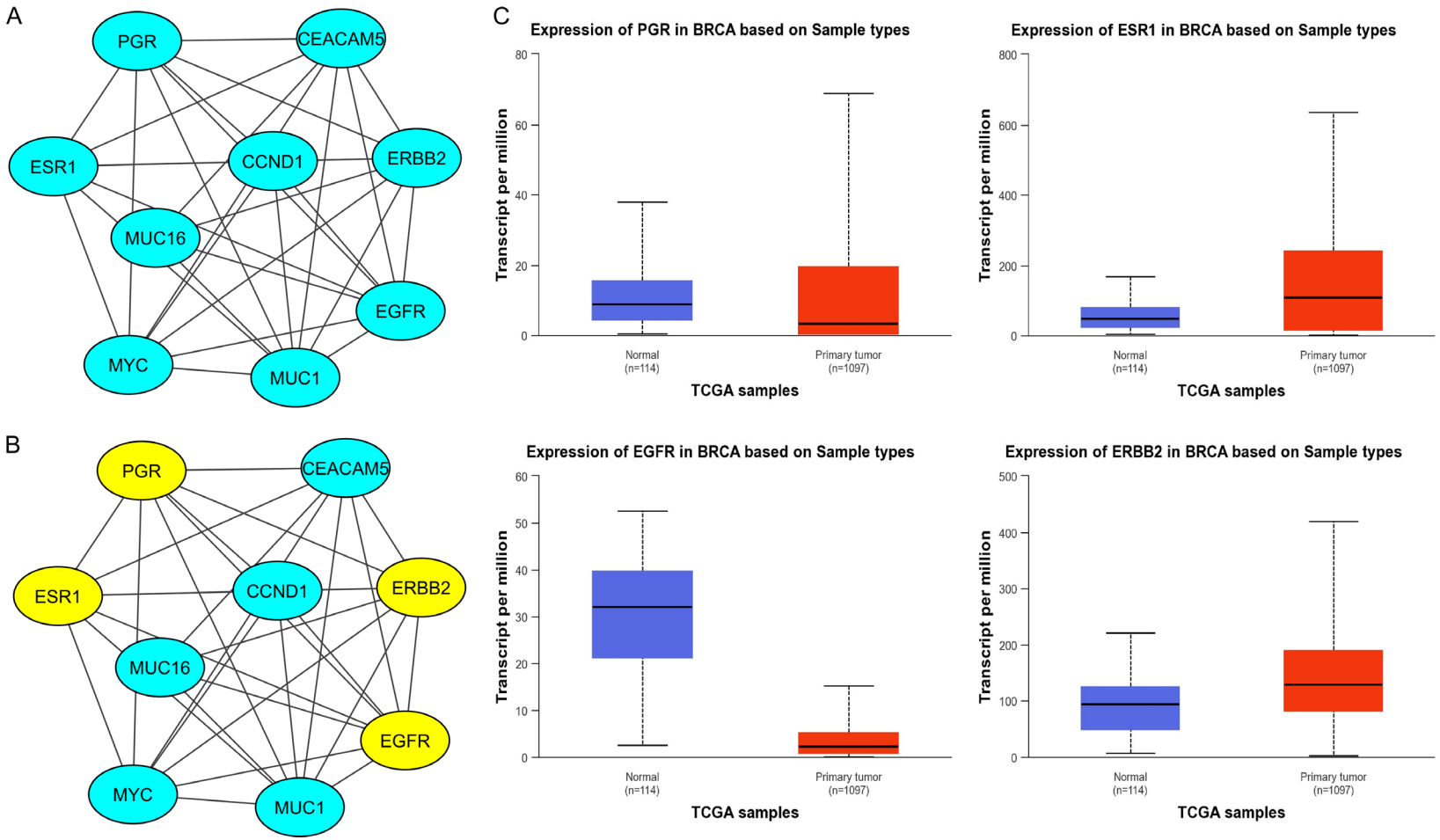

Figure 1. Construction of a Protein-Protein Interaction (PPI) network, with identification of key biomarkers, and the analysis of their expression profiles utilizing The Cancer Genome Atlas (TCGA) dataset by the UALCAN platform. (A) This portrays the initial step in this study, which involved constructing a PPI network of the extracted nine biomarkers, (B) This shows the PPI network constructed for the nine biomarkers, with a specific focus on the identification of four key biomarkers. The 4 markers were selected using the degree method within the PPI network, signifying their central role within the network, and (C) Expression profiling results obtained from analyzing the TCGA dataset using the UALCAN platform. A *p*-value < 0.05 was considered significant.

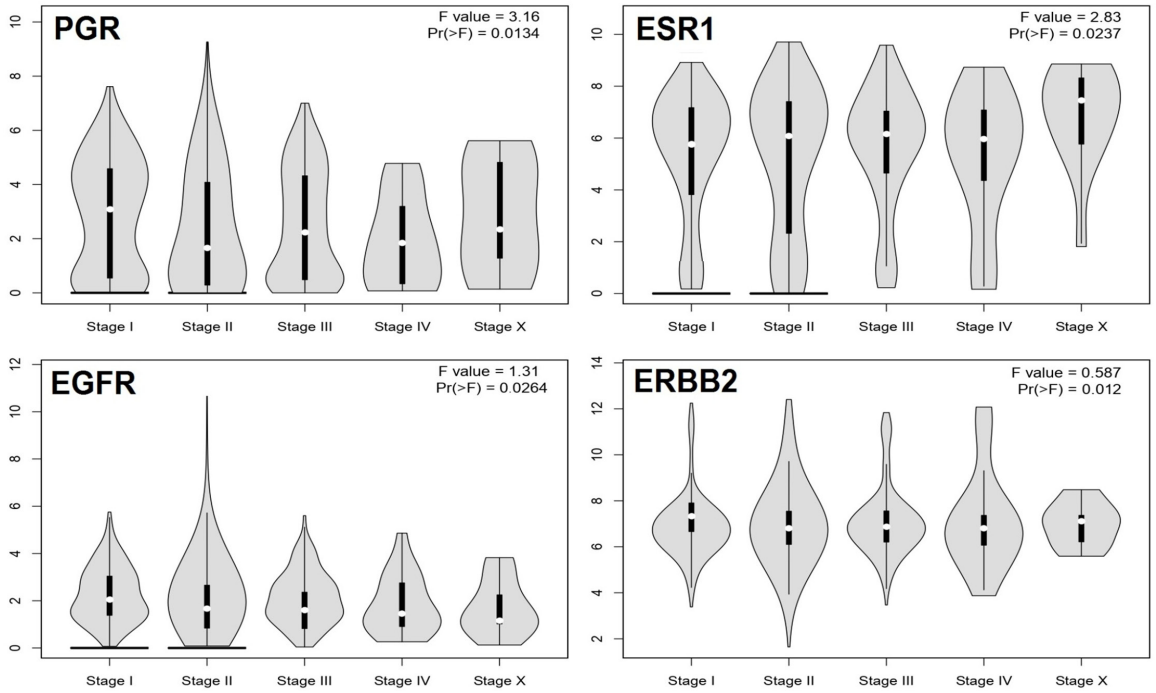

Figure 2. Detailed depiction of the expression profiles of key biomarkers in breast cancer (BC) patients at different cancer stages and in normal control samples. Analysis was conducted utilizing the GEPIA database. A *p*-value < 0.05 was considered significant.

targeted therapeutic interventions tailored to specific stages of BC.

Survival analysis of key biomarkers

KM Plotter tool was used to explore prognostic implications of key biomarkers in BC. Results showed the roles of PGR, ESR1, EGFR, and ERBB2 expression for overall survival (OS) of BC patients. Lower expression levels of PGR and EGFR were associated with poor OS among breast cancer patients (**Figure 3**). Higher expression levels of ESR1 and ERBB2 were linked to adverse OS outcome (**Figure 3**). These results underscore the prognostic use of the 4 key biomarkers for predicting survival in BC patients.

Expression analysis of key biomarkers using clinical BC samples

We conducted a comprehensive analysis of PGR, ESR1, EGFR, and ERBB2 expression using clinical BC tissue samples obtained from Pakistani patients. We meticulously followed a well-defined procedure, including RNA extraction, complementary DNA (cDNA) synthesis, and RT-qPCR analysis. Our results, consistent with TCGA dataset analysis, revealed significant (*p*-value < 0.05) down-regulation of PGR and EGFR (**Figure 4A**) and up-regulation (*p*-value < 0.05) of ESR1 and CRBB2 in BC samples compared to controls (**Figure 4A**).

Furthermore, the ROC curves based on RT-qPCR expression data for PGR, ESR1, EGFR, and ERBB2 provided additional evidence of their accuracy and sensitivity as biomarkers (**Figure 4B**). The curves demonstrated that these genes possess the discriminative power (**Figure 4B**) needed to serve as reliable biomarkers in breast cancer diagnosis and prognosis.

Gene enrichment analysis of key biomarkers

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, including the prediction of biological process (BP), cellular component (CC), and molecular function (MF) of the key biomarker genes (PGR, ESR1, EGFR, and ERBB2) were conducted using the Medscape tool. The identified key biomarker genes were highly enriched in

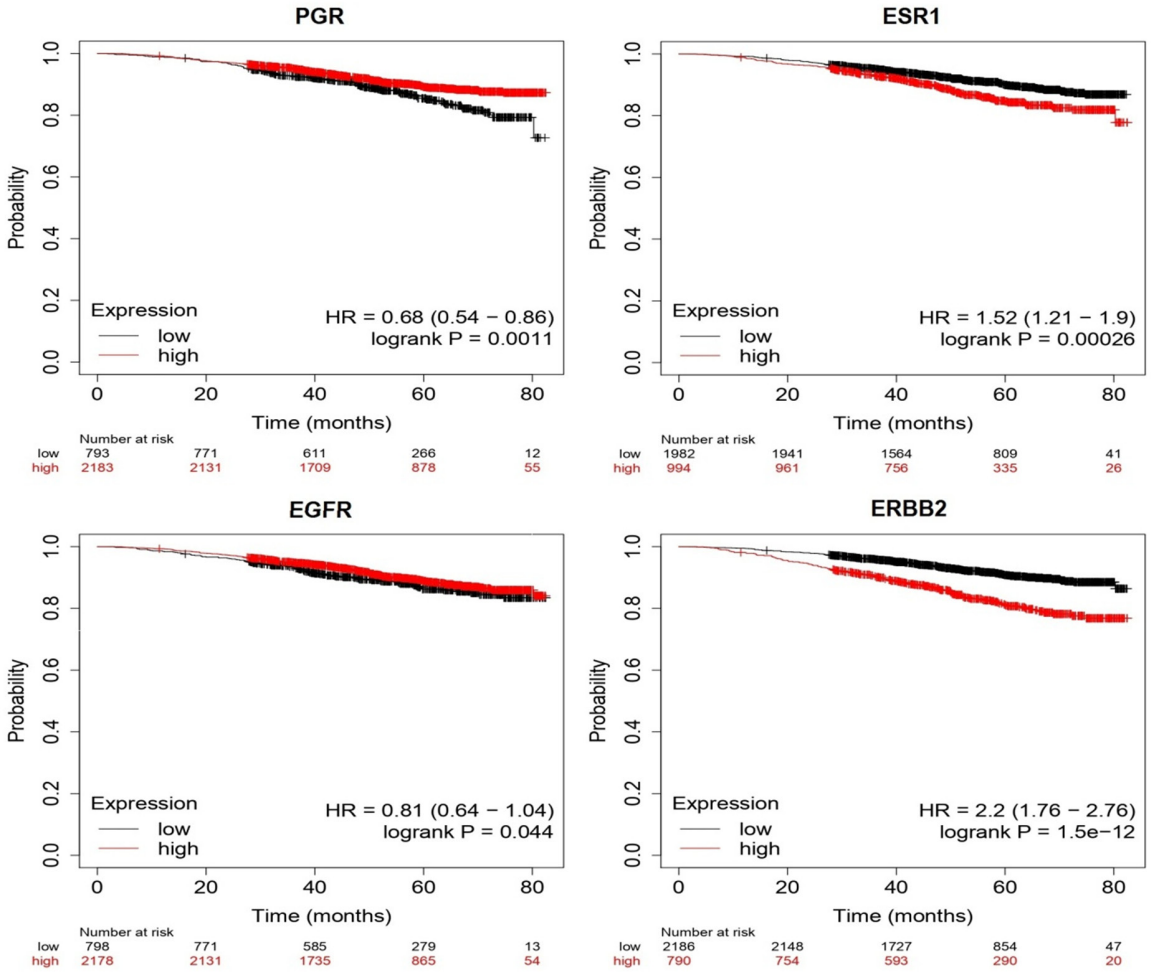

Figure 3. Survival analysis of key biomarkers using KM Plotter tool. This figure provides critical insight into the prognostic value of these key biomarkers in breast cancer (BC). A significant difference in survival probability between high and low expression groups highlights the clinical relevance of these biomarkers to predict patient outcome. A *p*-value < 0.05 was considered significant.

"ERBB3:ERBB2 complex, Shc-EGFR complex, and multivesicular body, internal vesicle" etc., CC terms (**Figure 5A**). Regarding MF, the identified key biomarker genes were mainly enriched in the "epidermal-growth factor-activated receptor activity, estrogen response element binding, and RNA polymerase I core binding" etc., terms (**Figure 5B**). Concerning BP, the identified hub genes were mainly involved in "mammary gland branching involved in pregnancy and branching involved in mammary gland duct morphogenesis" etc., terms (**Figure 5C**). Moreover, KEGG analysis revealed that PGR, ESR1, EGFR, and ERBB2 were associated with the dysregulation of important signaling pathways, including "cancer development, central carbon metabolism in cancer, and adherens junction", etc. (**Figure 5D**).

Retrieval of drugs from DrugBank

DrugBank database was used to determine which drugs were associated with PGR, ESR1, EGFR, or ERBB2. By queuing the DrugBank, we successfully identified a few important drugs, including Estradiol, Decitabine, and Carbamazepine (**Table 2**), that have the ability to modulate expression of these genes. This information deserves further exploration to discover targeted interventions that may impact the expression of these biomarkers in a clinical setting.

Discussion

Breast cancer (BC) remains a significant global health concern, characterized by its high inci-

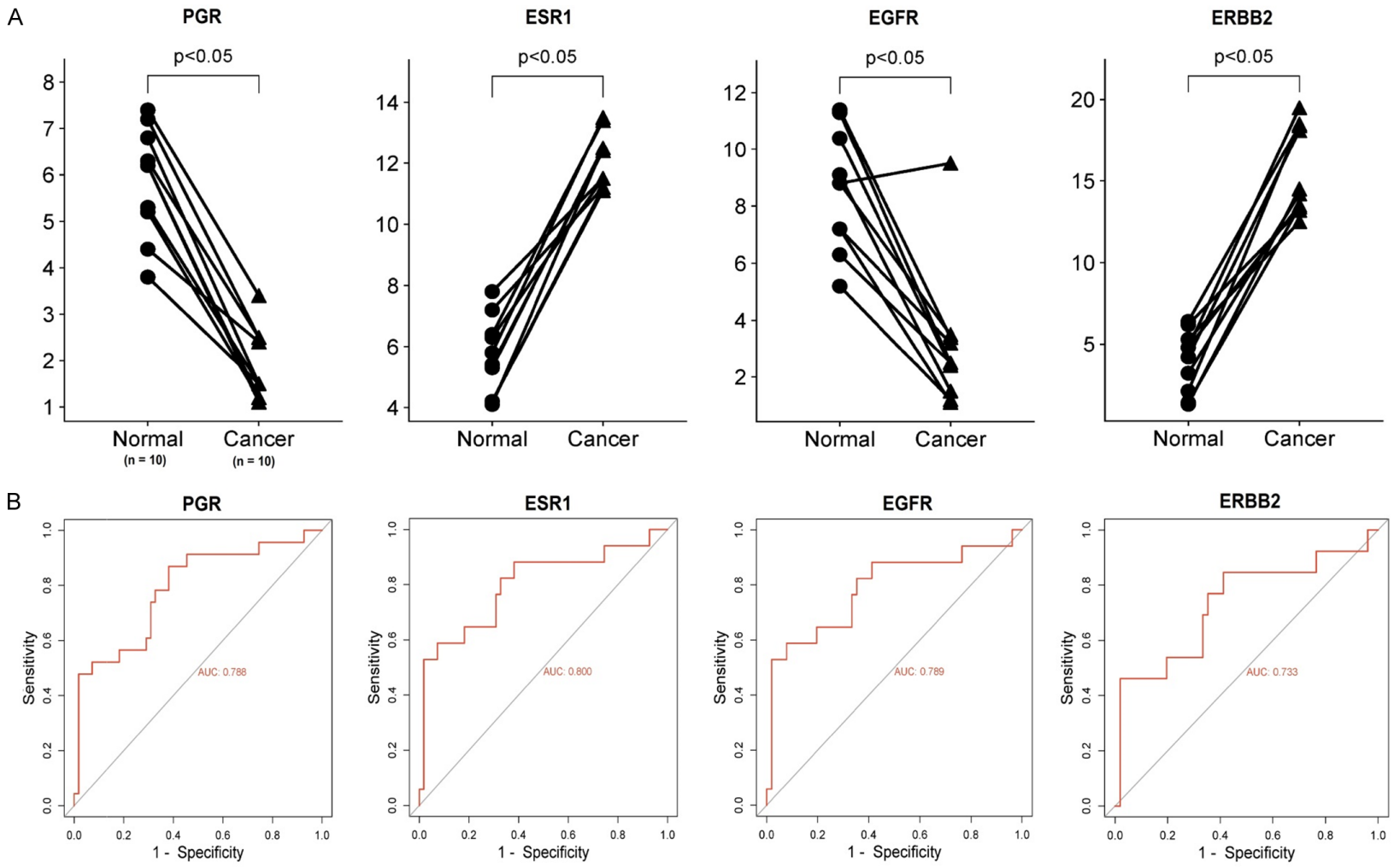

Figure 4. Expression profiling of the key biomarkers using breast cancer (BC) clinical tissue samples paired with control samples, and ROC curve analysis. (A) Expression profiling of key biomarkers using clinical tissue samples and control samples obtained from the Pakistani BC patients, and (B) RT-qPCR expression level-based ROC curves of the key biomarkers. A *p*-value < 0.05 was considered significant. ROC = Receiver operating curve, RT-qPCR = Reverse transcription quantitative real-time PCR.

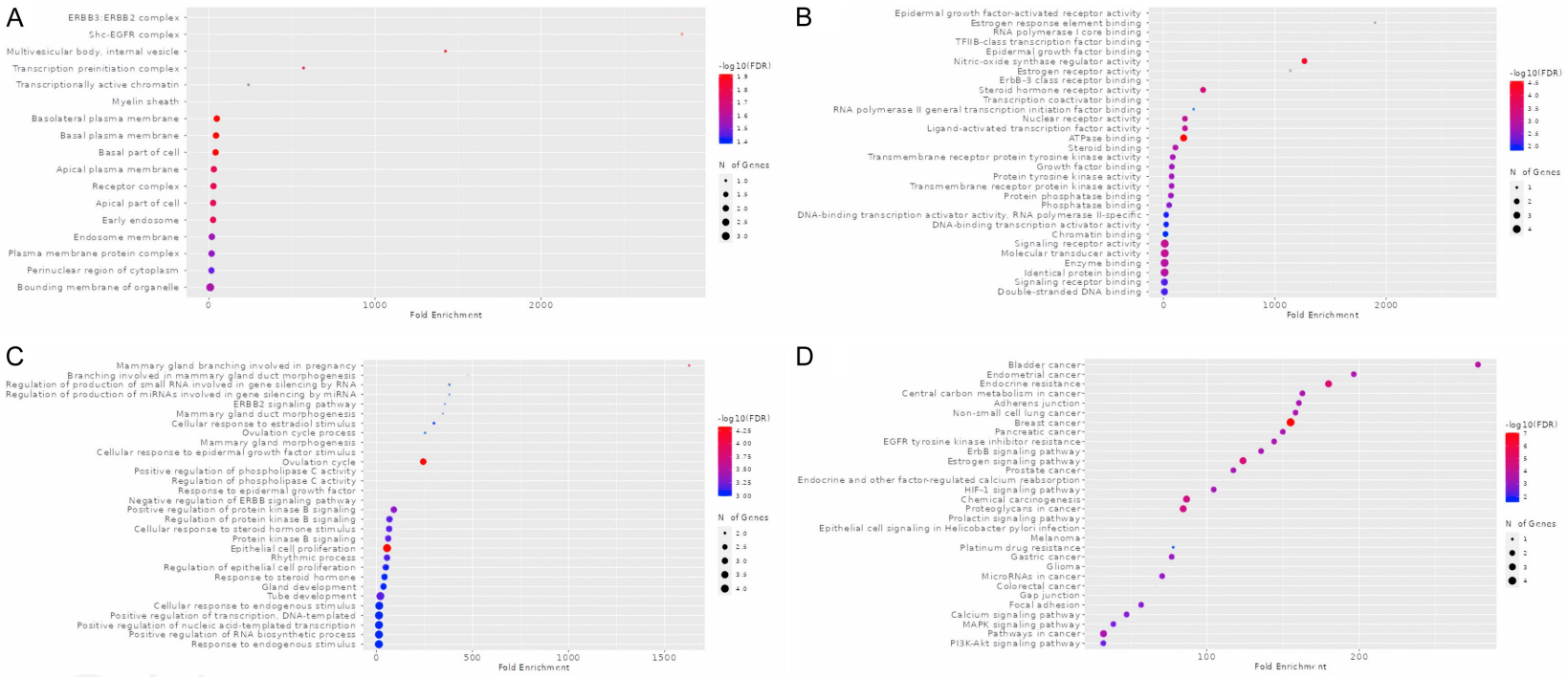

Figure 5. Gene enrichment analysis of key biomarkers by Metascape program. (A) CC terms, (B) BP terms, (C) MF terms, and (D) KEGG terms. A *p*-value < 0.05 was considered significant. CC = Cellular Component, BP = Biological Process, MF = Molecular Function, KEGG = Kyoto Encyclopedia of Genes and Genomes.

Table 2. DrugBank-based PGR, ESR1, EGFR, and ERBB2 gene-associated drugs

Sr. No	Hub gene	Drug name	Effect	Reference	Group
1	PGR	Estradiol	Increase expression of PGR mRNA	A21144	Approved
		Decitabine		A21918	
2	ESR1	Estradiol	Decrease expression of ESR1 mRNA	A21329	Approved
		Carbamazepine		A21542	
3	EGFR	Estradiol	Increase expression of EGFR mRNA	A21329	Approved
4	ERBB2	Estradiol	Decrease expression of ERBB2 mRNA	A21329	Approved

PGR = Progesterone receptor, ESR1 = Estrogen Receptor 1, EGFR = Epidermal growth factor receptor, ERBB2 = Erb-B2 receptor tyrosine kinase 2.

dence, mortality rate, and the challenges associated with early detection and effective treatment [31]. In this study, we aimed to identify all the already reported BC biomarkers from the literature and shortlist and validate a few more important biomarkers by multi-level methodology. The selection of biomarkers for this study was guided by a meticulous review of existing literature and the importance of these markers in BC pathogenesis and progression. A total of nine biomarkers were gathered from the literature, satisfying the criterion of exhibiting an accuracy level exceeding 20% in both specificity and sensitivity. Later, after applying the degree method, we focused on four key biomarkers: PGR, ESR1, EGFR, and ERBB2.

Our TCGA datasets analysis yielded valuable insight into the expression patterns of these biomarkers in breast cancer (BC). Results were consistent with previous studies, reinforcing the significance of these markers in BC. Specifically, we found that PGR and EGFR were significantly down-regulated in BC samples, aligning with research such as Suzuki et al. [32], which reported decreased PGR expression in BC. Contrary to this, down-regulation of PGR and EGFR was also observed in BC. One study by Lv et al. [33] investigated the expression of progesterone receptor (PGR) in BC. Interestingly, they found that in a subset of BC cases, PGR was overexpressed. This up-regulation of PGR was associated with a specific subtype of BC, suggesting that PGR expression can vary among different BC subtypes. While EGFR is often associated with a poor prognosis, there are instances of EGFR up-regulation in BC. For example, Baselga et al. [34] explored the role of EGFR in BC. They found that in a subset of BC patients, EGFR was indeed up-regulated. This up-regulation was associated with a more aggressive form of BC, highlighting the heterogeneity in EGFR expression among BC cases. Our observation of up-regulated ESR1 expression in BC samples was in line with findings from the study by Harrell et al. [35], that highlighted the importance of ESR1 in BC progression. Furthermore, the up-regulation of ERBB2 in BC samples was consistent with studies like the one conducted by Slamon et al. [36], emphasizing the role of ERBB2 in BC. However, down-regulation of ESR1 and ERBB2 expression in BC has also been reported in various studies, reflecting the diverse molecular characteristics of BC subtypes. For example, in a study by Rody et al. [37], the authors investigated BC subtypes and found that ESR1 expression was significantly down-regulated in triple-negative BC (TNBC), a subtype known for its lack of hormone receptor expression. This down-regulation underscores the distinct biology of TNBC compared to hormone receptor-positive BC. Moreover, a study by Guedj et al. [38] explored the heterogeneity of BC and reported that a subset of HER2-negative BC samples exhibited down-regulated expression of ERBB2.

To ensure the clinical relevance of our findings, we conducted a meticulous validation process using clinical samples obtained from BC patients in Pakistan. This validation process involved the extraction of total RNA, cDNA synthesis, and RT-qPCR analysis. The results obtained from this clinical validation mirrored the trends observed in the TCGA dataset analysis. This convergence of findings further strengthens the applicability of these biomarkers in real-world clinical settings.

To extend the significance of PGR, ESR1, EGFR, and ERBB2 beyond diagnosis, we analyzed the effect of their dysregulation on the OS of BC patients. Results revealed that lower expression of PGR and EGFR, coupled with higher expression of ESR1 and ERBB2, corre-

lated with poorer overall survival in BC patients. Our findings are in line with the previous studies reporting PGR, ESR1, EGFR, and ERBB2 as excellent prognostic biomarkers of BC [36, 39-41].

Moreover, one of the promising outcomes of our study was the identification of key drugs (Estradiol, Decitabine, and Carbamazepine) associated with the PGR, ESR1, EGFR, and ERBB2 biomarkers through the DrugBank database. This discovery opens avenues for the development of targeted therapeutic approaches in BC treatment. Tailoring treatments based on the expression patterns of these biomarkers may improve patient outcomes and reduce adverse effects.

While this study provides critical insight into BC diagnosis, prognosis, and treatment, it is essential to acknowledge its limitations. Larger-scale clinical validation studies are warranted to further validate the clinical utility of these biomarkers. Finally, preclinical and clinical trials should be conducted to translate these findings into practical application for BC patients.

Conclusion

This study expands our knowledge in the field of BC research. The identification and validation of PGR, ESR1, EGFR, and ERBB2 as key biomarkers holds significant clinical promise. These findings have the potential to revolutionize BC diagnostic and treatment strategies, ultimately improving patient outcome.

Acknowledgements

The authors extend their appreciation to the Researchers Supporting Project number (RSPD2023R1089) King Saud University, Riyadh, Saudi Arabia.

Disclosure of conflict of interest

None.

Address correspondence to: Muhammad Jamil, PARC Arid Zone Research Centre, Dera Ismail Khan 29050, Pakistan. E-mail: jamilmatrah@gmail.com

References

[1] Zipinotti Dos Santos D, de Souza JC, Pimenta TM, da Silva Martins B, Junior RSR, Butzene SMS, Tessarolo NG, Cilas PML Jr, Silva IV and Rangel LBA. The impact of lipid metabolism on breast cancer: a review about its role in tumorigenesis and immune escape. Cell Commun Signal 2023; 21: 161.

[2] Lukong KE. Understanding breast cancer - the long and winding road. BBA Clin 2017; 7: 64-77.

[3] Usman M, Okla MK, Asif HM, AbdElgayed G, Muccee F, Ghazanfar S, Ahmad M, Iqbal MJ, Sahar AM, Khaliq G, Shoaib R, Zaheer H and Hameed Y. A pan-cancer analysis of GINS complex subunit 4 to identify its potential role as a biomarker in multiple human cancers. Am J Cancer Res 2022; 12: 986-1008.

[4] Ferlay J, Héry C, Autier P and Sankaranarayanan R. Global burden of breast cancer. J Cancer Epidemiol 2010; 7: 1-19.

[5] Skakkebaek NE, Rajpert-De Meyts E, Buck Louis GM, Toppari J, Andersson AM, Eisenberg ML, Jensen TK, Jørgensen N, Swan SH, Sapra KJ, Ziebe S, Priskorn L and Juul A. Male reproductive disorders and fertility trends: influences of environment and genetic susceptibility. Physiol Rev 2016; 96: 55-97.

[6] Lammert J, Grill S and Kiechle M. Modifiable lifestyle factors: opportunities for (hereditary) breast cancer prevention - a narrative review. Breast Care (Basel) 2018; 13: 109-114.

[7] Ahmad M, Hameed Y, Khan M, Usman M, Rehman A, Abid U, Asif R, Ahmed H, Hussain MS, Rehman JU, Asif HM, Arshad R, Atif M, Hadi A, Sarfraz U and Khurshid U. Up-regulation of GINS1 highlighted a good diagnostic and prognostic potential of survival in three different subtypes of human cancer. Braz J Biol 2021; 84: e250575.

[8] Kelsey JL and Berkowitz GS. Breast cancer epidemiology. Cancer Res 1988; 48: 5615-5623.

[9] Goolsby MA and Boniquit N. Bone health in athletes. Sports Health 2017; 9: 108-117.

[10] Zhu X, Tang L, Mao J, Hameed Y, Zhang J, Li N, Wu D, Huang Y and Li C. Decoding the mechanism behind the pathogenesis of the focal segmental glomerulosclerosis. Comput Math Methods Med 2022; 2022: 1941038.

[11] Usman M, Hameed Y, Ahmad M, Jalil Ur Rehman, Ahmed H, Hussain MS, Asif R, Murtaza MG, Jawad MT and Iqbal MJ. Breast cancer risk and human papillomavirus infection: a Bradford Hill criteria based evaluation. Infect Disord Drug Targets 2022; 22: e200122200389.

[12] Mao J, Huang X, Okla MK, Abdel-Maksoud MA, Mubarak A, Hameed Z, Noreen R, Chaudhary A, Ghazanfar S, Liao Y, Hameed Y and Li C. Risk factors for TERT promoter mutations with papillary thyroid carcinoma patients: a meta-analysis and systematic review. Comput Math Methods Med 2022; 2022: 1721526.

[13] Lawler M, Alsina D, Adams RA, Anderson AS, Brown G, Fearnhead NS, Fenwick SW, Halloran SP, Hochhauser D, Hull MA, Koelzer VH, McNair AGK, Monahan KJ, Näthke I, Norton C, Novelli MR, Steele RJC, Thomas AL, Wilde LM, Wilson RH and Tomlinson I; Bowel Cancer UK Critical Research Gaps in Colorectal Cancer Initiative. Critical research gaps and recommendations to inform research prioritisation for more effective prevention and improved outcomes in colorectal cancer. Gut 2018; 67: 179-193.

[14] Sial N, Rehman JU, Saeed S, Ahmad M, Hameed Y, Atif M, Rehman A, Asif R, Ahmed H, Hussain MS, Khan MR, Ambreen A and Ambreen A. Integrative analysis reveals methylenetetrahydrofolate dehydrogenase 1-like as an independent shared diagnostic and prognostic biomarker in five different human cancers. Biosci Rep 2022; 42: BSR20211783.

[15] Zhang L, Sahar AM, Li C, Chaudhary A, Yousaf I, Saeedah MA, Mubarak A, Haris M, Nawaz M, Reem MA, Ramadan FA, Mostafa AAM, Feng W and Hameed Y. A detailed multi-omics analysis of GNB2 gene in human cancers. Braz J Biol 2022; 84: e260169.

[16] Xu W, Li H, Hameed Y, Abdel-Maksoud MA, Almutairi SM, Mubarak A, Aufy M, Alturaiki W, Alshalani AJ, Mahmoud AM and Li C. Elucidating the clinical and immunological value of m6A regulator-mediated methylation modification patterns in adrenocortical carcinoma. Oncol Res 2023; 31: 819-831.

[17] Bertoli G, Cava C and Castiglioni I. MicroRNAs: new biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. Theranostics 2015; 5: 1122-43.

[18] Li S, Yi M, Dong B, Tan X, Luo S and Wu K. The role of exosomes in liquid biopsy for cancer diagnosis and prognosis prediction. Int J Cancer 2021; 148: 2640-2651.

[19] Usman M, Hameed Y, Ahmad M, Iqbal MJ, Maryam A, Mazhar A, Naz S, Tanveer R, Saeed H, Bint-E-Fatima, Ashraf A, Hadi A, Hameed Z, Tariq E and Aslam AS. SHMT2 is associated with tumor purity, CD8+ T immune cells infiltration, and a novel therapeutic target in four different human cancers. Curr Mol Med 2023; 23: 161-176.

[20] Ullah L, Hameed Y, Ejaz S, Raashid A, Iqbal J, Ullah I and Ejaz SA. Detection of novel infiltrating ductal carcinoma-associated BReast CAncer gene 2 mutations which alter the deoxyribonucleic acid-binding ability of BReast CAncer gene 2 protein. J Cancer Res Ther 2020; 16: 1402-1407.

[21] Hossain MB, Islam MM, Abdulrazak LF, Rana MM, Akib TBA and Hassan M. Graphene-coated optical fiber SPR biosensor for BRCA1 and BRCA2 breast cancer biomarker detection: a numerical design-based analysis. Photonic Sens 2020; 10: 67-79.

[22] Evans DG, Lalloo F, Cramer A, Jones EA, Knox F, Amir E and Howell A. Addition of pathology and biomarker information significantly improves the performance of the Manchester scoring system for BRCA1 and BRCA2 testing. J Med Genet 2009; 46: 811-817.

[23] James CR, Quinn JE, Mullan PB, Johnston PG and Harkin DP. BRCA1, a potential predictive biomarker in the treatment of breast cancer. Oncologist 2007; 12: 142-150.

[24] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.

[25] Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.

[26] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK and Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 2017; 19: 649-658.

[27] Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98-W102.

[28] Lánczky A and Győrffy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. J Med Internet Res 2021; 23: e27633.

[29] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C and Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019; 10: 1523.

[30] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z and Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006; 34: D668-72.

[31] Wilkinson L and Gathani T. Understanding breast cancer as a global health concern. Br J Radiol 2022; 95: 20211033.

[32] Suzuki T, Miki Y, Nakamura Y, Moriya T, Ito K, Ohuchi N and Sasano H. Sex steroid-producing enzymes in human breast cancer. Endocr Relat Cancer 2005; 12: 701-720.

[33] Lv X, Dobrolecki LE, Ding Y, Rosen JM, Lewis MT and Chen X. Orthotopic transplantation of

breast tumors as preclinical models for breast cancer. J Vis Exp 2020; 10.3791/61173.
[34] Baselga J, Albanell J, Molina MA and Arribas J. Mechanism of action of trastuzumab and scientific update. Semin Oncol 2001; 28 Suppl 16: 4-11.
[35] Harrell JC, Dye WW, Allred DC, Jedlicka P, Spoelstra NS, Sartorius CA and Horwitz KB. Estrogen receptor positive breast cancer metastasis: altered hormonal sensitivity and tumor aggressiveness in lymphatic vessels and lymph nodes. Cancer Res 2006; 66: 9308-9315.
[36] Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A and McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987; 235: 177-182.
[37] Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, Gaetje R, Solbach C, Ahr A, Metzler D, Schmidt M, Müller V, Holtrich U and Kaufmann M. A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res 2011; 13: R97.
[38] Guedj M, Marisa L, de Reynies A, Orsetti B, Schiappa R, Bibeau F, MacGrogan G, Lerebours F, Finetti P, Longy M, Bertheau P, Bertrand F, Bonnet F, Martin AL, Feugeas JP, Bièche I, Lehmann-Che J, Lidereau R, Birnbaum D, Bertucci F, de Thé H and Theillet C. A refined molecular taxonomy of breast cancer. Oncogene 2012; 31: 1196-1206.
[39] Viale G, Regan MM, Maiorano E, Mastropasqua MG, Dell'Orto P, Rasmussen BB, Raffoul J, Neven P, Orosz Z, Braye S, Ohlschlegel C, Thürlimann B, Gelber RD, Castiglione-Gertsch M, Price KN, Goldhirsch A, Gusterson BA and Coates AS. Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98. J Clin Oncol 2007; 25: 3846-3852.
[40] Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J and Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351: 2817-2826.
[41] Carey LA, Rugo HS, Marcom PK, Mayer EL, Esteva FJ, Ma CX, Liu MC, Storniolo AM, Rimawi MF, Forero-Torres A, Wolff AC, Hobday TJ, Ivanova A, Chiu WK, Ferraro M, Burrows E, Bernard PS, Hoadley KA, Perou CM and Winer EP. TBCRC 001: randomized phase II study of cetuximab in combination with carboplatin in stage IV triple-negative breast cancer. J Clin Oncol 2012; 30: 2615-2623.
[42] Paik S, Bryant J, Tan-Chiu E, Yothers G, Park C, Wickerham DL and Wolmark N. HER2 and choice of adjuvant chemotherapy for invasive breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-15. J Natl Cancer Inst 2000; 92: 1991-1998.
[43] Zhu K, Bernard LJ, Levine RS and Williams SM. Estrogen receptor status of breast cancer: a marker of different stages of tumor or different entities of the disease? Med Hypotheses 1997; 49: 69-75.
[44] Tsai HW, Ho CL, Cheng SW, Lin YJ, Chen CC, Cheng PN, Yen CJ, Chang TT, Chiang PM, Chan SH, Ho CH, Chen SH, Wang YW, Chow NH and Lin JC. Progesterone receptor membrane component 1 as a potential prognostic biomarker for hepatocellular carcinoma. World J Gastroenterol 2018; 24: 1152-1166.
[45] Masuda H, Zhang D, Bartholomeusz C, Doihara H, Hortobagyi GN and Ueno NT. Role of epidermal growth factor receptor in breast cancer. Breast Cancer Res Treat 2012; 136: 331-345.
[46] Hou MF, Chen YL, Tseng TF, Lin CM, Chen MS, Huang CJ, Huang YS, Hsieh JS, Huang TJ, Jong SB and Huang YF. Evaluation of serum CA27.29, CA15-3 and CEA in patients with breast cancer. Kaohsiung J Med Sci 1999; 15: 520-528.
[47] Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value? Clin Chem 2006; 52: 345-351.
[48] Duffy MJ, Shering S, Sherry F, McDermott E and O'Higgins N. CA 15-3: a prognostic marker in breast cancer. Int J Biol Markers 2000; 15: 330-333.
[49] Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO and Botstein D. Molecular portraits of human breast tumours. Nature 2000; 406: 747-752.
[50] Rudas M, Lehnert M, Huynh A, Jakesz R, Singer C, Lax S, Schippinger W, Dietze O, Greil R, Stiglbauer W, Kwasny W, Grill R, Stierer M, Gnant MF and Filipits M; Austrian Breast and Colorectal Cancer Study Group. Cyclin D1 expression in breast cancer patients receiving adjuvant tamoxifen-based therapy. Clin Cancer Res 2008; 14: 1767-1774.