The Key Role and Potential Molecular Mechanism of HER2 (ERBB2) in Breast Cancer: A Bioinformatics Analysis Based on the TCGA-BRCA Database.

Yingaoran Lyu *

School of Life Health Information and Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

* Corresponding Author Email: yingaoraanlyu@gmail.com

Abstract. Breast cancer is one of the most common malignant tumors in women, with the HER2-positive subtype exhibiting strong invasiveness and poor prognosis. ERBB2, a key encoding gene of HER2, plays a crucial role in the development and progression of breast cancer, and its molecular mechanisms still require further in-depth analysis. This study, based on the TCGA-BRCA database, integrated transcriptome expression data, clinical information, and survival data to comprehensively analyze the molecular characteristics related to ERBB2. Differential expression analysis identified 334 significantly expressed genes, and GO and Hallmark pathway analyses revealed that they are mainly involved in the ERBB2 signaling pathway, estrogen response, and lipid metabolism. Further PPI network construction and module analysis revealed that ERBB2, GRB7, STARD3, PGAP3, and MIEN1 form a highly cohesive core functional module. Based on survival analysis, there is some correlation between ERBB2 expression and patient survival. The present study sheds lights on the synergistic interactions between ERBB2 and co-amplified genes in breast cancer, indicating that they may be involved in the synergistic progression of breast cancer and metabolic reprogramming. The results serve as a reference for the research of molecular mechanism of breast cancer that is positive to the HER2 receptor and screening for potential targets, but there is no experiment verification. Future studies could integrate multi-omics with functional experiments to further confirm the function of important genes and to study related molecular mechanisms.

Keywords: ERBB2; Breast Cancer; Bioinformatics Analysis; Protein–Protein Interaction Network; Functional Enrichment Analysis.

1. Introduction

Breast cancer is one of the most common cancers among women worldwide and a leading cause of cancer-related deaths, imposing a substantial social and economic burden [1]. The causes of breast cancer may be related to alcohol intake, long-term smoking, obesity, radiation exposure, and other factors. Breast cancer encompasses a variety of malignant tumors with distinct characteristics and is generally recognized to include at least four clinically relevant molecular subtypes: luminal A (ER+ and/or PR+/HER2−), luminal B (ER+ and/or PR+/HER2+), HER2-positive (ER−PR−HER2+), and basal-like (ER−PR−HER2−). Among these, HER2-positive breast cancer accounts for 15%–20% of all breast cancer cases.

The hallmark of HER2-positive breast cancer is HER2 gene amplification and/or overexpression of its corresponding kinase receptor protein. These molecular alterations render the tumor more aggressive, resulting in reduced disease-free survival and overall survival [1]. Studies have shown that HER2 gene expression levels are strongly correlated with breast cancer progression and patient prognosis, which has led to the development of HER2-targeted therapies [2–4]. Despite the remarkable success of HER2-targeted therapy, the molecular regulatory mechanisms mediated by HER2 remain incompletely understood.

In recent years, advances in high-throughput sequencing technology and bioinformatics analysis have opened new avenues for investigating the molecular mechanisms of tumorigenesis. However, HER2 expression exhibits intratumoral heterogeneity. In this study, based on the TCGA-BRCA database, we employed systematic bioinformatics analysis to explore the critical role and underlying molecular mechanisms of HER2 in breast cancer.

2. Materials and Methods

2.1. Data Collection and Processing

The breast cancer-related genes used in this study were retrieved from the TCGA database based on the term “breast cancer”.

First, HER2 (ERBB2) expression levels were used for grouping and screening, and genes significantly associated with high HER2 expression were extracted.

The gene expression matrix was obtained from the TCGA-BRCA project, using FPKM-normalized expression data. Through differential expression analysis with screening criteria of $|\log2FC|>1$ and $FDR<0.05$, 251 candidate genes were obtained, which were significantly associated with HER2 expression.

2.2. Differential Expression Analysis

Differential expression analysis was performed on the grouped data using the limma package to identify differentially expressed genes ($|\log2FC|>1$, $FDR<0.05$).

2.3. Survival Analysis

Survival curves were plotted using the Kaplan-Meier method. The Log-rank test was used to compare overall survival (OS) differences between different expression groups, and univariate Cox regression analysis was used to assess the impact of ERBB2 gene expression on patient prognosis.

2.4. GO Enrichment Analysis

To explore the potential biological functions of HER2-related genes, the clusterProfiler package in R software (4.6.0) was used for Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) analyses. An entry with P. adjusts < 0.05 was considered significantly enriched. To further illustrate the enrichment level and significance, a GO enrichment bubble chart was generated.

2.5. Hallmark Gene Enrichment Analysis

To explore the biological processes of breast cancer, GSEA analysis was performed using Hallmark Gene Sets from the MSigDB database to analyze which classic tumor biological processes HER2 act on breast cancer.

2.6. PPI Network Construction

The screened differentially expressed genes were imported into the STRING database, and the parameter Interaction Score > 0.7 was set to construct a protein network. After visualization analysis using Cytoscape software (3.10.4), the MCODE plugin was used to identify key functional modules, and the module with the highest score was selected as the core regulatory network.

3. Results

3.1. Relationship between ERBB2 Expression and Patient Prognosis

Data from 1073 breast cancer patients included in this study, of whom 923 survived and 150 died, were analyzed using Kaplan-Meier survival analysis. The results showed no statistically significant difference in overall survival between the high-expression group and the low-expression group ($p>0.05$). Further univariate Cox analysis revealed (HR=0.963, 95%CI=0.860-1.078, P=0.508) that ERBB2 expression level alone is not an independent predictor of overall survival in breast cancer patients.

3.2. Screening of Differentially Expressed Genes

Differential analysis was performed between the high- and low-expression groups of the ERBB2 gene, resulting in 334 differentially expressed genes. Of these, 251 were upregulated and 83 were downregulated. See Figure 1; the volcano plot shows a clear distribution of differentially expressed genes. (Notable genes include ERBB2, GRB7, STARD3, PGAP3, MIEN1, CDK12, PSMD3, and MED24, all of which are located near the HER2 amplification region.)

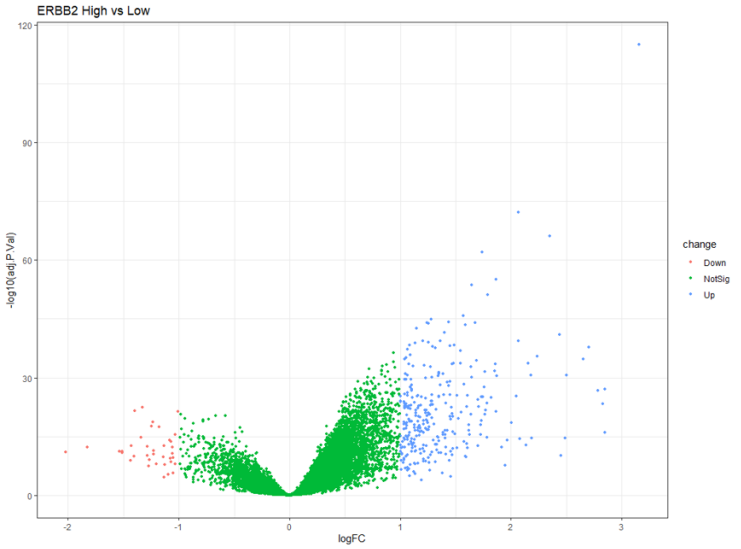


Fig 1. Differential gene distribution between ERBB2 high and low expression groups.

3.3. GO Enrichment Analysis

As shown in Figure 2, the GO results revealed that differentially expressed genes were mainly enriched in pathways such as Response to Estrogen, ERBB2 signaling pathway, and Peptidyl-tyrosine phosphorylation. As shown in Figure 3, the Response to Estrogen pathway indicates an interactive regulatory relationship between HER2 and estrogen signaling; the ERBB2 signaling pathway shows significant activation of HER2-related signaling pathways; and the Peptidyl-tyrosine phosphorylation pathway suggests that HER2 primarily regulates tumor progression through tyrosine kinase signaling.

The ERBB2 signaling pathway and cellular response to estrogen stimulus exhibited high gene ratios, further confirming that HER2-related genes are mainly involved in hormone regulation and receptor tyrosine kinase signal transduction.

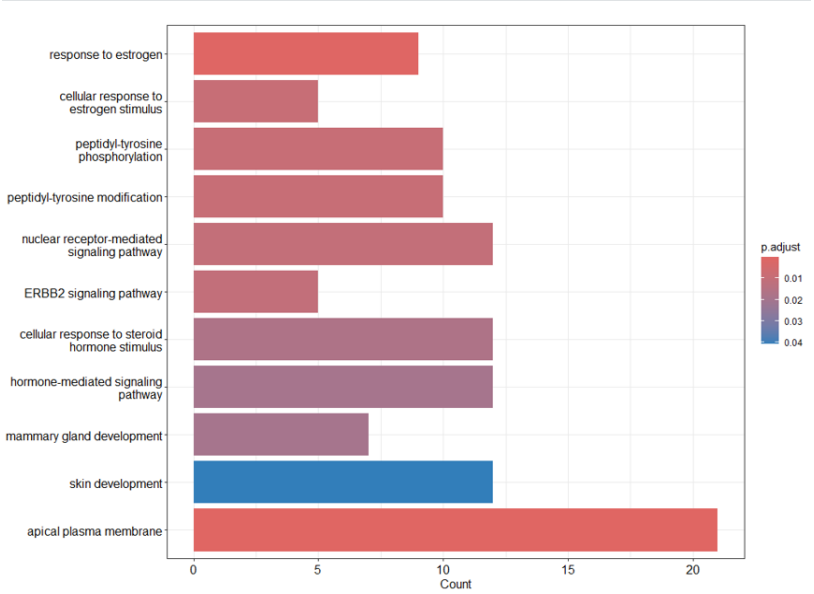


Fig 2. GO enrichment analysis of HER2-related genes

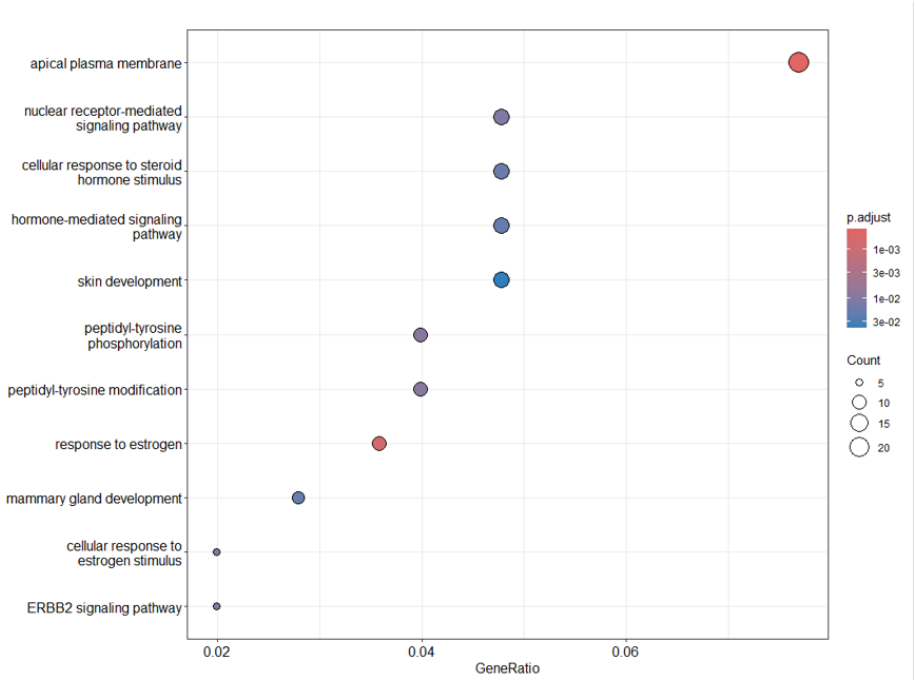


Fig 3. Bubble chart of GO functional enrichment analysis of differentially

3.4. Hallmark Pathway Enrichment Analysis

Hallmark analysis (Figure 4) showed significant enrichment of pathways such as Estrogen Response Early, Estrogen Response Late, Androgen Response, and Protein Secretion. Estrogen Response Early showed the most significant enrichment (P.adjust < 1×10^-7), suggesting an important synergistic effect between HER2 signaling and the estrogen receptor pathway.

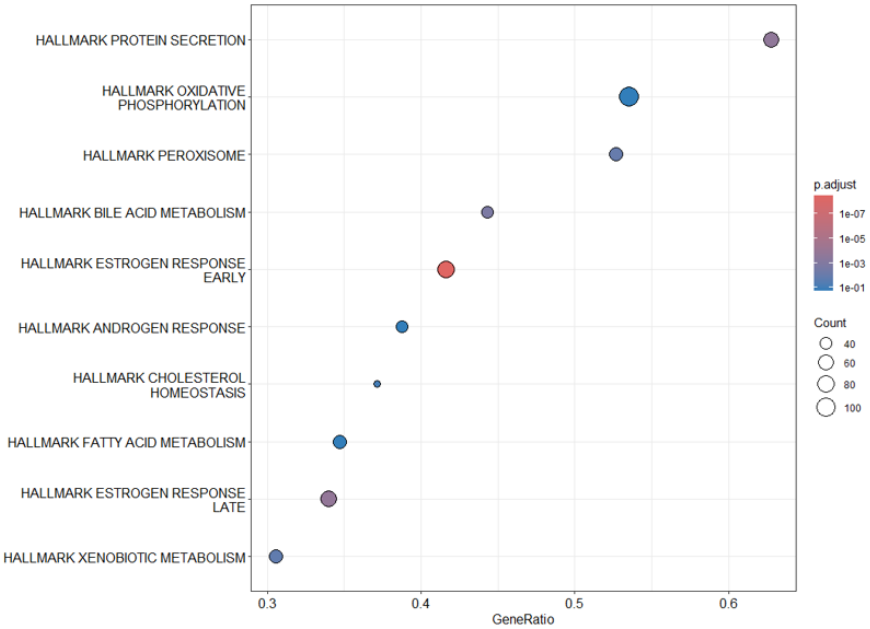


Fig 4. Hallmark enrichment analysis of HER2-related genes.

3.5. Core Module Screening

MCODE analysis identified multiple functional modules, as shown in Figure 5. The highest-scoring module (Cluster 1) included ERBB2, GRB7, STARD3, PGAP3, and MIEN1. This module is located in the 17q12 region of chromosome 17. Among them: ERBB2 is a core member of the HER2 receptor; GRB7 promotes RAS/MAKPK and PI3K-AKT signal transduction; STARD3 is involved

in cholesterol transport and membrane stability; PGAP3 is involved in the maturation of GPI-anchored proteins; and MIEN1 promotes tumor cell migration and invasion. These genes together form the core regulatory network for HER2 amplification.

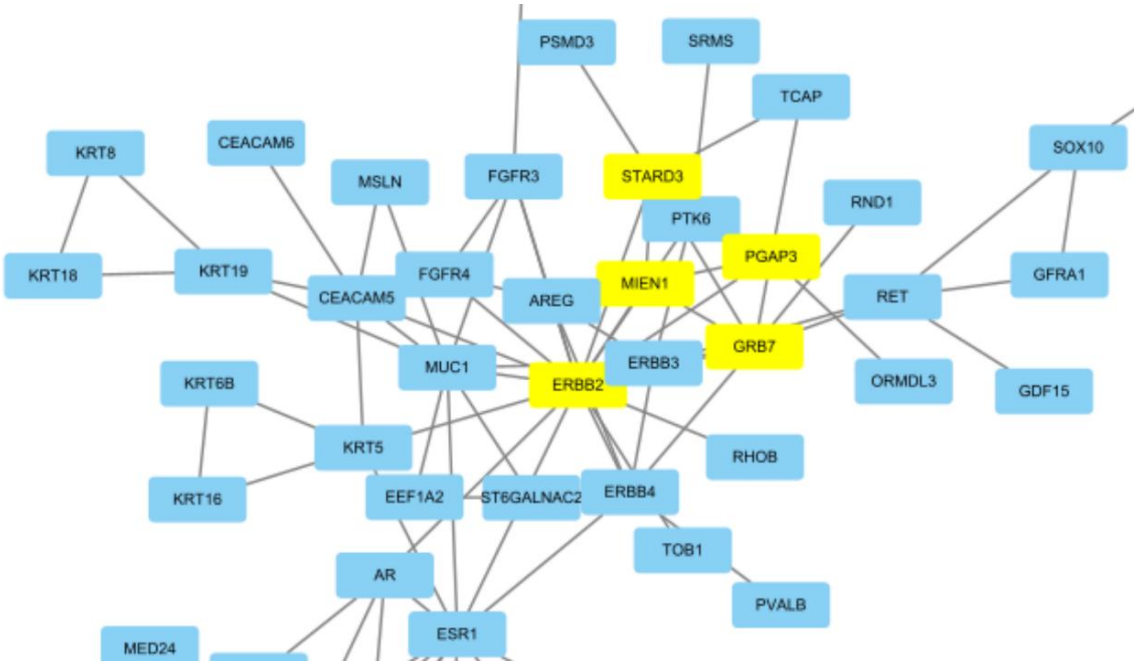


Fig 5. HER2-related differentially expressed gene-protein interaction network.

3.6. Expression Feature Analysis of Core Modules

To verify the expression characteristics of the core modules obtained from MCODE screening in HER2-positive breast cancer, expression heatmaps of ERBB2, GRB7, STARD3, PGAP3, and MIEN1 were constructed, as shown in Figure 6. The heatmaps show that ERBB2, GRB7, STARD3, PGAP3, and MIEN1 exhibit a synergistic upregulation trend in samples with high HER2 expressions.

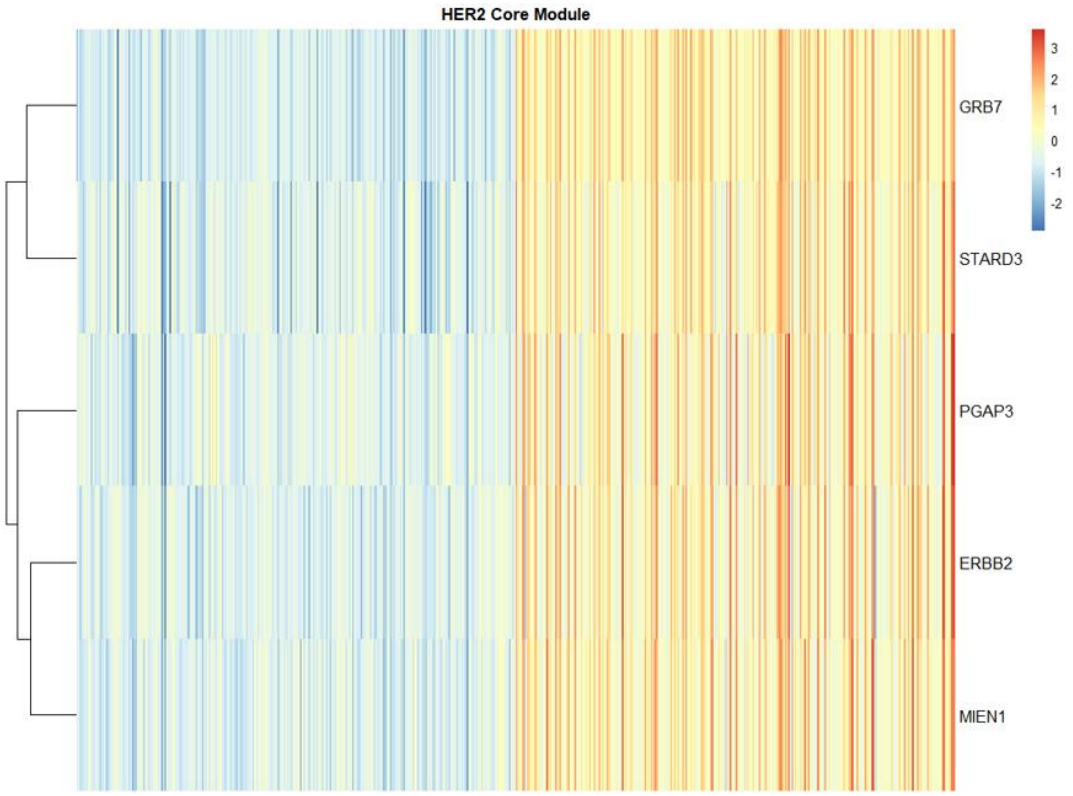


Fig 6. Heatmap of key expression of HER2 core modules in breast cancer.

4. Discussion

HER2 amplification is one of the most important molecular events in breast cancer [5]. This study analyzed the HER2-driven molecular network based on the TCGA-BRCA database system. Although survival curves plotted using the Kaplan-Meier method showed no significant correlation between ERBB2 expression levels and overall survival, differential analysis and network analysis revealed key molecular mechanisms driven by HER2.

First, GO analysis of the data revealed that samples with high ERBB2 expression were significantly enriched in the ERBB signaling pathway and tyrosine phosphorylation process, indicating that HER2 mainly promotes tumor progression by activating receptor tyrosine kinase signaling [6,7]. Second, this study identified the core functional modules composed of ERBB2, GRB7, STARD3, PGAP3, and MIEN1. Previous studies have shown that these genes are all located in the 17q12 amplification region and are often co-amplified with HER2. GRB7 can enhance downstream HER2 signal transduction, STARD3 promotes cholesterol metabolism and membrane receptor stability, and MIEN1 is closely related to tumor invasion and metastasis [8]. Furthermore, this study also found that high expressions of ERBB2 are closely related to the estrogen response process, suggesting that there may be a complex interaction between HER2 and the ER signaling pathway [9]. This finding provides a theoretical basis for combined endocrine therapy for HER2-positive breast cancer [10]. The transcriptome expression data and clinical information used in this study were all obtained from the TCGA database. Although TCGA has advantages such as large sample size and high data quality, it has not yet been validated using external databases such as GEO and METABRI, so this study still has limitations in terms of data sources.

In summary, the ERBB2-GRB7-STARD3-PGAP3-MIEN1 core module discovered in this study may constitute an important regulatory center for HER2-positive breast cancer. Future research should delve into the functional network of HER2 amplicon from multiple perspectives such as signal transduction, metabolic reprogramming, invasion and metastasis, and treatment resistance, in order to discover new molecular markers and potential therapeutic targets and provide a new theoretical basis and research direction for precision treatment of HER2-positive breast cancer.

5. Conclusion

This study systematically analyzed the expression characteristics and potential molecular mechanisms of ERBB2 in breast cancer based on transcriptomic and clinical data from the TCGA-BRCA database. Differential expression analysis identified 334 significantly differentially expressed genes. GO enrichment analysis showed that ERBB2 is mainly involved in biological processes such as the ERBB2 signaling pathway, tyrosine phosphorylation, estrogen response, and nuclear receptor-mediated signal transduction. Hallmark pathway analysis further revealed its close association with tumor-related pathways such as estrogen response, lipid metabolism, and oxidative phosphorylation. PPI network analysis revealed that ERBB2, GRB7, STARD3, PGAP3, and MIEN1 constitute a highly clustered core module, suggesting their potential joint involvement in the occurrence and progression of breast cancer.

These results indicate that the role of ERBB2 in breast cancer is not limited to the level of a single gene but rather forms a synergistic regulatory network with multiple co-amplified genes in the 17q12 chromosomal region, jointly influencing tumor cell proliferation, metabolic reprogramming, and invasion and metastasis. This study not only deepens our understanding of the molecular mechanisms of HER2-positive breast cancer but also provides a theoretical basis for identifying new diagnostic biomarkers and potential therapeutic targets, offering valuable insights for precision medicine research in breast cancer.

This study primarily relied on bioinformatics analysis based on public databases and lacks experimental and clinical sample validation. Future research could combine cell experiments, multi-omics data, and clinical studies to further explore the synergistic mechanisms between ERBB2 and

core genes such as GRB7, STARD3, and MIEN1, providing new research ideas for precision treatment of HER2-positive breast cancer.

References

[1] Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, Jemal A, Siegel RL. Breast Cancer Statistics, 2022. CA Cancer J Clin. 2022 Nov;72(6):524-541. doi: 10.3322/caac.21754. Epub 2022. PMID: 36190501.

[2] Marchiò C, Annaratone L, Marques A, Casorzo L, Berrino E, Sapino A. Evolving concepts in HER2 evaluation in breast cancer: Heterogeneity, HER2-low carcinomas and beyond. Semin Cancer Biol. 2021 Jul;72:123-135. doi: 10.1016/j.semcancer.2020.02.016. Epub 2020. PMID: 32112814.

[3] Kunte S, Abraham J, Montero AJ. Novel HER2-targeted therapies for HER2-positive metastatic breast cancer. Cancer. 2020 Oct 1;126(19):4278-4288. doi: 10.1002/cncr.33102. Epub 2020. PMID: 32721042.

[4] DeBusk K, Abeysinghe S, Vickers A, Nangia A, Bell J, Ike C, Forero-Torres A, Blahna MT. Efficacy of tucatinib for HER2-positive metastatic breast cancer after HER2-targeted therapy: a network meta-analysis. Future Oncol. 2021 Nov;17(33):4635-4647. doi: 10.2217/fon-2021-0742. Epub 2021. PMID: 34463120.

[5] Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987;235(4785):177-82. doi: 10.1126/science.3798106. PMID: 3798106.

[6] Raghav KPS, Moasser MM. Molecular Pathways and Mechanisms of HER2 in Cancer Therapy. Clin Cancer Res. 2023;29(13):2351-2361. doi: 10.1158/1078-0432.CCR-22-0283. PMID: 36574481; PMCID: PMC10293474.

[7] Pal SK, Pegram M. Targeting HER2 Epitopes. Semin Oncol. 2006;33(4):386-91. doi: 10.1053/j.seminoncol.2006.04.004. PMID: 16890794.

[8] Kauraniemi P, Kallioniemi A. Activation of multiple cancer-associated genes at the ERBB2 amplicon in breast cancer. Endocr Relat Cancer. 2006;13(1):39-49. doi: 10.1677/erc.1.01147. PMID: 16601278.

[9] Provenzano L, Dieci MV, Curigliano G, Giuliano M, Botticelli A, Lambertini M, Rizzo G, Pedersini R, Sirico M, La Verde N, Gennari A, Zambelli A, Toss A, Piras M, Giordano M, Tagliaferri B, Generali D, Sartori D, Miliziano D, Menichetti A, Ligorio F, Zurlo C, Griguolo G, Berton Giachetti PP, Faso V, Corti C, Chiappe E, Scagnoli S, Pisegna S, Capasso C, De Angelis C, Arpino G, Criscitiello C, Guarneri V, Pruneri G, Mariani L, Vernieri C; PALMARES-2 Study Group. Real-world effectiveness comparison of first-line palbociclib, ribociclib or abemaciclib plus endocrine therapy in advanced HR-positive/HER2-negative BC patients: results from the multicenter PALMARES-2 study. Ann Oncol. 2025;36(7):762-774. doi: 10.1016/j.annonc.2025.03.023. Epub 2025. PMID: 40204155.

[10] Cucciniello L, Garufi G, Di Rienzo R, Martinelli C, Pavone G, Giuliano M, Arpino G, Montemurro F, Del Mastro L, De Laurentiis M, Puglisi F. Estrogen deprivation effects of endocrine therapy in breast cancer patients: Incidence, management and outcome. Cancer Treat Rev. 2023;120:102624. doi: 10.1016/j.ctrv.2023.102624. Epub 2023. PMID: 37751658.