Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/2025.AU28758

Research Advances in Biomarkers and Prognostic Analysis for Breast Cancer Patients

Yongxi Xu

Cancer Institute, University College London, London, United Kingdom
wegmxuq@ucl.ac.uk

Abstract. Breast cancer exhibits biological heterogeneity, with prognosis significantly influenced by molecular subtypes, genetic alterations, and treatment types. Through further research, two types of translational biomarkers applicable within breast cancer patients have emerged: 1) pre-treatment tissue-based transcriptomic genes that encode intrinsic tumor biology and 2) post-treatment blood-based MRD markers (ctDNA) that capture residual systemic risk. This review establishes logical connections between these two categories, detailing the process of constructing pathways from treatment plan customization to prognostic follow-up assessment. We used literature analysis and comparative methods to extract key points from three research approaches, including data sources, analytical methods, and conclusions, then synthesized and connected them: a survival ranking of significantly associated genes within chemotherapy-treated ER+/HER2- and basal groups, a multivariable prognostic model constructed based on genes in TCGA dataset and a meta-analysis on ctDNA. We identified gaps between studies, how their findings complement each other, and ultimately provided critical progress toward realizing an implementable, end-to-end clinical treatment pathway: Early decision-making based on prognostic biomarkers, followed by ctDNA-guided dynamic risk assessment.

Keywords: breast cancer, biomarkers, ctDNA, prognosis outcome

1. Introduction

Breast cancer has topped the list of the most common cancers among females in 157 countries as the leading cancer source, while this condition has also topped the disease population cohort by its high incidence rate among them. In 2022, worldwide cases of diagnosis in females due to breast cancer accounted for about 2.3 million and resulted in around 670,000 deaths [1]. Breast cancer is a heterogeneous disease where differences between patients and various progression patterns of disease stages complicate the tailoring of precise treatments for this condition [2]. Patients have been classified as having specific molecular subtypes, such as ER/PR/HER2 in clinical practice; however, it remains undetermined whether they will respond differently to therapies or if there are different genotypic profiles between them and untreated breast cancer [3]. The establishment of suitable models to assess breast cancer is challenging, given that so many factors introduce disease heterogeneity and change over time. Beyond assessing risk, whether an evaluation of post-treatment breast cancer patients improves their prognostic outcomes or if there is rigidity in prognosis

DOI: 10.54254/2753-8818/2025.AU28758

assessments still needs to be determined. In order to improve the current situation, studies are being conducted on creating a correct pathway for the treatment of breast cancer, including the initiation of the risk evaluation before the start of the treatment and designing the personal and precise treatments according to each breast cancer's manifestation, followed by the final evaluation of physical status after the end of the treatment. This review brings together all current results to do a brief summarization about the related existing scientific results on the key points of the whole breast cancer treatment journey, which will provide valuable insights for prognostic research in breast cancer patients by clarifying the process for establishing comprehensive treatment protocols in clinical practice. For prognostic researchers, this review clearly outlines the breakthroughs and limitations of different types of prognostic biomarker studies, offering direction for future research.

2. Identification of specific biomarkers and application of statistical models

2.1. Significantly associated genes across different molecular subtypes: RPL22, TGT3, CAMSAP1, and PDLIM7

Győrffy B. categorized the breast cancer patient cohort based on molecular subtypes and treatment types received: the ER+/HER2- group receiving chemotherapy; the basal group receiving chemotherapy; the ER+/HER2- group excluded from systemic therapy (e.g., chemotherapy, to broaden the range of gene screening and enhance the persuasiveness of the results) and the basal group receiving adjuvant therapy. These groups were compared with the untreated group [4]. The Cox regression model was used to identify genes significantly associated with the prognosis effect across the four groups. The gene database was then ranked in descending order of significance. Lastly, pathway enrichment analysis was performed on these genes to identify the biological processes they represent.

Through screening, researchers confirmed that for breast cancer patients with the same molecular subtype, the treatment types significantly influence the associated genes: only 8.9% overlapped between the ER+/HER2- group without systemic therapy and the chemotherapy group, while the basal group without systemic therapy showed almost no overlap with the chemotherapy group (WARS, UBE2L6) [4]. More importantly, the significantly associated genes also differed across patients with distinct molecular subtypes. In the chemotherapy-treated ER+/HER2- group, high expression of the RPL22 gene was significantly associated with better recurrence-free survival, while high expression of the TGT3 gene was significantly associated with poorer overall survival [4]. In the chemotherapy-treated basal group, high expression of the CAMSAP1 and PDLIM7 genes was significantly associated with poorer overall survival. There was almost no overlap in the top-ranked significant genes between the two groups.

The conclusions drawn in this paper have substantial practical applicability: The pathway enrichment analysis on the significant genes from the chemotherapy-treated ER+/HER2- and basal groups identify which biologically over-represented processes the significant genes belong to. The findings ultimately revealed that ER+/HER2- patients exhibit heightened proliferation pathway activity, which means they are more likely to have poor prognosis, and greater susceptibility to chemotherapy. Therefore, risk stratification is necessary for ER+/HER2- patients: those with high scores respond better to chemotherapy, while low-score patients should have alternative therapies such as endocrine treatment. Basal patients demonstrate elevated activity in immune-related pathways and can achieve improved outcomes through chemotherapy-assisted immunogenic clearance [4].

DOI: 10.54254/2753-8818/2025.AU28758

These researchers have performed an extensive gene analysis on ER+/HER2- and basal molecular subtypes of the tumor, finding genes which had significant correlations to post-chemotherapy prognosis outcome. Following that, they correlated them with the follow-up data of the patients in order to formulate an accurate treatment scheme for every single molecular subtype.

2.2. Seven significant genes in the TCGA database and the establishment of a risk assessment model

Győrffy B. successfully provided a detailed analysis of the extent to which patients with specific molecular subtypes respond to chemotherapy, offering the initial framework for personalized treatment plans for patients with these two molecular subtypes. However, the range of the patient population and treatment types studied remains relatively narrow, failing to fully provide a strong foundation for improving the prognosis of breast cancer patients. Expanding the capacity of the analyzed gene database and correlating it with more comprehensive follow-up processes holds significant importance for developing more effective and safer treatment plans tailored to breast cancer patients. This approach aims to reduce the harm associated with breast cancer as a heterogeneous disease—a condition characterized by varying pathological features among individual patients, such as distinct molecular subtypes and genomic mutations, which creates substantial challenges in designing universally effective treatment strategies.

Liu L. expanded the range of data analysis and established a larger data model based on this foundation. Researchers downloaded 631 gene sample datasets from the TCGA database. They first compared tumor tissues with surrounding healthy tissues, using significance tests to identify pathogenic genes with markedly different expressions. This initial gene exclusion layer reduced interference from redundant genes caused by the massive dataset [5]. Subsequently, univariate Cox regression was used to evaluate the relationships between the expression levels of these key pathogenic genes and patients' final survival times. Due to the limitation of univariate Cox regression in controlling for other variables, the authors took the significantly associated results from the previous step and analyzed them again using multivariate Cox regression. This ensured that potentially confounding factors such as patient age and tumor stage were controlled. Ultimately, seven genes were found to be significantly associated with overall survival in breast cancer patients (TMEM190, LYVE1, LILRB5, RPL22, PDLIM7, CAMSAP1, CD209) [5].

Unlike the first paper, researchers not only found significantly associated genes but also mathematically integrated these seven genes to establish a risk scoring model. In this model, the risk score is calculated by summing the products of each significant gene's regression coefficient in the multivariate Cox regression and its expression value in patients. Patients are then categorized into high-risk and low-risk groups based on their scores. Finally, the Kaplan-Meier curve was used to validate the practical effectiveness of this model.

A major innovation of the study by Liu is the creation of a data model for the evaluation of patient risk by utilizing a larger database. Even though the model still needs to be enlarged and validated under varied clinical conditions, its beginning makes a valuable start point and provides a useful guideline to reach a generalized risk evaluation in breast cancer patients so as to adopt appropriate treatment for those who have high risk but avoid overtreatment for those who are low risk.

DOI: 10.54254/2753-8818/2025.AU28758

2.3. ctDNA detection: significantly associated with patient status after treatment

Beyond analyzing genes significantly associated with overall survival to enable more precise personalized treatment for patients, accurately assessing their physical condition post-treatment and providing better stabilized treatment plans accordingly is also crucial for extending overall survival and improving prognosis. Although the previous two papers provided a universal model and deep stratification for precision treatment evaluation in breast cancer patients, researchers did not link significant genes to post-treatment outcomes. This is also closely related to breast cancer's nature as a heterogeneous disease: not only do pathological manifestations vary between patients, but the same disease may present completely differently at different stages within the same patient. In other words, a patient's gene expression may differ before and after treatment. The previous two papers analyzed only pre- or post-treatment gene data, failing to analyze the interaction between pre- and post-treatment gene changes. This suggests that a patient's own gene data may not fully detect their physical state after treatment.

Papakonstantinou A. developed a way of using ctDNA levels to predict the treatment outcome in early-stage breast cancer patients receiving neoadjuvant therapy (NAT). He marked three time points: pre-NAT, during NAT and post-NAT. It was found that ctDNA levels before or after the treatment were linked to a high risk of relapse; however, the HR on OS was 19.1 when it came before the treatment, indicating that ctDNA detection at this stage had a big effect on poor OS [6]. On the contrary, there was no apparent connection between post-screening ctDNA status and the possibility of achieving pCR. Before ctDNA research began, it was a common practice among doctors to define pCR as the goal to reach a complete pathologic clearance from the disease on behalf of the patient, one of the most desired circumstances in prognosis. And yet pCR has been challenged by the emergence of evidence that revealed the association of ctDNA and the overall survival rate of patients, which pointed out that achieving pCR did not necessarily imply a prognosis so favorable as indicated before. What’s more, although pCR symbolizes local eradication, it cannot detect the remaining malignancy hidden elsewhere in the body. Thus, conducting ctDNA study along with conducting pCR measurement together can offer a more thorough view about the patient and the condition in these two different aspects - local pathologic clearance and systematic molecular residual and it will probably help form a personalized and specific treatment plan for the patient.

3. Conclusion

This review highlights major findings of biomarker identification to improve prognosis in breast cancer patients and identifies 3 notable advances as follows: a. chemo-treated patients with different molecular subtypes exhibit different genes that correlate with prognosis, thus revealing which molecular subtype requires chemo-treatment through risk scoring. b. Seven genes highly linked to breast cancer have been uncovered via large databases. A risk model was designed based on these genes, which enabled the determination of breast cancer patients’ risks independent of the type of therapy and molecular subtype. c. Post-treatment monitoring using ctDNA enables the determination of the physical status of the patient and thus to design the personalized treatment plan accordingly as well as post-treatment stabilization. As such, patients’ outcome was greatly improved.

However, there is still space for improvement in this review: although it provides a fairly complete review of different gene selections and standardizations carried out in each of these 3 articles, it focuses mostly on the biological meaning of the genes found by the statistical test, rather than explaining the concrete steps and methods. Moreover, no comparative analysis was conducted

DOI: 10.54254/2753-8818/2025.AU28758

for articles studying the same genes. Many advancements have been made in improving the prognosis of cancer patients. However, more work still needs to be done to extend patient-based data from multiple cancers and different treatment types to construct large-scale risk assessment data models and validate them with more data. In addition, while early ctDNA evidence looks very promising, uniform sampling standards should continue to be established for the current age of translational medicine for even greater robustness.

References

[1] Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. The Breast [Internet]. 2022 Sep; 66(66): 15–23. Available from: https: //www.sciencedirect.com/science/article/pii/S0960977622001448

[2] Guo L, Kong D, Liu J, Zhan L, Luo L, Zheng W, et al. Breast cancer heterogeneity and its implication in personalized precision therapy. Experimental Hematology & Oncology [Internet]. 2023 Jan 9; 12(1): 3. Available from: https: //pubmed.ncbi.nlm.nih.gov/36624542/

[3] Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanislawek A. Breast cancer—epidemiology, Risk factors, classification, Prognostic markers, and Current Treatment Strategies—an Updated Review. Cancers [Internet]. 2021; 13(17): 4287. Available from: https: //pmc.ncbi.nlm.nih.gov/articles/PMC8428369/

[4] Győrffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. Computational and Structural Biotechnology Journal. 2021; 19: 4101–9

[5] Liu L, Chen Z, Shi W, Liu H, Pang W. Breast cancer survival prediction using seven prognostic biomarker genes. Oncology Letters. 2019 Jul 18

[6] Papakonstantinou A, Gonzalez NS, Pimentel I, Suñol A, Zamora E, Ortiz C, et al. Prognostic value of ctDNA detection in patients with early breast cancer undergoing neoadjuvant therapy: A systematic review and meta-analysis. Cancer Treatment Reviews [Internet]. 2022 Mar 1; 104: 102362. Available from: https: //www.sciencedirect.com/science/article/pii/S0305737222000263?via%3Dihub