

CORRELATION OF GENOMIC INDICATORS WITH HORMONE RECEPTOR STATUS IN BREAST CANCER IN THE INDIAN POPULATION DATASET

Reena Lokare^{1,*}, Bharat Kulkarni², and Mansing Rathod¹

Received: April 23, 2025; Revised: May 14, 2025; Accepted: May 21, 2025

Abstract

Breast cancer is a highly heterogeneous disease characterized by gene alterations and changing hormone levels. Breast cancer require hormones to grow, which act through specific receptors on tumour cell surface such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). Based on the receptor status, breast cancer can be classified into ER-positive tumours and ER-negative tumours, PR-positive tumours and PR-negative tumours and HER2-positive tumours and HER2-negative tumours. Understanding the relationship between genomic indicators and hormone receptor status is crucial for improving diagnosis, prognosis, and personalized treatment strategies. This study investigates the correlation between key genomic alterations and hormone receptor status in breast cancer patients from the Indian population. With the help of genomic data derived from tumour samples, we analysed mutation frequencies and mutation patterns across 11 genomic indicators-ESR1 (Estrogen Receptor 1), CDH1 (Cadherin 1), TP53(Tumour Protein p53), CCND1(Cyclin D1), PIK3CA(Phosphatidylinositol-4, 5-bisphosphate 3-kinase catalytic subunit alpha), PTEN(Phosphatase and Tensin Homolog), GATA3(GATA Binding Protein 3), ERBB2(Erb-B2 Receptor Tyrosine Kinase 2 (also known as HER2)), FGFR1(Fibroblast Growth Factor Receptor 1), FOXA1(Forkhead Box A1) and NF1(Neurofibromin 1), and also evaluated their association with hormone receptor expression. It is observed that genes TP53, PTEN, ERBB2 and FOXA1 have overall high mutation frequency in all three ER, PR and HER2 positive breast cancer tumour samples. TP53 mutations were observed in overall 66% samples. PTEN mutations were observed in 63% overall samples. ERBB2 mutations were observed in 57% overall samples and FOXA1 mutations were observed in 48% overall samples. This study highlights the utility of 11 genomic biomarkers in guiding personalized therapy to Indian breast cancer patients.

Keywords: Breast Cancer; Correlation; Genomic Indicators; Hormone Receptor Status; Indian Population; Precision Medicine

¹ Department of Information Technology, K. J. Somaiya Institute of Technology, Somaiya Ayurvihar Complex, Eastern Express Highway, Near Everard Nagar, Sion (East), Mumbai – 400022, Maharashtra, India. E-mail: reena.l@somaiya.edu^{*}; rathodm@somaiya.edu

² Padmabhooshan Vasantraodada Patil Institute of Technology (PVPIT), Sangli-Tasgaon Road, A/P Budhgaon, Taluka Miraj, District Sangli, Maharashtra – 416304, India. E-mail: bk.2983@gmail.com

^{*} Corresponding author

DOI: <https://doi.org/10.55766/sujst9547>

Introduction

Background

Breast cancer is one of the most prevalent cancers affecting women worldwide and is a leading cause of cancer-related mortality. In recent decades, research has significantly advanced our understanding of breast cancer as a heterogeneous and complex disease with multiple subtypes characterized by distinct molecular features, clinical behaviour, and therapeutic responses (Jemal *et al.*, 2011; Bray *et al.*, 2018). Among the most clinically significant biomarkers used for the classification and treatment planning of breast cancer are estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) (Carey *et al.*, 2006) status. Based on these hormone receptor status, breast cancer is classified into ER+/-, PR+/-, and HER2+/- subtypes. Such classification helps in accurate treatment recommendation considering complex nature of the disease.

These biomarkers guide in recommendation of therapies like endocrine therapy or targeted therapies and also aid in predicting disease prognosis. Understanding the molecular nature of the disease and tailoring treatment based on breast cancer subtypes supports a personalized treatment approach which increases the likelihood of patient survival by providing more suitable treatments from the outset and avoiding the time often wasted in trying first-line, second-line therapies, and so on (Gillies *et al.*, 2016; Denkert *et al.*, 2017). In case of breast cancer which is mostly detected in advanced stage, significant time is wasted in understanding suitable therapy for individual patient.

Despite considerable progress, population-specific genomic insights remain underexplored particularly in the Indian subcontinent, where breast cancer incidence is rising and often presents with an invasive nature at a younger age. The genetic and environmental backgrounds of Indian breast cancer patients may contribute to distinct biomarker profiles, mutation spectra, and therapeutic responses. Hence, understanding how genomic alterations correlate with ER, PR, and HER2 status in this population is essential for developing precision medicine strategies tailored to Indian patients (Sahasrabudde *et al.*, 2021).

In order to recommend suitable treatment plan to patients, knowing the subtype of cancer is essential. Once the subtype is determined, a study related to genomic biomarkers can be done and accurate treatment can be recommended. These biomarkers are discussed in the following section.

Genomic Biomarkers and their Relevance in Breast Cancer

The presence, absence, or altered expression of genes can influence tumour behaviour. The most frequently mutated genes in breast cancer include TP53, PIK3CA, CDH1, GATA3, ERBB2, FOXA1, ESR1, and PTEN (Lokare and Patil, 2022). These genes are considered as biomarkers or genomic indicators in this study. These genes are involved in critical cellular processes such as DNA (Deoxyribonucleic Acid) repair, cell proliferation, apoptosis, and signal transduction. For instance, TP53 mutations are commonly associated with triple-negative breast cancer (TNBC), while PIK3CA mutations typically appear in ER-positive subtypes and are linked to activation of the PI3K-AKT(Phosphoinositide 3-Kinase- Protein Kinase B (PKB)) pathway. Several studies conducted in Western populations have demonstrated correlations between genomic mutations and hormone receptor status. However, such findings may not be universally applicable to Indian populations due to differences in ethnicity, lifestyle, environmental exposures, and genetic background. Therefore, population-specific genomic profiling is essential to uncover novel or enriched biomarkers that may influence receptor status, treatment response, and survival outcomes (Lokare and Patil, 2022).

Once genomic biomarkers are understood in tumour samples, hormone receptor status of the patient can be studied. On the basis of breast cancer subtype suitable treatment plan is decided for patients as discussed in next section.

Relevance of Hormone Receptors in Clinical Practice

Hormones like estrogen, progesterone, etc. are playing major role in cancer. Cancers grow in response to these hormones. Breast cancer cells have proteins as receptors. ER and PR are nuclear hormone receptors that, when activated, promote the proliferation of breast epithelial cells. The presence of these receptors particularly ER indicates potential responsiveness to hormonal therapies such as tamoxifen or aromatase inhibitors. HER2, a member of the Epidermal Growth Factor Receptor (EGFR) family, is a membrane-bound tyrosine kinase receptor that facilitates cell division and survival when overexpressed or amplified. While HER2 positivity can be effectively treated with monoclonal antibodies like trastuzumab, it typically

indicates more aggressive disease course. Understanding the molecular determinants of ER, PR, and HER2 status at the genomic level provides crucial insights into disease mechanisms, supports accurate subtype classification, and enables the development of more effective and individualized treatment protocols. Furthermore, identifying genomic correlates of receptor status may enhance diagnostic precision and aid in the discovery of novel targeted therapies, especially for hormone receptor-negative or HER2-enriched tumours (Malvia *et al.*, 2019; Mohanty *et al.*, 2022; Veeraraghavan *et al.*, 2025).

Due to diversity in the lifestyle, geographical differences, same treatment plan cannot be offered to patients in different countries. Recently Indian patients genomic profile was published in CBioPortal- a portal containing nearly 32 breast cancer type tumour profiles with data visualization and analysis facility. This study focuses on Indian patient's dataset which is discussed in next section.

Indian Breast Cancer Genomic Landscape

Breast cancer in India exhibits unique characteristics, including earlier age of onset, higher prevalence of Triple negative breast cancer (TNBC) in which all three hormones-ER, PR and HER2 are absent, and lower levels of awareness and screening rates. These factors contribute to delayed diagnosis and poorer clinical outcomes. In response, recent initiatives such as the Indian Breast Cancer Genomic Atlas (BCGA) have begun to collect valuable genomic data from Indian patients (Cerami *et al.*, 2012; Gao *et al.*, 2013; de Bruijn *et al.*, 2023). This presents a novel opportunity to investigate mutation profiles, pathway involvement, and gene-receptor associations in a regionally relevant cohort. Despite these advances, studies examining how specific gene mutations correlate with hormone receptor and HER2 status in Indian patients remain limited. Analysing these relationships can enhance early detection, enable stratified treatment approaches, and potentially uncover novel therapeutic targets specific to the Indian population.

Aim of the Study

This study aims to investigate the correlation between genomic biomarkers and the status of ER, PR, and HER2 receptors using an Indian breast cancer genomic dataset. Specifically, data from the BCGA-IITM 2025 cohort which includes 274 breast cancer tumour samples with detailed molecular profiles of breast cancer samples, is utilized. The reason behind selecting this dataset is it includes the genomic indicators in this study. Also, the dataset has features such as receptor status like ER, PR and HER2. These samples contain both clinical and

genomic data. Such molecular details of the cancer help in precision oncology. The specific objectives of the study are given below and are discussed in detail in the results section:

A. Identifying frequently mutated genes in the Indian dataset.

B. Determine the correlation between gene mutations and ER, PR, HER2 hormone expression.

C. Explore pathway associations and mutual exclusivity among key mutated genes.

D. Compare gene mutation profiles between Indian and U.S. breast cancer patients.

This research contributes to the expanding body of precision oncology studies in India and underscores the importance of biomarker-guided treatment planning based on population-specific genomic data.

Literature Review

Breast cancer remains a leading cause of cancer-related morbidity and mortality among women globally. Numerous studies have established regarding influence of gene alterations on disease progression, prognosis, and therapeutic response in breast cancer. Key hormone receptors Estrogen Receptor (ER), Progesterone Receptor (PR), and Human Epidermal Growth Factor Receptor 2 (HER2) are critical for breast cancer classification and treatment decisions. Landmark studies conducted by The Cancer Genome Atlas (TCGA) and METABRIC have explored the mutational landscape of breast cancer, identifying frequently altered genes including TP53, PIK3CA, GATA3, CDH1, PTEN, and ERBB2. These mutations have been shown to correlate with hormone receptor expression, thereby influencing subtype stratification and treatment response (The Cancer Genome Atlas Network, 2012; Curtis *et al.*, 2012; Toss *et al.*, 2018).

However, a significant gap exists in literature focused regionally, mainly on Indian patient populations. Studies by Kumar *et al.* (2024) and Anand *et al.* (2022) have shown that Indian breast cancer patients often exhibit a distinct mutational profile, with higher frequency of TP53 and ERBB2 mutations compared to Western cohorts. Moreover, Indian patients frequently present at younger ages and more advanced stages, potentially driven by genetic, lifestyle, and diagnostic disparities (Kumar *et al.*, 2024; Anand *et al.*, 2022).

Despite the findings by several studies discussed, publicly available genomic datasets from Indian patients remain limited, and few studies have directly examined the relationship between specific gene mutations and hormone receptor status within

this population. Investigating mutual exclusivity, pathway enrichment, and cross-cohort comparisons with global datasets can deepen our understanding of breast cancer heterogeneity among Indian women (Chikkala *et al.*, 2024).

This research addresses the existing gap by analysing the correlation between genomic indicators and hormone receptor expression (ER, PR, HER2) in an Indian breast cancer dataset, offering novel insights into regional molecular characteristics and potentially guiding personalized treatment approaches (Manjunath and Kaur, 2022).

Methodology

This section discusses methods used to obtain the results in the study such as gene mutation percentage, heat map, etc.

CBioPortal

CBioPortal is a portal which provides genomic, clinical and proteomic data of cancers. It provides nearly 32 breast cancer datasets. The tool has a provision to select dataset and submit query for targeted genes. It provides many data visualization tools like oncoprint, bar charts, pathway analysis, etc. The oncoprint is a heat map which shows overall gene mutation frequency as well as type of gene mutations in the dataset.

Various pathways are provided by the portal which allows to analyse the signalling cascade in a cell. It also gives the gene mutation frequency. The portal also has a provision to check mutually exclusive genes for more detailed analysis.

Python

Python programming language has rich set of built-in libraries for data analysis, machine learning algorithms, etc. The dataset used in the study has features such as breast cancer type, age, gender, receptor status, altered and non-altered genes, etc. The dataset has gene attribute having two possible values: gene mutated or not. It doesn't have information such as type of mutation, etc. The calculation of gene mutation frequency as per breast cancer subtype is done using built-in libraries of python. Also, the feature importance plot is obtained in python. The Sankey diagram which shows relation between mutually exclusive genes is also obtained in python.

As CBioPortal has very limited functionalities for doing data analysis and data visualization, python language is used for further analysis.

Results

The results of the objectives are discussed in this section.

Identifying Frequently Mutated Genes in the Indian Dataset

In our previous research, we identified a set of 11 genomic biomarkers using multi-omics hormonal therapy failure dataset of breast cancer patients available on CBioPortal with the help of machine learning algorithm such as feature importance given by random forest algorithm (Jemal *et al.*, 2011; Bray *et al.*, 2018). These biomarkers include 11 genomic biomarkers such as

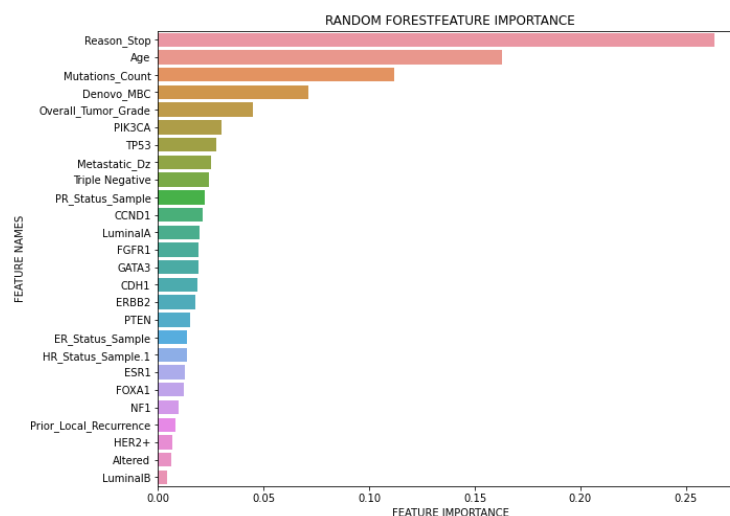
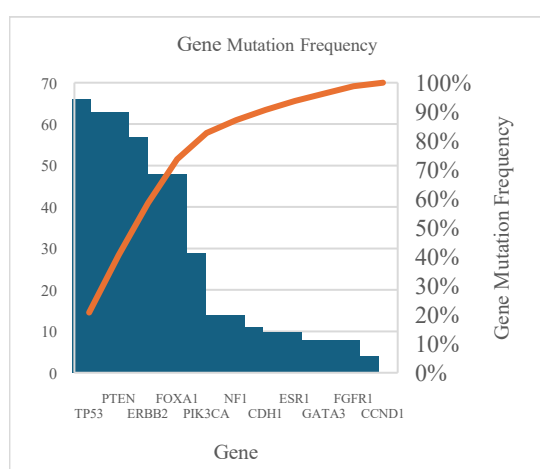


Figure 1. Feature importance of random forest algorithm

Table 1. Mutation Percentage

Sr. No.	Gene	Overall Mutation (%)	ER	PR	HER2
1	ESR1	10	66	66	69
2	CDH1	11	75	75	67
3	TP53	66	407	397	401
4	GATA3	8	63	62	40
5	CCND1	4	26	26	30
6	PTEN	63	416	409	407
7	ERBB2	57	380	373	351
8	FGFR1	8	53	53	46
9	FOXA1	48	315	307	309
10	PIK3CA	29	205	202	172
11	NF1	14	89	89	86

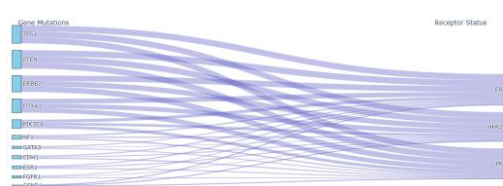
**Figure 2. Gene mutation frequency in Indian breast cancer genomic dataset**

ESR1, CDH1, TP53, GATA3, CCND1, PTEN, ERBB2, FGFR1, FOXA1, PIK3CA, and NF1, along with several clinical biomarkers. For the current study, only the previously identified genomic biomarkers were utilized. Figure 1 presents the feature importance plot generated using the random forest algorithm, which served as the primary reference for selecting the genomic biomarkers. Clinical biomarkers shown in the figure are currently omitted in this research.

Determining the Correlation between Gene Mutations and ER, PR, HER2 Expression

The Indian genomic dataset used in this study was analysed to investigate mutations in 11 selected genomic biomarkers, as shown in Table 1. The data is represented in diagrammatic form in Figure 2 to facilitate further analysis and better understanding of the disease's complexity and the impact of gene mutations on treatment outcomes.

The bar chart in Figure 2 represents the gene mutation frequencies from the Indian breast cancer patient dataset.

**Figure 3. Sankey diagram showing gene mutations in breast cancer receptor status****Table 2. Mutually exclusive genes**

Gene	Mutually Exclusive Gene
FOXA1	PIK3CA
PTEN	PIK3CA
ERBB2	PIK3CA
CDH1	FOXA1
GATA3	PTEN
GATA3	FOXA1
CCND1	FOXA1
CCND1	PTEN
FGFR1	FOXA1
GATA3	ERBB2
TP53	GATA3
CDH1	PTEN
CCND1	ERBB2
CDH1	ERBB2
FOXA1	NF1
GATA3	NF1
PTEN	FGFR1
ERBB2	NF1
ESR1	FOXA1

Figure 3 is a Sankey diagram illustrating the relationship between gene mutations and breast cancer receptor status. Such visualization helps in determining the correlation between genes and breast cancer receptor such as ER, PR and HER2.

Figure 4 is a oncoprint which is a heat map showing visual summary of genomic alterations.

Exploring Pathway Associations and Mutual Exclusivity Among Key Mutated Genes

An analysis of mutually exclusive genes was conducted on breast cancer tumour samples on

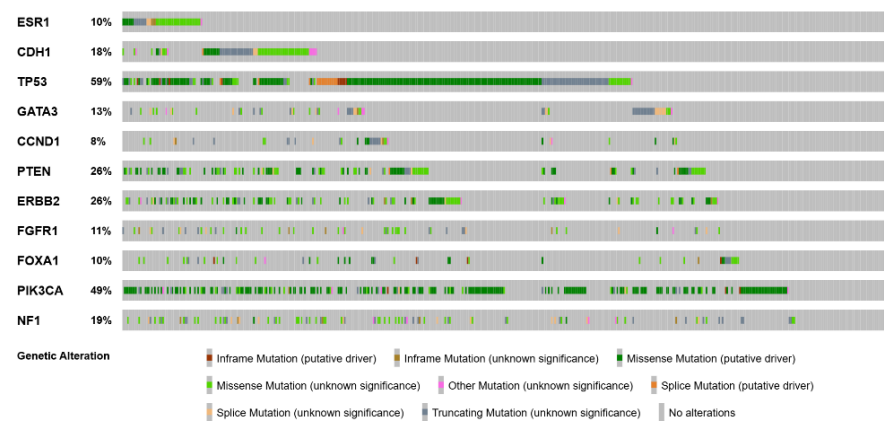


Figure 4. Gene mutation frequency in Indian patient’s dataset from cBioPortal

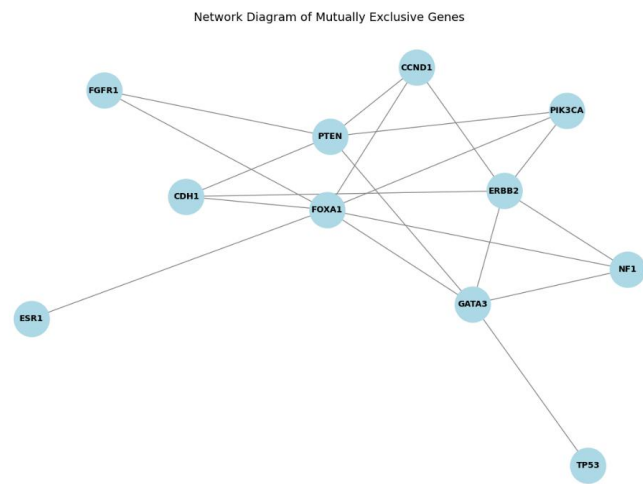


Figure 5. Network of Mutually exclusive genes in Indian breast cancer patient’s dataset

CBioPortal. Identifying such gene pairs provides deeper insights into breast cancer biology. In these cases, a mutation in one gene often implies that the other gene in the pair is not mutated.

The understanding of mutation patterns, a chain of mutually exclusive genes has been derived, as shown in Table 2. This table provides insight into gene relationships by identifying those that are mutually exclusive.

From the list in Table 3, it is evident that FOXA1 mutations drive a specific breast cancer subtype, independent of these pathways, and are likely involved in hormone regulation. The exclusivity of PTEN suggests its distinct role in regulating the PI3K/AKT pathway. HER-driven tumours tend to exclude alteration in genes such as PIK3CA, GATA3, CCND1, CDH1, and FGFR1. Among these, GATA3 is known to define a unique ER-positive subtype. The mutual exclusivity of CCND1 implies its involvement in a separate

proliferative pathway. Additionally, TP53, which is commonly mutated in triple-negative breast cancer, appears mutually exclusive with GATA3, reinforcing the distinction between ER-negative and ER-positive subtypes. A graphical representation of all mutually exclusive gene pairs is shown in Figure 5, offering a visual overview of these complex molecular relationships. These relationships are as shown in Table 3.

Table 3. List of Mutually Exclusive Genes Derived from Table 2

Gene	List of Mutually Exclusive Genes
FOXA1	PIK3CA, CDH1, GATA3, CCND1, FGFR1, NF1, ESR1
PTEN	PIK3CA, GATA3, CCND1, CDH1, FGFR1
ERBB2	PIK3CA, GATA3, CCND1, CDH1, NF1
GATA3	PTEN, FOXA1, ERBB2, NF1, TP53
CCND1	FOXA1, PTEN, ERBB2
TP53	GATA3

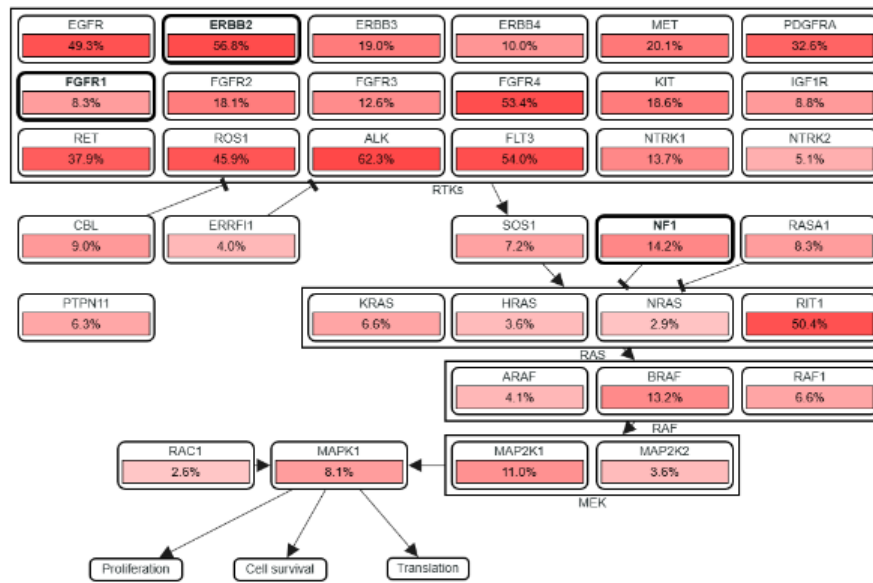


Figure 6. RTK-RAS Pathway with score: 3.00 Genes matched: ERBB2, FGFR1, NF1

The pathways show the series of actions in a cell which leads to specific outcome. Such study plays a major role in deciding target drugs, understanding tumour behaviour, etc. The following section studies such pathways available in CBioPortal to gain more understanding of the disease and possible treatment plans.

RTK-RAS –MAPK Pathway

The Receptor Tyrosine Kinase (RTK)-RAS(Rat Sarcoma)-Mitogen-Activated Protein Kinase (MAPK) pathway is frequently altered in breast cancer, as illustrated in Figure 6. It shows the genes ERBB2, FGFR1, NF1 are matched with score 3.

PI3K Pathway

Phosphoinositide 3 Kinase (PI3K) pathway plays a critical role in cell signalling, involving cell growth, survival, survival, metabolism and proliferation, and is often dysregulated in cancer. Figure 7 illustrates PIK pathway. It shows the genes PTEN PIK3CA are matched with score 2.

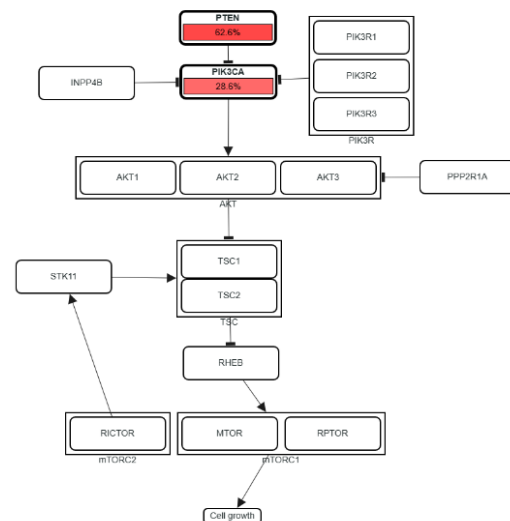


Figure 7. PI3K Pathway with score:2.00 Genes matched: PTEN, PIK3CA

Cell Cycle Pathway

The cell cycle pathway, illustrated in Figure 8 represents the signalling processes involved in cell cycle regulation. It demonstrates how gene mutations can disrupt normal cell cycle progression and ultimately lead to cancer development. It shows the genes TP53, CCND1 are matched with score 2.

TP53 Pathway

The TP53 signalling pathway for the studied dataset is illustrated in Figure 9. TP53 is a central tumour suppressor gene, with mutations observed in 66.2% of the samples, highlighting its major role in tumour progression. It shows the gene TP53 is matched with score 1.

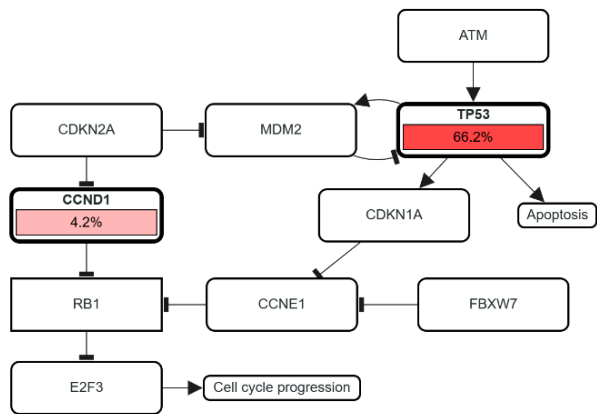


Figure 8. Cell Cycle Pathway with score:2.00 Genes matched: TP53, CCND1

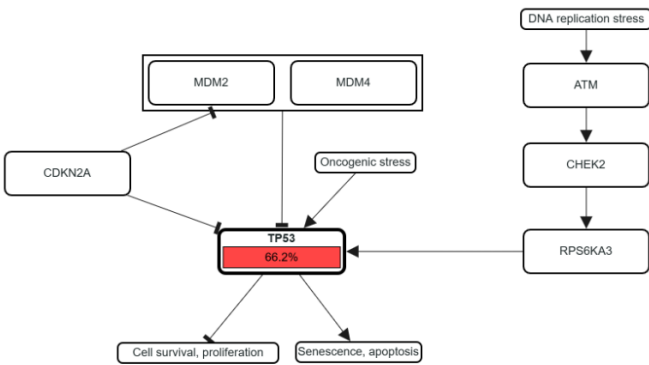


Figure 9. TP53 Pathway with score:1.00 Genes matched: TP53

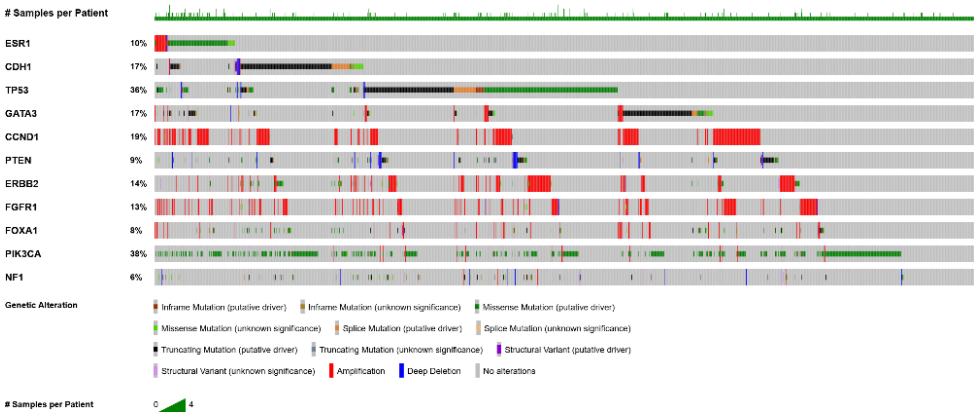


Figure 10. Gene mutation frequency in US Dataset from BioPortal

Comparing Gene Mutations in Indian Breast Cancer Patients vs. US Breast Cancer Patients

To compare gene mutation frequencies across different countries, breast cancer patient's data from India and the United States were analysed in this study. Such comparisons can offer insights onto

factors contributing to higher mutation rates and, consequently, disease prevalence. Datasets used for this analysis were both obtained from CBioPortal, and the comparison was focused on 11 genomic biomarkers. Figure 10 is a oncoprint which presents the gene mutation frequencies in Indian breast

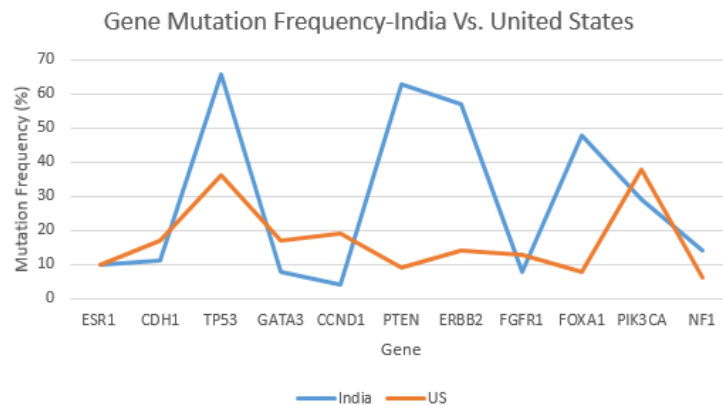


Figure 11. Gene mutation frequency Study-Indian Vs. US breast cancer patients

cancer patients, highlighting TP53 has 65% mutation frequency, PTEN has 53% mutation frequency, ERBB2 has 57% mutation frequency, and FOXA1 has 48% mutation frequency as the most frequently mutated genes. The figure also shows that missense mutations shown in green colour bars are the most common type of mutation in this cohort.

The oncoprint in Figure 11 shows gene mutation frequencies in U.S. breast cancer patients. It indicates that TP53 and PIK3CA are the most frequently mutated genes in this cohort. Additionally, gene amplifications are observed to be more common compared to the Indian dataset.

To further compare the patients across two countries, a line chart is presented in Figure 12, illustrating that U.S. patients have a lower overall number of gene mutations compared to Indian patients.

Discussions and Conclusions

This study provides a comprehensive analysis of the genomic landscape of breast cancer in the Indian population, with a specific focus on the correlation between gene mutations and hormone receptor status, as well as comparative mutational insights with the U.S. cohort. The study has used 11 genomic indicators for the research such as ESR1, CDH1, TP53, CCND1, PIK3CA, PTEN, GATA3, ERBB2, FGFR1, FOXA1 and NF1 genes.

First, in identifying frequently mutated genes, our analysis revealed a distinct pattern in the Indian dataset. The research found TP53 gene has 65% mutation frequency, PTEN has 63% mutation frequency, ERBB2 has 57% mutation frequency, FOXA1 has 48% mutation frequency, and PIK3CA

has 43% mutation frequency. This contrasts with global datasets, where PIK3CA tends to dominate mutation prevalence. The elevated frequency of tumour suppressor gene mutations (e.g., TP53, PTEN) may reflect unique biological or environmental factors influencing breast cancer in the Indian population.

Second, in examining the correlation between gene mutations and hormone receptor expression (ER, PR, HER2), our study uncovered significant associations. Table 1 presents the overall gene mutation frequency percentage, as well as the number of mutated genes across the breast cancer subtype defined by ER, PR and HER2 status. Figure 2 shows that TP53, PTEN, ERBB2 and FOXA1 are the most frequently mutated genes, while CCND1, GATA3, FGFR1, ESR1, CDH1, and NF1 exhibit lower mutation frequencies. The PIK3CA gene shows a moderate number of gene mutations. Additionally, the orange line in the chart represents the cumulative percentage of mutation frequency. A Sankey diagram in Figure 3 highlights a small number of genes (in this case, only about three) account for a large proportion of the total mutations. Specifically, four genes TP53, PTEN, ERBB2, and FOXA1 contribute to nearly 75% of the cumulative mutation frequency. It shows the gene TP53 has 1.21k samples, PTEN has 1.23k samples and ERBB2 has 1.10k samples and have thick flows toward all three receptors ER, PR, and HER2. Among these, PTEN displays the strongest association with 1.23k number of samples across all three receptors with the highest overall count. Additionally, ERBB2 shows a pronounced flow towards HER2, emphasizing its clinical relevance in HER2-positive breast cancer. The FOXA1 gene has over 300 associations with each receptor where ER has 315 samples, PR has 307 samples and HER2 has

309 samples, while GATA3 gene has 63 samples of ER, 62 samples of PR and 40 samples of HER2 and ESR1 gene has 66 samples of ER, 66 samples of PR and 69 samples of HER2 which exhibits moderate associations with ER and PR. These findings indicate TP53, PTEN and ERBB2 are key biomarkers across ER, PR, and HER2 subtypes. Understanding gene-receptor associations can help guide personalized therapy decisions. For example, ERBB2 is a critical biomarker for HER2- targeted therapies such as trastuzumab. While ESR1 plays a significant role in ER-targeted hormonal therapies like tamoxifen. Figure 4 displays gene mutation frequencies from the CBioPortal dataset. It shows that TP53 has a mutation frequency of 66%, PTEN has 63% mutation frequency, ERBB2 has 57% mutation frequency, and FOXA1 has 48% mutation frequency. In contrast, CCND1 has the lowest frequency of only 4%. This figure also illustrates the types of gene mutations, with missense mutations highlighted in green being the most common across all genes.

PIK3CA mutations were more prevalent in ER-positive tumours, supporting its known role in hormone-driven signalling pathways. Conversely, TP53 and PTEN mutations were frequently associated with ER-negative and HER2-positive subtypes, indicating a potential link to more aggressive, hormone-independent tumour behaviour. GATA3 mutations also correlated with ER-positive status, further affirming its role in luminal breast cancer biology.

Third, in exploring pathway associations and mutual exclusivity, the analysis revealed key oncogenic pathways such as PI3K/AKT/mTOR, MAPK, and cell cycle regulation being frequently altered. The mutual exclusivity analysis demonstrated that genes like FOXA1 and PIK3CA, or PTEN and FGFR1, rarely co-occurred in the same tumour, suggesting they may activate similar downstream pathways and alteration of one gene may negate the selective pressure for mutation in the other. Such exclusivity provides clues about pathway redundancy and tumour evolution mechanisms. This signalling cascade as shown in Figure 6 begins with mutations in RTK and progresses through RAS, RAF(Rapidly Accelerated Fibrosarcoma) and MAPK, ultimately affecting cell proliferation, survival, and translation. In the top RTK layer, ERBB2, ALK (Anaplastic Lymphoma Kinase), FGFR1, and RET are highly mutated, as indicated by their mutation percentages in the figure. These genes act as primary activators of the pathway, and their high mutation rates highlight their potential as biomarkers. In the middle RAS layer, NF1 a tumour suppressor negatively regulates RAS. Mutations in NF1 lead to RAS hyper

activation. KRAS, NRAS, and HRAS are oncogenes within the RAS sub-pathway, and RIT1 also shows high mutation levels, suggesting strong selective pressure in certain tumours. The MAK pathway, forming the bottom layer, promotes sustained signalling for cell growth and survival. For effective personalized therapy in breast cancer, the highly mutated genes in the top RTK layer ERBB2, ALK, FGFR1, and RET should be considered as therapeutic targets. In this pathway, three genes ERBB2, FGFR1, and NF1 matched with score 3, indicating that they not only interact at multiple points within the pathway but are also frequently mutated. This highlights their potential significance in guiding targeted treatment strategies.

PI3K pathway in Figure 7 highlights the PTEN gene with a mutation frequency of 62.6%. PTEN is a tumour suppressor gene that inhibits the PI3K pathway. Additionally, the PIK3CA gene,

which is frequently mutated in breast cancer, exhibits a mutation frequency of 28.6%. Hyper activation of PI3K signalling is a hallmark in many cancers, including endocrine-resistant breast cancer. A PIK3CA mutation indicates potential responsiveness to PI3K inhibitors. While PTEN loss could confer resistance to certain targeted therapies. In Figure 7, both PTEN and PIK3CA are shown with a score of 2. This score indicates that mutations in these genes were identified in two independent datasets or patients within the analysed cohort, suggesting a notable mutation frequency across samples.

The cell cycle pathway as shown in Figure 8 shows TP53 is a commonly mutated tumour suppressor gene across many cancers type, including breast cancer. It plays a critical role in regulating cell cycle arrest, apoptosis and DNA repair. In the dataset used for this study, TP53 mutations were observed in 66.2% of cases, underscoring its major role in cancer progression particularly in endocrine resistant breast cancer tumours. The CCND1 gene, which exhibited a 4.2% mutation frequency, is associated with uncontrolled cell cycle progression when altered. Since TP53 is a crucial gene for maintaining genomic integrity, its mutation leads to the collapse of the cell's protective mechanisms. Mutations in CCND1 and MDM2(Mouse Double Minute 2 homolog) further contribute to unchecked cellular proliferation. As shown in Figure 9, the TP53 pathway ATM (Ataxia Telangiectasia Mutated) → CHEK2 (Checkpoint Kinase 2) → RPS6KA3(Ribosomal Protein S6 Kinase A3) detects DNA replication stress. This cascade leads to the activation of TP53, which in turn triggers apoptosis. The oncogenic stress can also activate TP53, preventing cells from

becoming malignant. MDM2 and MDM4(Mouse Double Minute 4 homolog) inhibit TP53 by promoting its degradation. However, CDKN2A (Cyclin-Dependent Kinase Inhibitor 2A) can block MDM2, thereby stabilizing TP53 and preserving its tumour-suppressive function.

Finally, when comparing mutation frequencies between Indian and U.S. breast cancer patients, clear differences emerged. For example, TP53 and PTEN mutations were significantly more frequent in Indian patients, while PIK3CA mutations were more balanced or higher in the U.S. cohort. These differences could arise from a variety of factors including genetic diversity, environmental exposures, lifestyle factors, and healthcare access. The data underscore the importance of population-specific cancer genomics and caution against generalizing findings across ethnic groups without validation. Among the 11 genes analysed, TP53, PTEN, and ERBB2 shows the most notable differences in mutation frequency as shown in Figure 12. TP53 has a mutation frequency of 36%, PTEN has a mutation frequency of 9%, ERBB2 has a mutation frequency of 14%, and FOXA1 has a mutation frequency of 8% which exhibit higher mutation frequencies in the Indian cohort, while GATA3 has 17% mutation frequency and NF1 has 6% mutation frequency and display lower frequencies in both Indian and U.S. datasets. In contrast, PIK3CA has 38% mutation frequency which shows a higher mutation frequency in the U.S. cohort, suggesting the presence of different oncogenic drivers across populations. FGFR has 13% mutation frequency which shows low mutation frequency in both populations, with India reporting slightly lower levels than the U.S. Notably, FOXA1 has 8% mutation frequency which demonstrates a significant spike in mutation frequency in the Indian dataset.

Indian patients may exhibit higher gene mutation frequencies due to a combination of genetic and ethnic differences, environmental exposures, and lifestyle factors such as pollution, diet, and tobacco use. Additionally, late-stage diagnosis, limited access to early screening, and fewer personalized treatment options may result in more advanced tumours with higher mutational burden. Variations in data collection methods or sample characteristics between Indian and U.S. studies may also contribute to the observed differences. These factors collectively influence the higher prevalence of mutations seen in Indian cancer patients.

Overall, this study emphasizes the need for genomically informed breast cancer management tailored to regional and ethnic populations. The correlation of specific gene alterations with

hormone receptor status has implications for targeted therapies, such as PI3K inhibitors in ER+ PIK3CA-mutant tumours or HER2-targeted strategies in ERBB2-mutant subtypes. Moreover, understanding mutually exclusive mutations and pathway enrichment can help prioritize drug targets and guide future research into combinatorial or synthetic lethality-based therapies. The molecular profiling of tumour samples can help in deep understanding of the disease specific to particular patient. This will help in personalized recommendation of treatment to a patient. Such precision oncology will definitely help in reducing mortality. Also, the study is done using only 11 genomic indicators. Hence there is no need of whole human genome profiling. Only 11 genes can be tested using gene testing kits. This will aid in cost-effective treatment to patients.

The whole study is based on data available in CBioPortal which provides static data which is not updated in real-time. This can cause some delay in new research findings on data in the portal. Also, they provide predefined data visualization plots. Hence only simple exploratory analysis is possible. In order to perform advance data analysis, data must be exported in python or suitable programming language. They allow only predefined queries on the dataset which limits the data analysis. However, this portal is useful for the beginners or the researchers from computer background having limited understanding of biological data.

The study can further be extended to understand impact of all hormones on the breast cancer and find out causes behind abnormal hormones secretion. Also, detailed analysis of patient data can be done for drug discovery.

Acknowledgement

We thank our colleagues and the principal, Dr. Vivek Sunnapwar for supporting our research work.

Ethical Considerations

This study involved the use of publicly available genomic data from the CBioPortal for Cancer Genomics, which complies with all ethical data-sharing standards. No direct interaction with human participants was involved. For any data originally obtained from patients, informed consent was acquired by the original study authors, as stated in the CBioPortal documentation. All data were anonymized and used solely for academic research.

References

- Anand, A., Mishra, A., Damde, H., Saxena, A., Yadav, S.K., and Sharma, D. (2022). Molecular profile and clinico-pathological characteristics of breast cancer in Central India: First investigative report. *Indian Journal of Surgical Oncology*, 13(2):421–425. <https://doi.org/10.1007/s13193-022-01502-0>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424. <https://doi.org/10.3322/caac.21492>
- Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., and Millikan, R.C. (2006). Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*, 295(21):2492–2502. <https://doi.org/10.1001/jama.295.21.2492>
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5):401–404. <https://doi.org/10.1158/2159-8290.CD-12-0095>
- Chikkala, R., Bhayal, D., Rani, N., Modali, R., Bhatia, K., and Dubey, B. (2024). Mutational landscape of BRCA gene mutations in Indian breast cancer patients: Retrospective insights from a diagnostic lab. *Egyptian Journal of Medical Human Genetics*, 25(1):Article101. <https://doi.org/10.1186/s43042-024-00567-6>
- Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., METABRIC Group, Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowitz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., and Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346–352. <https://doi.org/10.1038/nature10983>
- de Bruijn, I., Kundra, R., Mastrogiacomio, B., Tran, T.N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., Abeshouse, A., Baiceanu, D., Ciftci, E., Dogrusoz, U., Dufilie, A., Erkoc, Z., Garcia Lara, E., Fu, Z., Gross, B., Haynes, C., Heath, A., Higgins, D., Jagannathan, P., Kalletla, K., Kumari, P., Lindsay, J., Lismann, A., Leenknegt, B., Lukasse, P., Madela, D., Madupuri, R., van Nierop, P., Plantalech, O., Quach, J., Resnick, A.C., Rodenburg, S.Y.A., Satravada, B.A., Schaeffer, F., Sheridan, R., Singh, J., Sirohi, R., Sumer, S.O., van Hagen, S., Wang, A., Wilson, M., Zhang, H., Zhu, K., Rusk, N., Brown, S., Lavery, J.A., Panageas, K.S., Rudolph, J.E., LeNoue-Newton, M.L., Warner, J.L., Guo, X., Hunter-Zinck, H., Yu, T.V., Pilai, S., Nichols, C., Gardos, S. M., Philip, J., AACR Project GENIE BPC Core Team, AACR Project GENIE Consortium, Kehl, K. L., Rieley, G. J., Schrag, D., Lee, J., Fiandalo, M. V., Sweeney, S.M., Pugh, T.J., Sander, C., Cerami, E., Gao, J., and Schultz, N. (2023). Analysis and visualization of longitudinal genomic and clinical data from the AACR Project GENIE Biopharma Collaborative in CBioPortal. *Cancer Research*, 83(23):3861–3867. <https://doi.org/10.1158/0008-5472.CAN-23-0816>
- Denkert, C., Liedtke, C., Tutt, A.N.J., and von Minckwitz, G. (2017). Molecular alterations in triple-negative breast cancer - the road to new treatment strategies. *The Lancet*, 389(10087):2430–2442. [https://doi.org/10.1016/S0140-6736\(16\)32454-0](https://doi.org/10.1016/S0140-6736(16)32454-0)
- Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the CBioPortal. *Science Signaling*, 6(269):pl1. <https://doi.org/10.1126/scisignal.2004088>
- Gillies, R.J., Kinahan, P.E., and Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577. <https://doi.org/10.1148/radiol.2015151169>
- Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61(2):69–90. <https://doi.org/10.3322/caac.20107>
- Kumar, R., Awasthi, S., Pradhan, D., Kumar, R., Goel, H., Singh, J., Jain, A., Meena, J., Sharma, A., Mathur, S., & Tanwar, P. (2024). Somatic mutational landscape across Indian breast cancer cases by whole exome sequencing. *Scientific Reports*, 14(1):Article18679. <https://doi.org/10.1038/s41598-024-65148-4>
- Lokare, R. and Patil, S. (2022). A Study of Genetic Mutations, Amplification, Deletion and Fusion in Endocrine Resistant Advanced Breast Cancer. In: Bhalla, S., Bedekar, M., Phalnikar, R., and Sirsikar, S. (eds). In: *Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems*. Springer, Singapore, p. 25–51. https://doi.org/10.1007/978-981-19-0863-7_3
- Malvia, S., Bagadi, S.A.R., Pradhan, D., Chintamani, C., Bhatnagar, A., Arora, D., Sarin, R., and Saxena, S. (2019). Study of gene expression profiles of breast cancers in Indian women. *Scientific Reports*, 9(1):46261. <https://doi.org/10.1038/s41598-019-46261-1>
- Manjunath, M., Nirgude, S., Mhatre, A., Vemuri, S.G., Nataraj, M., Thumsi, J., and Choudhary, B. (2022). Transcriptomic profiling of Indian breast cancer patients revealed subtype-specific mRNA and lncRNA signatures. *Frontiers in Genetics*, 13:Article932060. <https://doi.org/10.3389/fgene.2022.932060>
- Mohanty, S.S., Sahoo, C.R., and Padhy, R.N. (2022). Role of hormone receptors and HER2 as prospective molecular markers for breast cancer: An update. *Genes & Diseases*, 9(3):648–658. <https://doi.org/10.1016/j.gendis.2020.12.005>
- Sahasrabudde, N.A., Korlimarla, A., Kulkarni, M., Kusuma, V., Prabhu, J.S., Dixit, S., Badwe, R., and Koppiker, C. (2021). NGS-based profiling of key cancer genes in Indian triple-negative breast cancer patients reinforces molecular heterogeneity of the disease. *Indian Journal of Cancer*, 58(4), 598–602. https://doi.org/10.4103/ijc.IJC_432_20
- The Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70. <https://doi.org/10.1038/nature11412>
- Toss, A., Piacentini, F., Cortesi, L., Artuso, L., Bernardis, I., Parenti, S., Molino, A., Conte, P., and Tagliafico, E. (2018). Genomic alterations at the basis of treatment resistance in metastatic breast cancer: Clinical applications. *Oncotarget*, 9(60):Article31606. <https://doi.org/10.18632/oncotarget.25810>
- Veeraraghavan, J., De Angelis, C., Gutierrez, C., Liao, F.-T., Sabotta, C., Rimawi, M.F., Osborne, C.K., and Schiff, R. (2025). HER2-positive breast cancer treatment and resistance. In: Yu, D.S., Balk, S.P., and DiGiovanna, J. (eds.). *Advances in Experimental Medicine and Biology*, Springer, 1464:495–525. https://doi.org/10.1007/978-3-031-70875-6_24