International Journal for Research in Biology and Pharmacy

GREEN PUBLICATION

ISSN (Online): 2208-2093
Volume 11 Issue 01 March 2025
Available on-line at: https://ijrbp.com/index.php/bp/issue/view/539

DOI:*https://doi.org/10.69980/bp.v11i1.2503* *Received: 26-Dec-2024 Accepted: 27-Feb-2025*

CLINICOPATHOLOGICAL, MOLECULAR SUBTYPE, AND SOMATIC MUTATION PROFILING OF BREAST CANCER SURVIVAL OUTCOMES

Dr. Aisha Al Mansoori[1], Dr. Omar Khalid Al Nuaimi[2], Dr. Sara Ahmed Al Kaabi[3]

[1] Department of Oncology and Molecular Medicine, United Arab Emirates University, Al Ain, United Arab Emirates
[2] Department of Medical Genetics and Genomics, Khalifa University, Abu Dhabi, United Arab Emirates
[3] College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

Corresponding Author: Dr. Aisha Al Mansoori
Email: aisha.almansoori.oncology@researchmail.org

Abstract
Breast cancer survival is influenced by complex interactions among clinicopathological characteristics, molecular subtype, and somatic mutation patterns. This study aimed to evaluate the relationship of clinical features, molecular subtype distribution, and selected somatic mutations with survival outcomes in breast cancer patients. A quantitative retrospective observational design was adopted using de-identified breast cancer patient records. The analysis included 1,904 cases with clinicopathological variables, receptor status, molecular subtype classification, somatic mutation profiles, and survival data. Descriptive statistics, frequency distributions, cross-tabulations, independent-samples t-tests, and chi-square tests were used for analysis. The cohort included 801 living patients and 1,103 deceased patients. Deceased patients had significantly higher age at diagnosis, tumor size, lymph-node positivity, mutation count, and Nottingham Prognostic Index than living patients. Tumor stage and histologic grade showed significant survival-related differences. Molecular subtype was also significantly associated with survival status, with LumA being the most frequent subtype and claudin-low showing the highest living proportion. Somatic mutation profiling identified PIK3CA and TP53 as the most frequent mutations. TP53, MUC16, KMT2C, and GATA3 mutation status showed significant associations with survival outcome. Subtype-specific mutation patterns indicated high TP53 mutation frequency in basal and HER2 tumors and high PIK3CA mutation frequency in LumA tumors. These findings support the value of integrated clinicopathological, molecular subtype, and somatic mutation profiling for understanding breast cancer survival outcomes.

Keywords: *Breast cancer; clinicopathological profiling; molecular subtype; somatic mutation; survival outcome; PAM50*

1. Introduction

Breast cancer srepresents one of the most significant cancers in terms of impact on female health and poses a considerable challenge in regard to timely detection and appropriate management and diagnosis. The clinical picture of breast cancer is characterized by considerable diversity even among tumors with a similar morphological appearance. Such diversity is associated with the biological complexity of breast cancer, where survival rates depend on various aspects related to the demographic characteristics of the patients, tumor pathology, receptors, molecular classification, and genetic profile. Therefore, a combination of all of these elements in an integrative manner is crucial for the analysis of breast cancer populations (Xu & Xu, 2023).

Classic clinicopathological factors like age of diagnosis, tumor size, tumor stage, histological grade, presence of lymph node metastasis, and receptors' status continue to be crucial for the assessment of breast cancers. They are crucial for understanding the severity of disease, aggressiveness of tumors, and expected outcomes. Large-scale population-based studies have found significant correlation between the prognosis of patients and histological and clinicopathological characteristics, thus emphasizing their importance (Han et al., 2020). Nonetheless, clinicopathological characteristics might be insufficient in accounting for differences in survival outcomes since there is no similarity between molecular levels of tumor cells with comparable clinicopathological profiles.

Thus, molecular classification has emerged as a critical element in the study of breast cancer. The PAM50 intrinsic subtype classification system defines several molecular subtypes of breast cancers, namely luminal A, luminal B, HER2-enriched, basal-like, among others. Different subtypes are characterized by differences in gene expression profiles, drug response, tumor recurrence, and survival rate. The PAM50 classification system has been assessed using varied techniques of expression, highlighting its potential utility as a reliable molecular classification system for breast cancer studies (Picornell et al., 2019). Further improvements in PAM50-based classifications have been shown to enhance the agreement between molecular intrinsic subtyping and clinical subtype definitions, suggesting that molecular classification might offer additional insights into prognosis beyond clinical parameters (Raj-Kumar et al., 2019).

Subtyping using gene expression has further proved that there can be different biological processes involved in the development of subtypes of breast cancer, which include immunology-based molecular signatures and variations in the tumor microenvironment (Mei et al., 2020). The recent real-world comparison between immunohistochemistry-based categorization and that from PAM50 assay has further revealed the point that molecular subtypes and clinicopathological categories are not the same all the time. This shows the need for molecular assessment. Further, subtype purity in luminal A breast cancers is known to affect patient clinical profile and survival, indicating the biological heterogeneity in seemingly favorable subtypes (Kumar et al., 2023). It should also be noted that PAM50 signature may change to more aggressive subtypes with disease progression (Jørgensen et al., 2021).

Apart from the molecular subtype, the somatic mutation profile is a significant source of knowledge about the biology of breast cancer. Mutations in genes such as PIK3CA and TP53 are some of the most investigated mutation events in breast cancer due to their effects on signaling pathways, genome integrity, cancer phenotype, and treatment significance. For instance, PIK3CA mutations have been documented in invasive breast cancer, but the rate differs depending on the population and background of the patients (Jia et al., 2021). On the other hand, TP53 mutations have been implicated in aggressive tumor phenotypes with poor prognosis in operable breast cancers (Park et al., 2022). Another field where genetic markers are particularly important is triple negative breast cancer subtype, since the molecular classification could potentially aid in better prognosis and therapy (Lu et al., 2023). The use of computational analysis based on RNA sequencing has proven that the molecular classification of breast cancer types can indeed be done (Yu et al., 2020).

In light of the above background information, the current study seeks to investigate the correlation between the clinical profile, molecular subtype prevalence, specific mutational patterns, and survival in patients diagnosed with breast cancer. Specifically, the current study adopts the approach of profiling to determine how both clinical pathologic features and molecular factors play a role in survival differences. Such an approach would be important for improving biological inference, prognosis prediction, and breast cancer research.

2. Methodology

2.1 Study Design

The research used a quantitative study design to investigate clinicopathological features, molecular subtyping, mutation profile, and association with breast cancer prognosis outcomes. Data were analyzed from a public dataset comprising of anonymized breast cancer data with clinicopathological, molecular, and mutation information (Alharbi, 2020).

2.2 Data Source and Study Population

The data consisted of 1,904 cases of patients diagnosed with breast cancer, along with 693 predictors covering clinicopathological parameters, receptor status, molecular subtyping, gene expression, somatic mutations, and survival outcome. Each case referred to a single patient. The data were de-identified and had been collected before, hence no new sample collection was done.

2.3 Study Variables

The outcome variable of interest was overall survival status, classified as alive or dead. The clinicopathological factors used were the patient's age at time of diagnosis, tumor size, tumor stage, tumor histology, lymph node involvement, Nottingham prognostic index, ER status, PR status, and HER2 status. The molecular profile was mostly determined using

the PAM50 plus claudin-low subtypes. The somatic mutations considered were PIK3CA, TP53, GATA3, MAP3K1, MUC16, KMT2C, SYNE1, and AHNAK2.

2.4 Data Preprocessing

Cleaning of the data set involved checking the presence of duplicate observations, missing values, inconsistent coding, and the relevance of variables used in the study. Duplicate patient observations were absent. Available-case analysis was used to handle missing values by excluding missing values from analyses based on specific variables. Binary recoding of somatic mutations was used, where the value "0" was used to denote that there was no mutation, while non-zero values of somatic mutations denoted the presence of mutation. Variables such as survival time and cancer death class that could cause outcome leakage were not included in the survival predictors.

2.5 Data Analysis

Descriptive statistics were performed on continuous variables as mean, median, standard deviation, minimum, and maximum. Frequencies and percentages were obtained for categorical variables. Comparison of clinicopathological features was done among living and deceased cases. Analysis of molecular subtypes distribution was performed, along with comparison of survival rates among subtype groups. Frequency of mutations was estimated for chosen genes, as well as comparisons by survival status and molecular subtype.

2.6 Statistical Analysis

Independent-samples t-test was applied in the comparison of continuous variables in terms of their differences based on survival, including age of diagnosis, size of tumor, status of positive lymph nodes, mutation burden, and Nottingham Prognostic Index. The chi-square test was conducted for the analysis of categorical data in relation to survival, involving the tumor stage, grade, receptors, subtype, and somatic mutation status. A p-value of < 0.05 was considered statistically significant.

3. Results

3.1 Descriptive Profile of the Study Cohort

The dataset of breast cancer patients involved for analysis comprised 1,904 patients' medical records, including clinicopathological characteristics, receptor status, molecular subtype, somatic mutation, and survival outcomes. The distribution of survival outcome indicated that there were 801 surviving patients while 1,103 deceased patients.

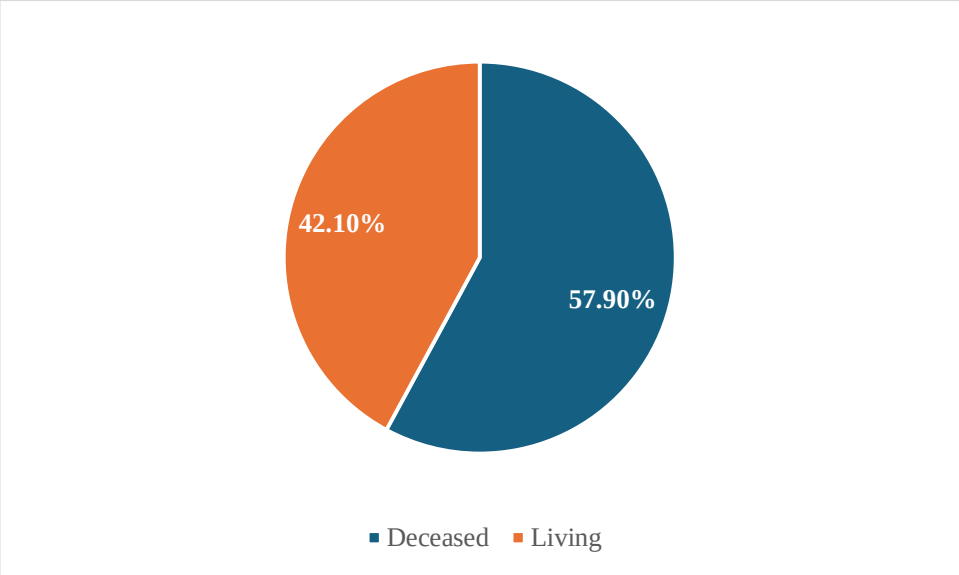

Figure 1. Survival outcome distribution

In Figure 1, it is shown that the percentage of death patients is 57.9%, whereas the percentage of alive patients is 42.1%.

Table 1. Descriptive Profile of Continuous Clinicopathological Variables

Variable	N	Mean	SD	Median	Minimum	Maximum
Age at diagnosis	1,904	61.09	12.98	61.77	21.93	96.29
Tumor size	1,884	26.24	15.16	23.00	1.00	182.00
Tumor stage	1,403	1.75	0.63	2.00	0.00	4.00
Histologic grade	1,832	2.42	0.65	3.00	1.00	3.00
Positive lymph nodes	1,904	2.00	4.08	0.00	0.00	45.00
Mutation count	1,859	5.70	4.06	5.00	1.00	80.00

Nottingham Prognostic Index	1,904	4.03	1.14	4.04	1.00	6.36
Overall survival months	1,904	125.12	76.33	115.62	0.00	355.20

Based on Table 1 above, the average age at which cancer patients were diagnosed was 61.09 years, while the average length of overall survival period was 115.62 months. The average tumor size was 23.00, whereas the average positive lymph nodes were zero, suggesting that there was a considerable number of cases with no lymph nodes positivity.

3.2 Clinicopathological Characteristics by Survival Status

The clinicopathological correlation indicated distinct variations between living and deceased individuals. The deceased exhibited a higher average age, larger tumor size, more positive lymph nodes, more mutations, and a higher Nottingham Prognostic Index compared to the living subjects.

Table 2. Comparison of Continuous Variables by Survival Status

Variable	Deceased Mean	Living Mean	Deceased Median	Living Median	p-value
Age at diagnosis	64.44	56.46	66.36	56.74	<0.001
Tumor size	28.36	23.32	25.00	20.00	<0.001
Positive lymph nodes	2.57	1.21	1.00	0.00	<0.001
Mutation count	5.96	5.32	5.00	5.00	<0.001
Nottingham Prognostic Index	4.17	3.85	4.05	4.03	<0.001

From Table 2, it is clear that the median age at diagnosis for deceased subjects was significantly higher compared to that of surviving subjects (66.36 years vs. 56.74 years). In addition, the median size of the tumor was higher for deceased subjects compared to surviving subjects (25.00 vs. 20.00). Lymph node positivity was higher in the case of deceased subjects, and there were statistically significant differences in the continuous variables considered.

3.3 Tumor Stage, Grade, and Survival Outcome

The difference between tumor staging and grading was evident in relation to survival outcomes. The number of individuals who were alive decreased with an increase in tumor staging. The same trend was noticed for tumor grading, with grade 1 having the highest number of survivors and grade 3 having the least.

Table 3. Survival Outcome by Tumor Stage and Histologic Grade

Variable	Category	Deceased n	Living n	Total n	Living %	p-value
Tumor stage	Stage 0	1	3	4	75.0	<0.001
	Stage 1	215	260	475	54.7	
	Stage 2	482	318	800	39.8	
	Stage 3	86	29	115	25.2	
	Stage 4	8	1	9	11.1	
Histologic grade	Grade 1	75	90	165	54.5	<0.001
	Grade 2	414	326	740	44.1	
	Grade 3	564	363	927	39.2	

Table 3 indicates that the living percentage decreased from 54.7% at stage 1 to 11.1% at stage 4. Similarly, the histological grade was associated with survival in such a way that the living percentages were 54.5%, 44.1%, and 39.2% at grades 1, 2, and 3, respectively.

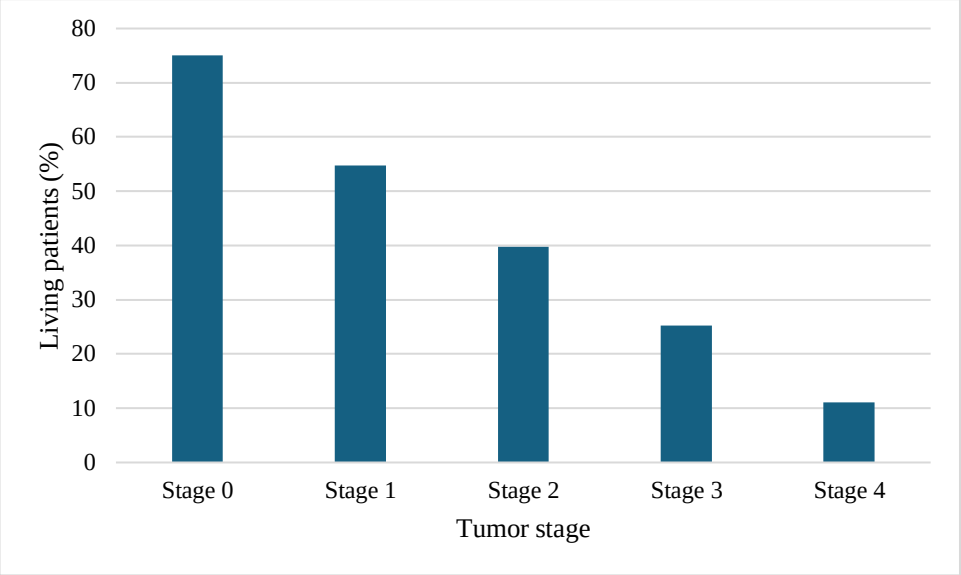

Figure 2. Survival proportion across tumor stages

Figure 2 shows the decreasing survival proportion with the increasing tumor stage, suggesting a poorer survival distribution for patients with advanced tumor stages.

3.4 Receptor and Treatment Profile

The study sample comprised mostly ER-positive and HER2-negative receptors. However, receptor status variables were not associated significantly with the survival status, while radiotherapy status variables were significantly associated with survival distribution.

Table 4. Receptor and Treatment Characteristics by Survival Status

Variable	Category	Deceased n	Living n	Total n	Living %	p-value
ER status	Negative	250	195	445	43.8	0.424
	Positive	853	606	1,459	41.5	
PR status	Negative	529	366	895	40.9	0.351
	Positive	574	435	1,009	43.1	
HER2 status	Negative	956	712	1,668	42.7	0.168
	Positive	147	89	236	37.7	
Chemotherapy	No	891	617	1,508	40.9	0.053
	Yes	212	184	396	46.5	
Hormone therapy	No	409	321	730	44.0	0.201
	Yes	694	480	1,174	40.9	
Radiotherapy	No	496	271	767	35.3	<0.001
	Yes	607	530	1,137	46.6	

As depicted in Table 4, the majority of the subjects were positive for estrogen receptors. Negative cases for HER2 were also common compared to the HER2-positive subjects. Even though there was no statistical significance between the status of receptors and survival status, radiotherapy exhibited a statistical correlation with survival status, where more patients receiving radiotherapy were alive.

3.5 Molecular Subtype Distribution and Survival Outcome

Subtype distribution based on molecular subtypes revealed that the most common subtype was LumA, followed by LumB, HER2, basal, Claudin-low, normal-like, and NC. Survival outcome varied significantly between molecular subtype groups.

Table 5. Molecular Subtype Distribution and Survival Outcome

Molecular subtype	Deceased n	Living n	Total n	Living %
LumA	364	315	679	46.4
LumB	303	158	461	34.3
HER2	155	65	220	29.5
Basal	111	88	199	44.2
Claudin-low	89	110	199	55.3
Normal-like	76	64	140	45.7

NC	5	1	6	16.7

According to Table 5, LumA was the most common subtype with 679 patients, followed by LumB with 461 patients. The claudin-low subtype demonstrated the largest number of survivors at 55.3%, followed by LumA with 46.4% and normal-like subtype with 45.7%. The percentages of survivors were relatively low for the HER2 subtype (29.5%) and LumB subtype (34.3%). The relationship between molecular subtype and survival status was significant ($p < 0.001$).

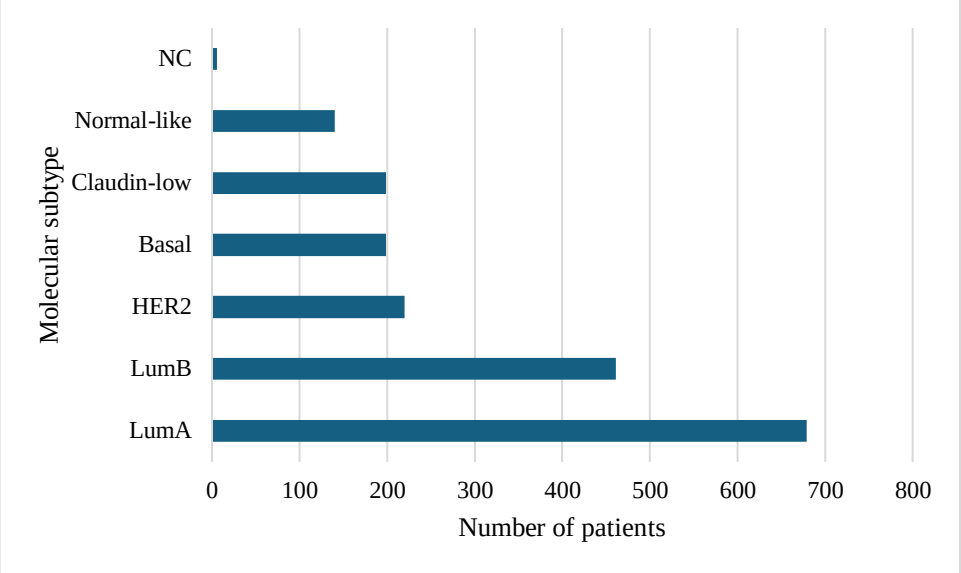

Figure 3. Molecular subtype distribution

Figure 3 gives the distribution of molecular subtypes with the prevalence of LumA being the highest one followed by LumB.

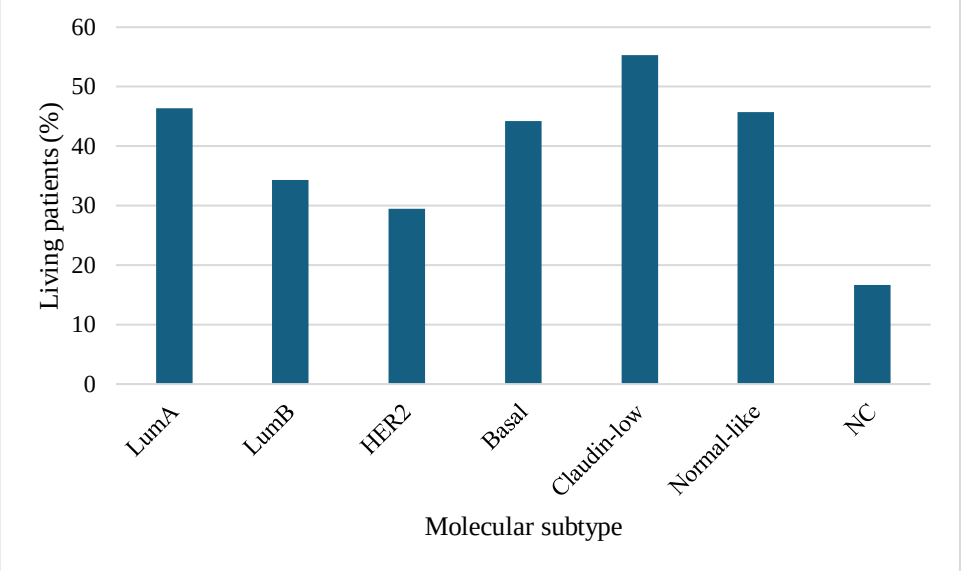

Figure 4. Survival proportion across molecular subtypes

Figure 4 provides subtype-specific survival rates where the highest survival rate was for claudin-low, while the lowest was for NC and HER2 subtypes.

3.6 Somatic Mutation Frequency

Somatic mutation profiling revealed that the PIK3CA and TP53 genes were mutated at the highest frequency. Commonly mutated genes apart from PIK3CA and TP53 were MUC16, AHNAK2, KMT2C, SYNE1, GATA3, MAP3K1, AHNAK, and DNAH11.

Table 6. Top Somatic Mutation Frequencies

Gene mutation	Mutated n	Mutation %
PIK3CA	795	41.8
TP53	659	34.6

MUC16	326	17.1
AHNAK2	311	16.3
KMT2C	234	12.3
SYNE1	232	12.2
GATA3	230	12.1
MAP3K1	198	10.4
AHNAK	176	9.2
DNAH11	175	9.2

As shown in Table 6, PIK3CA mutations occurred in 41.8% of patients, while TP53 mutations occurred in 34.6% of patients. Other common mutations included MUC16, which was seen in 17.1%, and AHNAK2, which was found in 16.3% of patients, respectively.

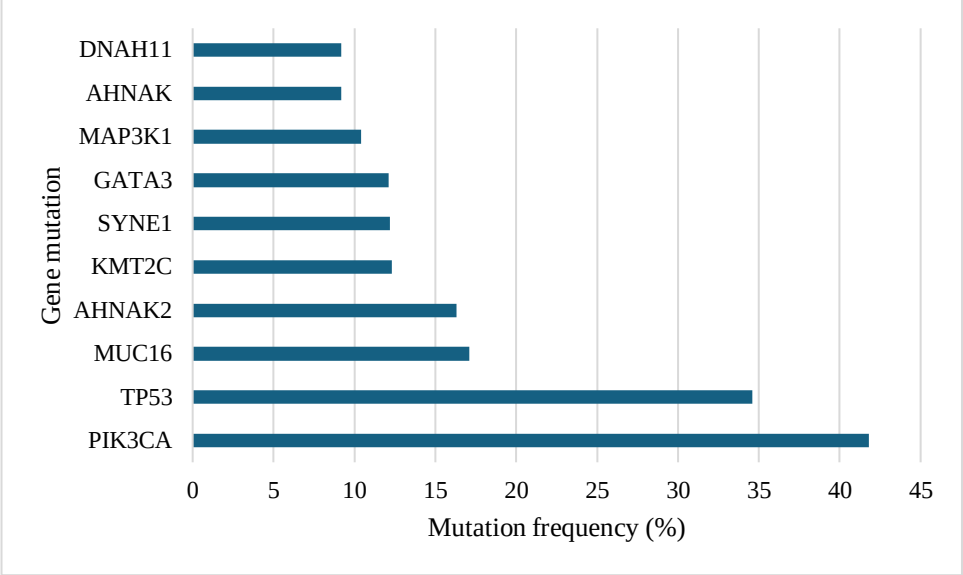

Figure 5. Frequency of selected somatic mutations

The frequency of the common somatic mutations is presented in Figure 5, showing that PIK3CA and TP53 are the most frequently mutated genes.

3.7 Somatic Mutation Status and Survival Outcome

Somatic mutations had distinct survival curves. Mutations of TP53, MUC16, KMT2C, and GATA3 genes had significant associations with survival status.

Table 7. Selected Somatic Mutations by Survival Status

Gene mutation	Mutation status	Deceased n	Living n	Total n	Living %	p-value
PIK3CA	Absent	622	487	1,109	43.9	0.060
	Present	481	314	795	39.5	
TP53	Absent	699	546	1,245	43.9	0.034
	Present	404	255	659	38.7	
MUC16	Absent	893	685	1,578	43.4	0.011
	Present	210	116	326	35.6	
KMT2C	Absent	953	717	1,670	42.9	0.049
	Present	150	84	234	35.9	
GATA3	Absent	1,002	672	1,674	40.1	<0.001
	Present	101	129	230	56.1	
MAP3K1	Absent	981	725	1,706	42.5	0.301
	Present	122	76	198	38.4	

As seen in Table 7, the proportion of TP53 mutation patients was lower than that of the TP53 non-mutation group (38.7% versus 43.9%). Lower proportions of patients with mutations in MUC16 and KMT2C were also seen. However, patients with GATA3 mutation had a higher proportion of living patients compared to patients without GATA3 mutation (56.1% versus 40.1%). PIK3CA mutation was found to have an inconclusive relationship with the outcome of survival.

3.8 Mutation Patterns Across Molecular Subtypes

Mutation rate varied significantly between different molecular subtypes. The highest TP53 mutation rate was observed among basal cancers, whereas LumA cancers had high mutation rates in genes PIK3CA, GATA3, and MAP3K1.

Table 8. Selected Mutation Frequencies Across Molecular Subtypes

Molecular subtype	N	PIK3CA %	TP53 %	GATA3 %	MAP3K1 %	MUC16 %	KMT2C %
Basal	199	16.1	88.4	0.0	4.5	23.1	7.5
HER2	220	41.4	70.0	7.7	6.8	27.3	14.5
LumA	679	57.4	11.9	19.6	16.2	16.6	15.6
LumB	461	34.9	24.3	13.9	8.7	16.1	11.3
NC	6	50.0	0.0	0.0	33.3	0.0	33.3
Normal-like	140	50.0	22.9	8.6	10.0	10.0	7.9
Claudin-low	199	24.1	52.3	2.0	4.0	9.5	8.0

According to Table 8, TP53 mutation rate was highest among basal cancers (88.4%) and HER2 cancers (70.0%). PIK3CA mutation rate was highest among LumA cancers (57.4%) and normal-like cancers (50.0%). Mutation rates in GATA3 and MAP3K1 were higher among LumA cancers compared to basal or claudin-low cancers.

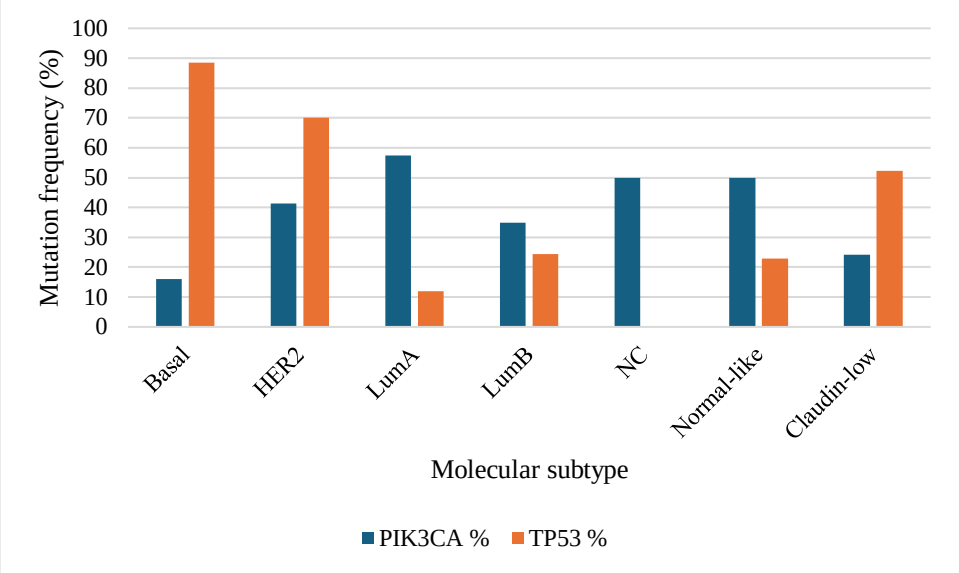

Figure 6. TP53 and PIK3CA mutation frequency across molecular subtypes

Figure 6 shows the different mutation pattern among subtypes with high rates of TP53 mutations in basal and HER2 subtypes, while PIK3CA mutations occur at high rates in LumA and normal-like subtypes.

4. Discussion

This current study assessed the clinical, pathological features, distribution of molecular subtypes, mutational profile, and their impact on breast cancer survival outcomes. According to the results, survival status is not defined by a singular variable but represents an aggregate of the patient's clinical presentation, tumor growth, and molecular subtype and mutational profile selected. Indeed, from the larger biological perspective, the heterogeneity of breast cancer indicates the complex nature of its prognosis, including both traditional pathological and molecular features.

Based on the descriptive results, it was observed that dead patients had advanced age, larger tumors, more positive lymph nodes, high mutations, and high Nottingham Prognostic Index scores compared with surviving patients. These variations could mean that patient-related variables and aggressive tumors play significant roles in determining survival outcomes. Advanced age at the time of diagnosis might have a bearing on survival by virtue of patient biological susceptibility, co-morbid conditions, adverse tolerance to therapies, and late diagnosis. Likewise, having large tumors and being positive for lymph node metastasis would indicate more disease load and increased chances of spreading. Notably, the differences seen between the Nottingham Prognostic Index scores provide additional support for the application of a holistic approach to prognosis.

Survival relationship was evident when tumor stage and histologic grade were analyzed. The percentage of patients that were alive declined consistently, starting from an early stage tumor to an advanced stage tumor, suggesting that tumor staging continues to play a critical role. It can be seen that the percentage of patients that were alive in stage 1 tumors was significantly higher than stage 3 and 4 tumors. With respect to histologic grade, it was found that the distribution of patients that were alive was better in grade 1 tumors compared to grade 3 tumors. This finding is biologically consistent because high-grade tumors generally display greater cellular atypia, higher proliferative activity, and more aggressive clinical behavior.

Results from the receptor status analysis demonstrated that the proportion of ER-positive and HER2-negative tumors was higher in the current patient population. Nevertheless, none of the parameters of ER, PR, and HER2 status reached statistical significance in terms of prognosis in the survival outcomes of this analysis. This could be explained by the dichotomous nature of the survival endpoints, differences in therapies, variability in follow-up, and the impact of other molecular parameters aside from receptor status on survival. While receptor status remains crucial in the clinical stratification and treatment of patients, it seems that receptor status must be evaluated in conjunction with other parameters. Radiotherapy status was strongly linked to the distribution of survival, where there were more alive patients in the radiotherapy group. It must be noted carefully that allocation of patients for therapy treatment is dependent on the stages of illness and tumor properties and is not based on randomness.

Another significant observation in the current research is related to the subtype analysis. The most common subtype among breast cancer patients was LumA, followed by LumB, HER2, basal, claudin-low, normal-like, and NC. Living proportions were not similar for each of the molecular subtypes since some subtypes had relatively high living proportions compared to others. For example, LumA and claudin-low cases were characterized by higher living proportions. These results support the theory that breast cancer molecular subtypes may vary in their biological behavior, response to treatment, and prognosis. There is evidence at the population level demonstrating that survival from breast cancer differs based on its molecular subtype, with hormone receptor-positive breast cancers generally having better prognoses compared to biologically aggressive breast cancers (Fallahpour et al., 2017). Variations in survival rates among different molecular subtypes of breast cancer have also been documented in population studies conducted in the United States (Howlader et al., 2018).

The current analysis highlights the significance of using PAM50 molecular classification as well. PAM50 subtypes are different biological classes of tumors, rather than clinically defined categories. Differences between LumA, LumB, HER2, basal, and claudin-low tumors are explained by variations in gene expression profiles, cell proliferation, hormonal regulation, immunological characteristics, and the tumor microenvironment. Indeed, studies indicate that molecular classification of tumors by their intrinsic characteristics could help to uncover heterogeneous tumor characteristics that have a significant impact on survival (Jaber et al., 2020). Moreover, the multi-omic study of PAM50 subtyping shows that molecular subtypes are shaped by complex interactions of various genomic and transcriptomic factors (Ochoa et al., 2020). Thus, the molecular subtype data presented in this article contribute to the biological significance of survival disparities rather than clinicopathological factors only.

The results of somatic mutation analysis indicated that PIK3CA and TP53 are the most frequently mutated genes followed by MUC16, AHNAK2, KMT2C, SYNE1, GATA3, MAP3K1, AHNAK, and DNAH11. It is logical to assume that the frequent changes in the PIK3CA and TP53 genes are associated with the breast cancer genomic profile, which typically implies mutations in cell growth regulation and genome stability genes. Breast cancer genome-wide mutational analysis has shown that somatic mutations clarify the genomics and transcriptome heterogeneity and provide clinical subgroup distinctions (Pereira et al., 2016). In this current study, there were significant correlations between the presence of TP53, MUC16, KMT2C, and GATA3 mutations and survival rate. The living fraction in TP53-mutation was lower compared to those without TP53-mutation. This can be justified as TP53 mutation is often correlated with poor DNA damage repair, genomic instability, and malignancy. Similarly, patients with mutations in MUC16 and KMT2C also exhibited lower living fractions, indicating that certain mutations could be an indicator of other molecular instabilities or malignancies.

Intriguingly, cases with mutated GATA3 had a greater survival rate compared to those without the mutation. This observation could be due to the relationship between GATA3 and luminal differentiation with hormone receptor-positive tumor characteristics. There is literature from systematic review and meta-analysis that demonstrates how GATA3 is related to positive clinical outcomes in breast cancer cases, which include hormone receptor positivity and better differentiation (Guo et al., 2017). Thus, the increased survival rate in cases with the GATA3 mutation is probably linked to subtype prevalence and luminal tumor characteristics.

Subtype-specific mutations further showed unique molecular features. For example, TP53 mutations occurred most frequently in basal and HER2 subtypes, but PIK3CA mutations occurred most frequently in LumA and normal-like subtypes. This represents a biological difference between aggressive and luminal-like breast cancers. The frequent occurrence of TP53 mutations in the basal and HER2 subtypes probably leads to poor survival patterns in those types of breast cancer. In addition, the frequency of PIK3CA mutations in LumA probably relates to its frequent alteration in hormone receptor breast cancer. Further studies have also demonstrated that mutations in the genes PIK3CA and TP53 might also be of importance in treatment-related issues for endocrine receptor-positive metastatic breast cancer (Chen et al., 2023).

In conclusion, the data presented above confirms the importance of taking into account both the integration of clinicopathological factors, as well as the molecular subtype and somatic mutations in terms of breast cancer prognosis. Although traditionally considered factors like age, tumor size, stage, grade, and lymph-node status are vital, an extra biological dimension can be obtained using the information concerning molecular subtype and mutational pattern. The study therefore emphasizes that breast cancer prognosis is best understood through a combined profiling approach rather than by relying on a single clinical or molecular indicator.

5. Conclusion

This study illustrated that outcomes of breast cancer patients' survivals are correlated with both clinicopathological features, molecular subtype composition, and somatic mutations. The deceased individuals were older at the time of

diagnosis, had bigger tumors, lymph-node positivity, more mutations, and high Nottingham Prognostic Index than the living group. This implies that the tumor's aggressiveness and burden continue to play roles in the difference in survival rates. Additionally, tumor stages and histologic grades exhibited significant survival trends with less living rate in the advanced stage and high grade groups. Subtype profiling at the molecular level yielded additional information regarding prognosis. The LumA group was found to be the most common subtype among others, and there were significant differences in survival rate distribution between different subtypes. There were higher survival rates observed in the claudin-low and LumA subtypes than the HER2 and LumB subtypes. This illustrates the significance of biological aspects in predicting patient prognosis based on their breast cancer subtype. Somatic mutation analysis revealed that PIK3CA and TP53 were the most common mutations. There was a statistically significant relationship between mutation status and survival. The presence of TP53, MUC16, and KMT2C mutations was significantly related to reduced survival rates, whereas GATA3 mutation had a positive relationship with the living proportion. There is evidence of subtype-specific mutation profiles, which indicated that there were high frequencies of TP53 mutations in basal and HER2 subtypes as well as PIK3CA mutations in LumA and normal-like subtypes. In summary, results obtained from clinicopathological and molecular evaluation contribute to the better understanding of survival in breast cancer patients.

References

1. Alharbi, R. (2020). *Breast Cancer Gene Expression Profiles (METABRIC)* [Data set]. Kaggle. https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric
2. Chen, T. W. W., Hsiao, W., Dai, M. S., Lin, C. H., Chang, D. Y., Chen, I. C., ... & Lu, Y. S. (2023). Plasma cell-free tumor DNA, PIK3CA and TP53 mutations predicted inferior endocrine-based treatment outcome in endocrine receptor-positive metastatic breast cancer. *Breast Cancer Research and Treatment*, *201*(3), 377-385.
3. Fallahpour, S., Navaneelan, T., De, P., & Borgo, A. (2017). Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *Canadian Medical Association Open Access Journal*, *5*(3), E734-E739.
4. Guo, Y., Yu, P., Liu, Z., Maimaiti, Y., Chen, C., Zhang, Y., ... & Huang, T. (2017). Prognostic and clinicopathological value of GATA binding protein 3 in breast cancer: a systematic review and meta-analysis. *PloS one*, *12*(4), e0174843.
5. Han, Y., Wang, J., & Xu, B. (2020). Clinicopathological characteristics and prognosis of breast cancer with special histological types: a surveillance, epidemiology, and end results database analysis. *The Breast*, *54*, 114-120.
6. Howlader, N., Cronin, K. A., Kurian, A. W., & Andridge, R. (2018). Differences in breast cancer survival by molecular subtypes in the United States. *Cancer Epidemiology, Biomarkers & Prevention*, *27*(6), 619-626.
7. Jaber, M. I., Song, B., Taylor, C., Vaske, C. J., Benz, S. C., Rabizadeh, S., ... & Szeto, C. W. (2020). A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Breast Cancer Research*, *22*(1), 12.
8. Jia, M., Liao, N., Chen, B., Zhang, G., Wang, Y., Li, X., ... & Balch, C. M. (2021). PIK3CA somatic alterations in invasive breast cancers: different spectrum from Caucasians to Chinese detected by next generation sequencing. *Breast Cancer*, *28*(3), 644-652.
9. Jørgensen, C. L. T., Larsson, A. M., Forsare, C., Aaltonen, K., Jansson, S., Bradshaw, R., ... & Ryden, L. (2021). PAM50 intrinsic subtype profiles in primary and metastatic breast cancer show a significant shift toward more aggressive subtypes with prognostic implications. *Cancers*, *13*(7), 1592.
10. Kumar, N., Gann, P. H., McGregor, S. M., & Sethi, A. (2023). Quantification of subtype purity in Luminal A breast cancer predicts clinical characteristics and survival. *Breast Cancer Research and Treatment*, *200*(2), 225-235.
11. Lu, B., Natarajan, E., Balaji Raghavendran, H. R., & Markandan, U. D. (2023). Molecular classification, treatment, and genetic biomarkers in triple-negative breast cancer: a review. *Technology in cancer research & treatment*, *22*, 15330338221145246.
12. Mei, J., Zhao, J., & Fu, Y. (2020). Molecular classification of breast cancer using the mRNA expression profiles of immune-related genes. *Scientific Reports*, *10*(1), 4800.
13. Ochoa, S., de Anda-Jauregui, G., & Hernández-Lemus, E. (2020). Multi-omic regulation of the pam50 gene signature in breast cancer molecular subtypes. *Frontiers in oncology*, *10*, 845.
14. Park, J. H., Kwon, M. J., Seo, J., Kim, H. Y., Min, S. K., & Kim, L. S. (2022). Somatic mutations of TP53 identified by targeted next-generation sequencing are poor prognostic factors for primary operable breast cancer: a single-center study. *Journal of Breast Cancer*, *25*(5), 379.
15. Pereira, B., Chin, S. F., Rueda, O. M., Vollan, H. K. M., Provenzano, E., Bardwell, H. A., ... & Caldas, C. (2016). The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature communications*, *7*(1), 11479.
16. Picornell, A. C., Echavarria, I., Alvarez, E., López-Tarruella, S., Jerez, Y., Hoadley, K., ... & Martin, M. (2019). Breast cancer PAM50 signature: correlation and concordance between RNA-Seq and digital multiplexed gene expression technologies in a triple negative breast cancer series. *BMC genomics*, *20*(1), 452.
17. Raj-Kumar, P. K., Liu, J., Hooke, J. A., Kovatich, A. J., Kvecher, L., Shriver, C. D., & Hu, H. (2019). PCA-PAM50 improves consistency between breast cancer intrinsic and clinical subtyping reclassifying a subset of luminal A tumors as luminal B. *Scientific reports*, *9*(1), 7956.
18. Xu, H., & Xu, B. (2023). Breast cancer: Epidemiology, risk factors and screening. *Chinese journal of cancer research*, *35*(6), 565.

19. Yu, Z., Wang, Z., Yu, X., & Zhang, Z. (2020). RNA-Seq-Based Breast Cancer Subtypes Classification Using Machine Learning Approaches. *Computational intelligence and neuroscience*, *2020*(1), 4737969.