



GENE EXPRESSION AND SOMATIC MUTATION PROFILES ACROSS MOLECULAR SUBTYPES OF BREAST CANCER

Sarah N. Namusoke^{1*}, Joseph K. Mugisha², Esther A. Nakato³

^{1*}Department of Biochemistry, Makerere University, Kampala, Uganda. sarahnamusoke@hotmail.com

²Department of Pathology, Makerere University, Kampala, Uganda

³Department of Biological Sciences, Mbarara University of Science and Technology, Mbarara, Uganda

***Corresponding Author:** Sarah N. Namusoke, sarahnamusoke@hotmail.com

Received: 04/02/2026, Revised: 11/04/2026, Accepted: 18/05/2026, Published: 28/05/2026

ABSTRACT

Breast cancer is a molecularly heterogeneous disease characterized by distinct gene-expression patterns and somatic alterations across intrinsic subtypes. Integrated assessment of transcriptomic and genomic features can improve understanding of subtype-specific tumor biology. A retrospective secondary-data analysis was conducted using 1,898 breast cancer cases classified as Luminal A, Luminal B, HER2-enriched, Basal, Claudin-low, or Normal-like. Gene-expression differences were evaluated using the Kruskal–Wallis test, while mutation–subtype associations were assessed using chi-square or Fisher’s exact tests. Mann–Whitney U testing was used to compare expression between mutation-positive and mutation-negative groups, with false discovery rate correction applied for multiple testing. Significant subtype-associated expression differences were identified, with GATA3, EGFR, MAPT, AURKA, CCNE1, CHEK1, IGF1R, CDC25A, CDK1, and E2F2 among the most strongly differentiated genes. PIK3CA and TP53 were the most frequently mutated genes, while TP53, PIK3CA, GATA3, MAP3K1, CDH1, and other alterations showed significant subtype associations. Mutation status was also linked with corresponding expression changes for several genes, including CDH1, GATA3, AKT1, PTEN, TP53, MAP3K1, and ERBB2. Integrated transcriptomic and genomic profiling revealed marked molecular heterogeneity across breast cancer subtypes and highlighted potentially important subtype-associated biomarkers.

Keyword: Breast cancer; Gene expression; Somatic mutation; Molecular subtype; METABRIC

1. Introduction

Breast cancer is a very heterogeneous disease and it differs in morphology, molecular changes, clinical behavior and response to treatment to a great extent. Histopathological classification is still vital in diagnosis, but it is not quite reflective of the molecular diversity found amongst tumors of similar pathological appearance. Genomic and transcriptomic profiling has thus enhanced breast cancer classification to reveal biologically distinct groups. Combining both morphological and molecular data has enhanced the characterization of tumors and aided in more accurate stratification of the diseases (Binder et al., 2021). The disease of breast cancer has now been recognized as a set of molecularly different diseases that might need subtype-specific treatment (Testa et al., 2020). The molecular intrinsic classification has also demonstrated significant benefits compared to clinical classification in the definition of biologically meaningful tumor groups (Szymiczek et al., 2021). The classification by gene-expression has been in the forefront in the characterization of the key molecular subtypes of breast cancer. Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like tumors have different transcriptional profiles, whereas Claudin-low tumors are another molecular group with a distinct expression profile. Consensus systems via gene-expression have been created to enhance the reproducibility of subtype assignment of breast cancer samples (Horr & Buechler, 2021). Multi-omics techniques have also shown that the different molecular features can be combined to enhance the differentiation of the breast cancer subtypes (Lupat et al., 2023). Specific molecular changes in the breast cancer brain metastases have also been observed, and they emphasize the further applicability of molecular classification as the disease advances (Cosgrove et al., 2022).

Gene-expression profiling offers valuable data on tumor proliferation-related pathways, hormone signaling, growth-factor signaling, differentiation, and therapeutic response. Expression signatures can unveil molecular patterns potentially not visible with histopathological evaluation alone and play a role in biologically informed interpretation of treatments (Creighton, 2023). In comparative studies, oncogenic expression signatures have been demonstrated to indicate significant molecular mechanisms that relate to the development of breast tumors (Kim et al., 2020). Paired primary and recurrent triple-negative breast cancers have also been found to differ in terms of gene-expression and immune-related features, which suggests that tumor transcriptional profiles can change as a disease progresses (Hutchinson et al., 2020). Another significant cause of heterogeneity of breast cancer is somatic mutations. The frequent mutations in genes including PIK3CA, TP53, GATA3, MAP3K1, CDH1, PTEN, BRCA1, BRCA2, and ERBB2 have been known to cause different genomic patterns in tumors. Molecular profiling on a scale has shown that there is a great deal of variation in the somatic changes of primary breast cancers and their transcriptomic profiles (Brueffer et al., 2020). The genetic variations between primary tumors and recurrences also suggest that breast cancer may experience significant molecular changes in the course of time (Fumagalli et al., 2020). A population-specific molecular variation has been cited in the Asian breast

cancers (Pan et al., 2020) and distinct tumor signatures have been associated with germline and somatic alterations of BRCA-related genes (Bodily et al., 2020).

Genomic changes associated with subtypes can have an impact on the biology of tumors and possible therapeutic weaknesses. Specific genomic profiles were identified among metastatic special histologic types of breast cancer (Pareja et al., 2020), whereas metaplastic breast carcinoma has demonstrated to exhibit typical molecular characteristics within the wider range of triple-negative (Gonzalez-Martinez et al., 2020). Precision-medicine strategies have also been used in refractory metastatic triple-negative breast cancer with the use of molecular subtyping and genomic profiling (Jiang et al., 2021). Recent studies have further implicated certain mutational processes in HER2 enriched breast cancer, indicating that further subtype-specific genomic pathways remain to be discovered (Jeong et al., 2026). The integrated analysis of gene expression and somatic mutations may offer a more integrated picture of the heterogeneity of breast cancer as compared to molecular layers individually. Combination of various types of molecular data, such as gene expression, somatic mutation, and other molecular characteristics, has been shown to be valuable in characterizing tumors (Liu et al., 2021). Nevertheless, the study of gene-expression and mutation patterns is usually considered in isolation, with this restricting the appreciation of the co-occurrence of transcriptomic and genomic changes in molecular subtypes. As such, the research was designed to compare the patterns of gene-expression, to characterize the distributions of somatic-mutations, and to evaluate the mutation-associated differences in expression in molecular subtypes of breast cancer.

Objectives of the study

1. To compare gene-expression profiles and identify significantly differentially expressed genes across the major molecular subtypes of breast cancer.
2. To characterize subtype-specific somatic mutation patterns and evaluate the association between mutation status and corresponding gene-expression levels in breast cancer.

2. Methodology

2.1 Study Design

The secondary-data analytical study was a retrospective type that was used to describe the profile of gene-expression and somatic-mutation in molecular subtypes of breast cancer. Molecular and clinical patient-level data previously gathered was analyzed, without actual participant recruitment, intervention, or follow-up. The comparative approach was an analytical framework on comparisons between known molecular subtypes of breast cancer. A combination of gene-expression patterns, somatic-mutation frequencies, and correlations between mutation status and the related gene expression was used to determine subtype-specific molecular variations and overall patterns of heterogeneity in tumors in the study population.

2.2 Data Source

Data were obtained from the publicly available Breast Cancer Gene Expression Profiles (METABRIC) dataset available on Kaggle (Alharbi, 2020). The data consisted of 1,904 cases of breast cancer and 693 variables, such as clinical features, molecular subtype data, gene-expression data, and somatic-mutation variables. In particular, there were 489 gene-expression variables and 173 gene-specific mutation variables that could be analyzed on the molecular level. The data had been in the CSV format and had to be downloaded to prepare and analyze the data using statistics in the Kaggle repository.

2.3 Study Population and Molecular Subtype Classification

A total of all breast cancer cases where PAM50 plus Claudin-low subtype data were available were first to be analyzed. The PAM50+ Claudin-low subtype variable that was part of the data set was used to molecularly classify it. The groups were Luminal A, Luminal B, HER2-enriched, Basal, Claudin-low, Normal-like and NC. Six cases that were NC were excluded as they were not a well-defined molecular subtype. This resulted in 1,898 cases that were retained to make comparative analyses across the six molecular subtype groups that were defined.

2.4 Gene-Expression and Somatic-Mutation Variables

The analysis considered a total 489 continuous gene-expression variables and 173 gene-specific somatic-mutation variables. Measurements of expressions were left in their processed numerical form, with valid negative and zero values. Genes in mutation fields were PIK3CA, TP53, GATA3, MAP3K1, CDH1, PTEN, AKT1, BRCA1, BRCA2, and ERBB2. To analyse, mutation state was transformed into binary form with 0 denoting the lack of mutation record and non-zero mutation annotation indicating mutation-positive cases.

2.5 Data Preparation and Quality Assessment

Duplicate records, duplicate patient identifiers, missing values, format of variables, and consistency of molecular subtype classification were checked and then statistical analysis was performed on the dataset. There were no duplicate identifiers or observations of patients. There were missing values in the selected clinical variables but the main gene-expression and mutation variables used in the molecular analyses were complete. Available-case analysis was used to deal with missing clinical observations. Zero values in mutation variables and negative or zero gene-expression values were not discarded since these values indicated something meaningful about the molecules, as opposed to missing values.

2.6 Statistical Analysis

Frequencies and percentages were used to describe categorical variables whereas mean $\pm$ SD or median with interquartile range were used to describe continuous variables. The Kruskal -Wallis test was used to

evaluate the difference in gene-expression between the molecular subtypes. Chi-square or Fisher exact tests were used to assess the associations between mutation status and molecular subtype. Mann–Whitney U tests were used to compare gene expression in mutation-positive and mutation-negative cases, and Spearman correlation was used to measure the selected continuous molecular variables. It was adjusted by BenjaminiHochberg false discovery rate correction, where adjusted $p < 0.05$ is defined as statistically significant.

3. Results

3.1 Distribution of Molecular Breast Cancer Subtypes

One thousand and eight hundred and ninety-eight cases of breast cancer were obtained after the elimination of six cases that were classified as NC. The most prevalent molecular subtype was luminal A, with 679 cases (35.77%), and the second one was Luminal B, which included 461 cases (24.29%). HER2-enriched tumors were found in 220 cases (11.59%), Basal and Claudin-low subtypes had 199 cases (10.48% each). Normal-like tumors were the least in number with 140 cases (7.38%). All subtypes of the molecule are given in Table 1.

Table 1. Distribution of breast cancer cases according to molecular subtype

Molecular subtype	Number of cases	Percentage (%)
Luminal A	679	35.77
Luminal B	461	24.29
HER2-enriched	220	11.59
Basal	199	10.48
Claudin-low	199	10.48
Normal-like	140	7.38
Total	1,898	100.00

Figure 1 presents the distribution of the cases of breast cancer by the molecular subtypes.

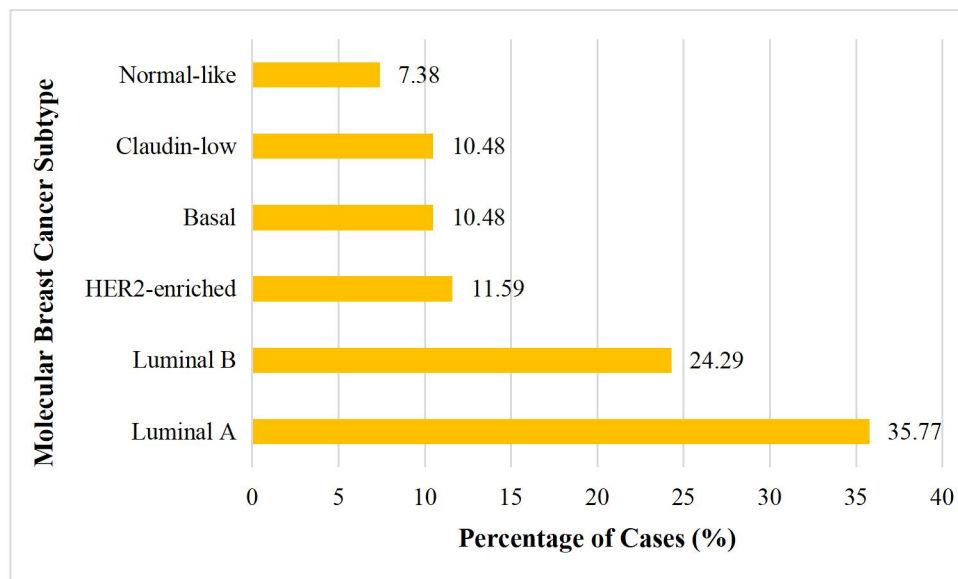


Figure 1. Distribution of Molecular Breast Cancer Subtypes

The figure indicates that there is evident difference in the representation of molecular breast cancer subtypes in the cohort. Luminal subtypes were predominant as compared to HER2-enriched, Basal and Claudin-low and the Normal-like groups that constituted relatively small percentages of the study population.

3.2 Gene-Expression Differences Across Molecular Subtypes

There was also a significant variation in gene-expression pattern between the six molecular subtypes. The most significant variation on subtype was GATA3, then EGFR, MAPT, AURKA, CCNE1, CHEK1, IGF1R, CDC25A, CDK1 and E2F2. These genes were all statistically significant with benchmarking of the Benjamini-Hochberg false discovery rate. Other important genes were APH1B, BCL2, ERBB3, CDK6, NPNT, FANCD2, E2F3, HSD17B4, CCNB1, and NOTCH1. Table 2 summarizes the strongest differences in gene-expression.

Table 2. Major gene-expression differences across molecular subtypes

Gene	Kruskal–Wallis H	p-value	FDR-adjusted p-value
GATA3	979.770	1.44×10^{-209}	7.04×10^{-207}
EGFR	895.973	1.98×10^{-191}	4.84×10^{-189}
MAPT	839.581	3.16×10^{-179}	5.15×10^{-177}
AURKA	803.827	1.72×10^{-171}	2.10×10^{-169}
CCNE1	783.886	3.54×10^{-167}	3.46×10^{-165}

CHEK1	737.350	4.12×10^{-157}	3.35×10^{-155}
IGF1R	726.232	1.04×10^{-154}	7.30×10^{-153}
CDC25A	725.410	1.57×10^{-154}	9.61×10^{-153}
CDK1	711.151	1.91×10^{-151}	1.04×10^{-149}
E2F2	688.009	1.92×10^{-146}	9.40×10^{-145}

3.3 Somatic Mutation Profile of the Cohort

PIK3CA and TP53 were the most altered genes in the cohort and determined by somatic mutation analysis. The PIK3CA mutations were found in 792 cases (41.73%), and the TP53 mutations were found in 659 cases (34.72%). Other frequent changes were MUC16, AHNAK2, KMT2C, SYNE1, GATA3, MAP3K1, DNAH11, AHNAK, and CDH1. The other genes had relatively lower mutation frequencies. The distribution of the most frequent somatic mutations among the 1,898 cases is shown in Table 3.

Table 3. Most frequent somatic mutations in the study population

Gene	Mutated cases	Mutation frequency (%)
PIK3CA	792	41.73
TP53	659	34.72
MUC16	326	17.18
AHNAK2	310	16.33
KMT2C	232	12.22
SYNE1	232	12.22
GATA3	230	12.12
MAP3K1	196	10.33
DNAH11	175	9.22
AHNAK	174	9.17
CDH1	172	9.06

The prevalence of the most prevalent somatic mutations is depicted in Figure 2.

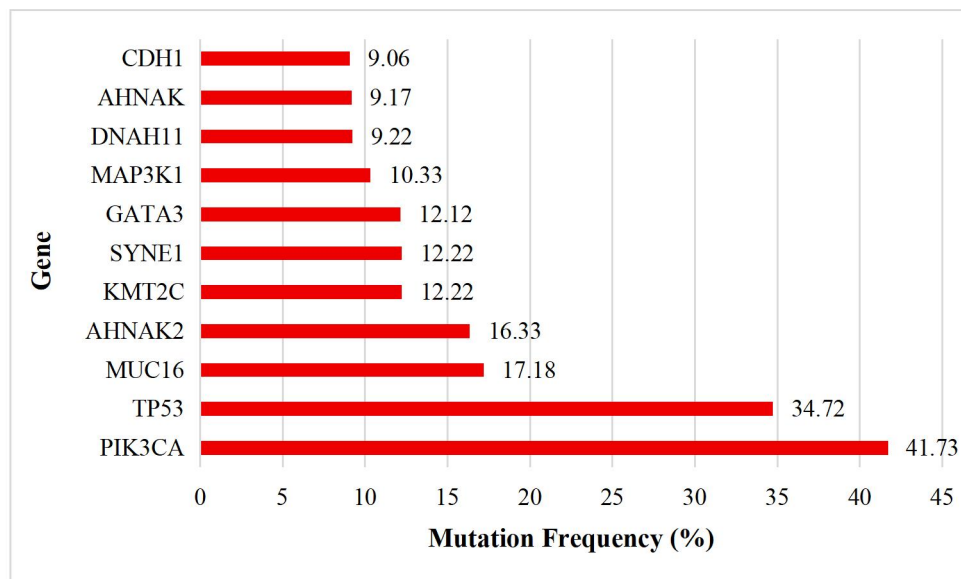


Figure 2. Frequency of Major Somatic Mutations in Breast Cancer

As depicted in the figure, the frequencies of somatic mutations differed significantly among the genes that were studied. PIK3CA and TP53 were commonest mutated genes with less common mutations in MUC16, AHNAK2, KMT2C, SYNE1, GATA3, MAP3K1, DNAH11, AHNAK and CDH1.

3.4 Association of Somatic Mutations with Molecular Subtypes

The frequency of somatic mutation varied greatly with the breast cancer molecular subtypes. TP53 was found to have the best association with the subtypes then PIK3CA and GATA3. After FDR adjustment, significant associations were also found with MAP3K1, CBFB, CDH1, MUC16, SYNE1, AKAP9, DNAH11, RB1, AHNAK, KMT2C, SMAD4 and USH2A. Conversely, THSD7A, LIFR, SMARCB1, PIK3R1, and AFDN were not significant after multiple-testing correction. Table 4 below shows the key mutation-subtype correlations.

Table 4. Association between somatic mutations and molecular subtype

Gene	Chi-square	p-value	FDR-adjusted p-value
TP53	587.624	9.54×10^{-125}	1.65×10^{-122}
PIK3CA	160.848	6.53×10^{-33}	5.65×10^{-31}
GATA3	89.091	1.04×10^{-17}	6.02×10^{-16}
MAP3K1	45.378	1.22×10^{-8}	5.26×10^{-7}
CBFB	38.914	2.47×10^{-7}	8.55×10^{-6}
CDH1	35.973	9.62×10^{-7}	2.77×10^{-5}

MUC16	34.454	1.93×10^{-6}	4.78×10^{-5}
SYNE1	25.812	9.71×10^{-5}	2.10×10^{-3}
AKAP9	23.849	2.32×10^{-4}	3.73×10^{-3}
RB1	23.802	2.37×10^{-4}	3.73×10^{-3}

3.5 Integrated Mutation and Gene-Expression Patterns

Combined analysis revealed that there were significant differences in the levels of relevant gene-expression between mutation-positive and mutation-negative tumors. The median CDH1 expression was significantly lower in cases with CDH1-mutations, and median GATA3 was high in GATA3-mutated cases. CBFB, AKT1, TBX3, GPS2, RB1, PTEN, MAP2K4, NCOR1, SMAD4, TP53, MAP3K1, ERBB2, PDE4DIP, and RUNX1 were also found to have significant differences in their expression after FDR correction. The results show quantifiable associations between the status of mutations and transcriptomic variations, as described in Table 5.

Table 5. Integrated analysis of mutation status and corresponding gene expression

Gene	Mutated n	Non-mutated n	Median expression, mutated	Median expression, non-mutated	FDR-adjusted p-value
CDH1	172	1,726	-1.819	0.195	2.63×10^{-57}
GATA3	230	1,668	0.842	0.178	6.48×10^{-45}
CBFB	91	1,807	-0.637	0.034	2.47×10^{-12}
AKT1	76	1,822	0.742	-0.128	6.66×10^{-10}
TBX3	101	1,797	0.945	-0.101	1.23×10^{-9}
GPS2	21	1,877	-1.471	0.009	2.02×10^{-7}
RB1	49	1,849	-0.840	0.055	1.55×10^{-6}
PTEN	78	1,820	-0.561	0.129	2.86×10^{-6}
MAP2K4	64	1,834	-0.585	-0.122	2.97×10^{-6}
TP53	659	1,239	-0.154	0.011	6.42×10^{-3}

MAP3 K1	196	1,702	0.097	−0.067	9.64×10^{-3}
ERBB2	58	1,840	0.159	−0.183	2.27×10^{-2}

Figure 3 shows the distribution of mutation-positive and mutation-negative cases of the selected genes.

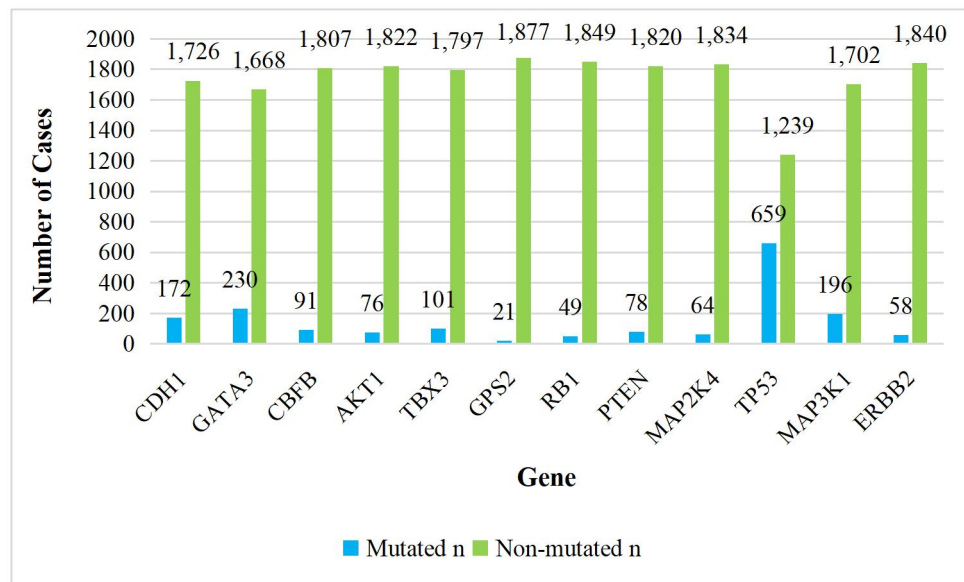


Figure 3. Mutation Status Across Selected Breast Cancer Genes

The figure demonstrates that most of the genes had non-mutated cases as the highest, and cases with mutations were much less across the chosen genes. The TP53 presented the most cases of mutations, then GATA3, MAP3K1 and CDH1.

4. Discussion

The current research revealed a significant molecular heterogeneity in the breast cancer subtypes both at the transcriptomic and genomic scale. Gene-expression analysis revealed significant subtype-associated differences, especially those of GATA3, EGFR, MAPT, AURKA, CCNE1, CHEK1, IGF1R, CDC25A, CDK1, and E2F2, suggesting that hormone-related signaling, proliferation, growth-factor pathways, and cell-cycle regulation differed significantly across molecular groups. These data confirm the opinion that breast cancer subtypes are not biologically similar entities that have varying molecular features. This finding is aligned with the existing molecular classification systems which classify Luminal, HER2-enriched and Basal-like tumors based on their typical transcriptional signatures and clinical phenotype (Tsang & Gary, 2020). The significant difference in gene-expression patterns also indicates that there may be subtype-specific molecular signatures with prognostic and therapeutic implications. Some of the highly differentiated genes in the analysis performed today are those that are linked to tumor proliferation, tumor signaling, and cell-cycle regulation. Other studies have previously utilized breast cancer gene-expression

profiles, and in these studies, transcriptomic analysis identified potential prognostic biomarkers of aggressive disease, especially in triple-negative breast cancer (Thalor et al., 2022). The current results thus support the practicality of expression profiling to differentiate biologically significant groups of breast cancer and to determine molecular characteristics, which can be related to disease behavior.

Somatic mutation analysis also revealed that PIK3CA and TP53 were the most commonly mutated genes with recurrent mutations in MUC16, AHNK2, KMT2C, SYNE1, GATA3, MAP3K1, and CDH1. Frequencies of mutations varied considerably by molecular subtype with TP53 presenting the highest subtype relationship, followed by PIK3CA and GATA3. These results are not new since prior evidence has shown that breast tumors with various HER2-expression levels have different somatic mutational profiles (Zhang et al., 2022). This kind of fluctuation suggests that genomic changes are not evenly distributed among molecular subtypes and can help them to differ in the way they behave biologically. The combined mutation-expression analysis was another piece of evidence that genomic and transcriptomic changes are interrelated. Gene-positive mutations were observed in genes such as CDH1, GATA3, CBFB, AKT1, PTEN, TP53, MAP3K1 and ERBB2 with significant variation in the levels of the respective genes-expressions. This implies that there is the possibility of certain somatic changes having quantifiable transcriptional effects. The same complexity of the molecules has been found in breast cancers that exhibit heterogeneous HER2 amplification, with somatic mutations and copy-number changes playing a role in generating significant intratumoral and intertumoral diversity (Van Bockstal et al., 2020).

These findings may have important clinical implications. The pattern of expression and mutation in subtypes may be useful in providing more sophisticated molecular stratification, identification of biomarkers and targeting of therapeutic strategies. Potential genomic profiling of metastatic breast cancer has shown that molecular characterization has the potential to determine clinically significant changes and assist in treatment decisions (van Geelen et al., 2020). The specifically unique molecular patterns of aggressive subtypes can also be applicable to triple-negative breast cancer, which is very heterogeneous and difficult to treat even with the current molecular characterization (Varzaru et al., 2024). There are some limitations that are to be taken into account. The research was grounded on secondary observational findings and thus causal relationships were not made. There were missing data on some clinical variables and the analysis was mostly based on the aspect of gene-expression and somatic-mutation information but not proteomic, epigenetic and longitudinal treatment-response information. Moreover, both biological and technical variability may affect molecular classification, an issue that has been acknowledged when it comes to breast cancer classification (Zhang, 2023). The results of the study should be confirmed in future research using independent cohorts, combining more layers of omics data, and determining whether the observed molecular patterns are associated with prognosis or response to therapy.

5. Conclusion

This study has shown that there is a high level of molecular heterogeneity among breast cancer subtypes in terms of the joint analysis of gene-expression and somatic-mutation profiles. It was found that there are clear subtype-related variations in genes pertaining to hormone signaling, growth-factor pathways, cell-cycle regulation, and tumor proliferation, with GATA3, EGFR, MAPT, AURKA, CCNE1, CHEK1, IGF1R, CDC25A, CDK1, and E2F2 displaying the strongest variation. Somatic mutation analysis also revealed that PIK3CA and TP53 were the most commonly mutated genes, whereas TP53, PIK3CA, GATA3, MAP3K1 and CDH1 and a number of others exhibited significant associations with molecular subtype. The combined analysis also revealed that mutation status was associated with associated gene-expression alterations of several genes, such as CDH1, GATA3, AKT1, PTEN, TP53, MAP3K1 as well as ERBB2. These results suggest that genomic and transcriptomic changes have a direct relationship and play a combined role in the molecular heterogeneity of breast cancer. These findings are in favor of the importance of combined molecular profiling in enhancing the characterization of the subtypes and the detection of possibly significant biomarkers. But independent cohort validation and additional molecular layer will count. The next round of research should also consider the possibility of these subtype-specific molecular patterns enhancing prognostic evaluation and assist in governing more specific therapeutic approaches in breast cancer.

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. Binder, A., Bockmayr, M., Hägele, M., Wienert, S., Heim, D., Hellweg, K., ... & Klauschen, F. (2021). Morphological and molecular breast cancer profiling through explainable machine learning. *Nature Machine Intelligence*, 3(4), 355-366.
3. Bodily, W. R., Shirts, B. H., Walsh, T., Gulsuner, S., King, M. C., Parker, A., ... & Piccolo, S. R. (2020). Effects of germline and somatic events in candidate BRCA-like genes on breast-tumor signatures. *PloS one*, 15(9), e0239197.
4. Brueffer, C., Gladchuk, S., Winter, C., Vallon-Christersson, J., Hegardt, C., Häkkinen, J., ... & Saal, L. H. (2020). The mutational landscape of the SCAN-B real-world primary breast cancer transcriptome. *EMBO molecular medicine*, 12(10), EMM202012118.
5. Cosgrove, N., Varešlija, D., Keelan, S., Elangovan, A., Atkinson, J. M., Cocchiglia, S., ... & Young, L. S. (2022). Mapping molecular subtype specific alterations in breast cancer brain metastases identifies clinically relevant vulnerabilities. *Nature communications*, 13(1), 514.
6. Creighton, C. J. (2023). Gene expression profiles in cancers and their therapeutic implications. *Cancer Journal (Sudbury, Mass.)*, 29(1), 9.
7. Fumagalli, C., Ranghiero, A., Gandini, S., Corso, F., Taormina, S., De Camilli, E., ... & Barberis, M. (2020). Inter-tumor genomic heterogeneity of breast cancers: comprehensive genomic profile of primary early breast cancers and relapses. *Breast Cancer Research*, 22(1), 107.
8. González-Martínez, S., Pérez-Mies, B., Carretero-Barrio, I., Palacios-Berraquero, M. L., Perez-García, J., Cortés, J., & Palacios, J. (2020). Molecular features of metaplastic breast carcinoma: an infrequent subtype of triple negative breast carcinoma. *Cancers*, 12(7), 1832.
9. Horr, C., & Buechler, S. A. (2021). Breast Cancer Consensus Subtypes: A system for subtyping breast cancer tumors based on gene expression. *NPJ breast cancer*, 7(1), 136.
10. Hutchinson, K. E., Yost, S. E., Chang, C. W., Johnson, R. M., Carr, A. R., McAdam, P. R., ... & Yuan, Y. (2020). Comprehensive profiling of poor-risk paired primary and recurrent triple-negative breast

- cancers reveals immune phenotype shifts. *Clinical Cancer Research*, 26(3), 657-668.
11. Jeong, J. Y., Kim, C. Y., Nam, A. R., & Cho, J. Y. (2026). FGFR4-associated APOBEC3 mutagenesis characterizes HER2-enriched subtype of human breast cancer: Molecular Diagnostics. *British Journal of Cancer*, 1-16.
 12. Jiang, Y. Z., Liu, Y., Xiao, Y., Hu, X., Jiang, L., Zuo, W. J., ... & Shao, Z. M. (2021). Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: the FUTURE trial. *Cell research*, 31(2), 178-186.
 13. Kim, T. M., Yang, I. S., Seung, B. J., Lee, S., Kim, D., Ha, Y. J., ... & Kim, S. (2020). Cross-species oncogenic signatures of breast cancer in canine mammary tumors. *Nature communications*, 11(1), 3616.
 14. Liu, H., Qiu, C., Wang, B., Bing, P., Tian, G., Zhang, X., ... & Yang, J. (2021). Evaluating DNA methylation, gene expression, somatic mutation, and their combinations in inferring tumor tissue-of-origin. *Frontiers in cell and developmental biology*, 9, 619330.
 15. Lupat, R., Perera, R., Loi, S., & Li, J. (2023). Moanna: multi-omics autoencoder-based neural network algorithm for predicting breast cancer subtypes. *Ieee Access*, 11, 10912-10924.
 16. Pan, J. W., Zabidi, M. M. A., Ng, P. S., Meng, M. Y., Hasan, S. N., Sandey, B., ... & Teo, S. H. (2020). The molecular landscape of Asian breast cancers reveals clinically relevant population-specific differences. *Nature communications*, 11(1), 6433.
 17. Pareja, F., Ferrando, L., Lee, S. S., Beca, F., Selenica, P., Brown, D. N., ... & Reis-Filho, J. S. (2020). The genomic landscape of metastatic histologic special types of invasive breast cancer. *NPJ breast cancer*, 6(1), 53.
 18. Szymiczek, A., Lone, A., & Akbari, M. R. (2021). Molecular intrinsic versus clinical subtyping in breast cancer: A comprehensive review. *Clinical genetics*, 99(5), 613-637.
 19. Testa, U., Castelli, G., & Pelosi, E. (2020). Breast cancer: a molecularly heterogenous disease needing subtype-specific treatments. *Medical Sciences*, 8(1), 18.
 20. Thakor, A., Kumar Joon, H., Singh, G., Roy, S., & Gupta, D. (2022). Machine learning assisted analysis of breast cancer gene expression profiles reveals novel potential prognostic biomarkers for triple-negative breast cancer.
 21. Tsang, J. Y., & Gary, M. T. (2020). Molecular classification of breast cancer. *Advances in anatomic pathology*, 27(1), 27-35.
 22. Van Bockstal, M. R., Agahozo, M. C., van Marion, R., Atmodimedjo, P. N., Sleddens, H. F., Dinjens, W. N., ... & van Deurzen, C. H. (2020). Somatic mutations and copy number variations in breast cancers with heterogeneous HER2 amplification. *Molecular Oncology*, 14(4), 671-685.
 23. van Geelen, C. T., Savas, P., Teo, Z. L., Luen, S. J., Weng, C. F., Ko, Y. A., ... & Loi, S. (2020). Clinical implications of prospective genomic profiling of metastatic breast cancer patients. *Breast Cancer Research*, 22(1), 91.
 24. Varzaru, V. B., Vlad, T., Popescu, R., Vlad, C. S., Moatar, A. E., & Cobec, I. M. (2024). Triple-negative breast cancer: molecular particularities still a challenge. *Diagnostics*, 14(17), 1875.
 25. Zhang, G., Ren, C., Li, C., Wang, Y., Chen, B., Wen, L., ... & Liao, N. (2022). Distinct clinical and somatic mutational features of breast tumors with high-, low-, or non-expressing human epidermal growth factor receptor 2 status. *BMC medicine*, 20(1), 142.
 26. Zhang, X. (2023). Molecular classification of breast cancer: relevance and challenges. *Archives of pathology & laboratory medicine*, 147(1), 46-51.

This article is licensed under a [Creative Commons Attribution 4.0 International \(CC BY NC ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/) License; copyright remains with the author(s).