ORIGINAL ARTICLE

Refining prognostic tools for luminal breast cancer: genetic insights and comprehensive analysis

L. A. Sinberger[1], T. Zahavi[1], N. Keren-Khadmy[2], Y. Dugach[2], A. Sonnenblick[2,3]* & M. Salmon-Divon[1,4]*

[1]Department of Molecular Biology, Ariel University, Ariel; [2]Institute of Oncology, Tel Aviv Sourasky Medical Center, Tel Aviv; [3]Faculty of Medicine, Tel Aviv University, Tel Aviv; [4]Adelson School of Medicine, Ariel University, Ariel, Israel

CHECK FOR UPDATES

Available online xxx

Background: Luminal breast cancer (BC) is generally associated with a lower risk of recurrence compared with other subtypes. However, patients with luminal BC can still experience recurrence, which remains a significant concern and contributes to BC-related mortality. Current clinical practice for recurrence risk prognosis relies on prognostic tests based on tumor gene expression profiles.

Materials and methods: In this study, we aimed to investigate the association between different genetic alterations with the likelihood of recurrence and gene expression prognostic prediction (Oncotype DX®, MammaPrint®, and PAM50-ROR) in luminal BC patients. We constructed three transcriptome-based predictive models, based on these widely used clinical tests, to evaluate the recurrence risk of patients with luminal BC, using RNA-seq data from 1527 samples across 11 datasets. We further classified 1780 patients from the TCGA and METABRIC datasets into risk groups and detected distinct recurrence risk patterns.

Results: Our analysis revealed that low-risk groups had higher frequencies of mutations in *PIK3CA*, *MAP3K1*, *CDH1*, *KMT2C*, and *CBFB*, as well as co-mutations in *PIK3CA-MAP3K1*, *PIK3CA-CBFB*, and *KMT2C-MAP3K1*. In contrast, high-risk groups showed enrichment of *TP53*, *RB1*, and *PTPN22* mutations compared with the whole cohort, with notable co-mutations in *TP53-PIK3CA* and *TP53-KMT2C*. Furthermore, mutations in *TP53* and *BRCA2*, and deletions in the 7p22.3 region were at least threefold more frequent in high-risk patients compared with low-risk patients. Using an independent dataset, we validated our finding of higher frequency of *BRCA2* mutations in Oncotype DX® high-risk patients. Notably, *PIK3CA* mutations had an unexpected negative impact on recurrence and survival among high-risk patients.

Conclusion: Our study reveals key genetic factors associated with recurrence risk in luminal BC. Identifying these mutations and copy number alterations provides a basis for refined prognostic models and suggests avenues for further research, potentially improving treatment strategies and follow-up care for patients with luminal BC.

Key words: luminal BC, PIK3CA mutation, recurrence, Oncotype DX, PAM50, MammaPrint

INTRODUCTION

Breast cancer (BC) is the most frequently diagnosed cancer among females worldwide, with 2.2 million new cases reported in 2020.[1-3] In the United States, the incidence of BC demonstrates a gradual annual increase of ∼0.5%. This upward trend primarily stems from the diagnosis of localized and hormone receptor-positive cases.[4] Globally, BC exhibits a significant recurrence rate, ranging from 20% to 30%. Notably, disease recurrence constitutes the primary cause of BC-related mortality. None the less, over the past three decades, advancements in diagnostic and treatment options have led to an increasing survival rate.[5-7]

BC tumors can be classified using several methods including transcriptomic signature, tumor histology, and the TNM (tumor—node—metastasis) staging system.[3,8] One approach to classifying BC is based on the gene expression patterns of the cells, which classify BC into five distinct subgroups: (i) luminal A, (ii) luminal B, (iii) the human epidermal growth factor receptor 2 (HER2) subgroup characterized by *HER2* overexpression or amplification, (iv) basal-like tumors typically lacking hormone receptor or HER2 expression, and (v) normal-like tumors.[9-12]

Luminal BC constitutes ∼60%-70% of BC diagnoses and is associated with increased patient survival rates and reduced relapse compared with other BC subtypes. Gene expression signatures allow for distinguishing between luminal A and B tumors. Luminal A tumors exhibit high expression of

 1

progesterone receptors (PR) and estrogen-related genes, and low expression of proliferation genes like Ki67. Conversely, luminal B tumors are characterized by low expression of estrogen receptor (ER) and PR and high expression of genes related to cell cycle, proliferation, and growth factor receptors. Luminal B tumors are more aggressive, with a higher 5-year recurrence rate compared with luminal A tumors.[3,12-14] Overall, luminal tumors have a slower growth rate than ER-negative tumors, resulting in a better patient prognosis. In recent years various genomic prognostic tests have emerged to predict prognosis and treatment options for patients with BC. These include Oncotype DX®, MammaPrint®, PAM50, Endopredict, Breast Cancer Index, and others, all based solely on the gene expression profiles of tumors.[3,8,15,16] The development of BC involves the accumulation of various genetic changes, encompassing germline mutations,[17-19] driver mutations,[20] and copy number alterations (CNAs).[21,22] Advancements in next-generation sequencing technologies, which provide insights into gene sequences and CNAs, have enabled the integration of such data into routine clinical practices for individuals with cancer.[23,24]

In this article, we aim to investigate whether genomic alterations are associated with recurrence risk in patients with luminal BC. By correlating the presence of genetic alterations with genomic prognostic prediction, there is potential for a more comprehensive assessment of a patient's risk of recurrence. This, in turn, could enhance the precision of treatment strategies and follow-up protocols.[22,25,26]

MATERIALS AND METHODS

Data collection

We obtained gene expression data of 3083 luminal samples from 11 BC datasets (CAL,[27] DFHCC,[28] GSE58644,[29] MAINZ,[30] METABRIC,[31] NKI,[32] TCGA,[21] TRANSBIG,[33] UNT,[34] UPP,[35] and VDX[36]) using the MetaGxBreast R package (version 1.18.0),[37] together with their clinical information (Supplementary Tables S1 and S2, respectively, available at https://doi.org/10.1016/j.esmoop.2025.105080). The microarray gene expression data from each study were downloaded and integrated into a unified database. Batch effect correction according to the study was done using the removeBatchEffect function in the limma R package (version 3.54.1).[38] This database was used to construct the recurrence risk models as described in the Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2025.105080. PAM50 classification, mutation, and CNA data were available for 1780 patients with luminal BC sourced from the TCGA and METABRIC datasets and were downloaded via the cBioPortal.[39,40] PAM50 classification for the remaining samples in the MetaGxBreast studies was carried out using the genefu R package (version 2.30.0).[41]

Copy number alteration (CNA)

Affymetrix 6.0 SNP array segment data file (seg) for luminal patients in the TCGA PanCanAtlas study was obtained from the cBioPortal.[39,40] The GISTIC 2.0 software (Broad Institute,

Cambridge, MA)[42-44] accessed through the GenePattern portal,[45] was employed to analyze the segment data separately for each risk group against the reference genome GRCh37 (hg19), with a confidence level of 0.75 as suggested in the software documentation. To identify reported CNAs, we compared the resulting peak locations with focal amplifications or deletions found in the TCGA[21] and BIG-1-98[46] studies (38 amplification peaks and 71 deletion peaks, Supplementary Table S3, available at https://doi.org/10.1016/j.esmoop.2025.105080). To determine the frequency of TCGA and BIG reported structural variations in different risk groups, we used BEDTools[47] to find the intersection between GISTIC result frequency and the examined peaks. Fisher's exact test was applied to assess the statistical significance of the differences in the number of samples with specific CNA between the high- and low-risk groups. CNA frequencies were converted into categorical sample counts before carrying out the test.

Identification of differentially expressed genes (DEGs)

The limma[38] (version 3.54.1) and edgeR[48] (version 3.40.2) R packages were used for detection of differentially expressed genes (DEGs). Unless specified otherwise, a false discovery rate (FDR) of ≤ 0.05 and log fold change (logFC) ≥ 1.5 were used as the statistical cut-off for differential expression.

Pathways enrichment analysis

Pathways enrichment analysis was carried out against the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database through the WebGestalt 2019[49] tool. DEGs found in the comparison of patients classified as high-risk across all models versus patients classified as low-risk across all models were used as input for over-representation analysis. Ranked gene lists with their corresponding logFC from the above comparison were used as input for gene set enrichment analysis (GSEA). Only ontologies and pathways with FDR ≤ 0.05 are shown.

BRCA2 validation cohort

For the independent validation, we accessed the Sourasky institutional Oncotype DX registry. Clinical and pathological data were extracted from the electronic medical records of patients diagnosed with luminal BC between 2004 and 2020. Our inclusion criteria focused on luminal patients who underwent mutation panels, and we collected relevant data on primary tumor characteristics, including Oncotype DX results and the presence of BRCA2 mutation. The study design adhered to the tenets of the Declaration of Helsinki and was approved by the local institutional review board (TLV 19-449 and TLV 18-426).

Statistical analysis and plotting

All statistical analyses were carried out using the R statistical framework (version 4.2.2).[50] The ggplot2[51] (version 3.4.1) R package was used for generating plots, and the plotly[52] (version 4.10.1) R package was used for generating

interactive plots. Venn diagrams were build using the VennDiagram[53] R package (version 1.7.3). Upset diagrams were prepared using the ComplexHeatmap[54] (version 2.14.0) R package. Survival curves were estimated using the Kaplan—Meier method and visualized by the Survival[55] (version 3.5.3) and Survminer[56] (version 0.4.9) R packages and P values were calculated using the log-rank test. Forest plots were visualized using the forestploter[57] (version 1.1.0) R package. Mutation enrichment was calculated by dividing the observed mutation frequency to the expected frequency. Enrichment significance was tested using Fisher's exact test, followed by multiple testing correction using the FDR method to control for multiple comparisons. FDR $\leq$ 0.05 was used to determine statistical significance.

RESULTS

Development of models to predict the recurrence risk of patients with luminal BC

In this work, we sought to find genomic alterations associated with recurrence risk in luminal BC. To overcome limited access to commercial prognostic test information, we created three transcriptome-based predictive models, to evaluate the recurrence risk of patients with luminal BC based on the commercial tests Oncotype DX®, MammaPrint®, and PAM50-ROR, which are commonly used in clinical practice for patients with luminal BC. We utilized these models on 1527 samples using RNA-seq data of patients from 11 datasets (Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2025.105080) and for each patient, we calculated three different scores based on the genes that are used in the original commercial tests. The clinicopathological characteristics of the patients included in the model building are summarized in Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2025.105080. The median age of the patients was 58.4 years, with 69.6% of patients being >50 years old. Tumors of grade 2 comprised 50.1% of cases, while 19.8% were grade 1. Additionally, 99% of the tumors measured <5 cm. The median follow-up duration was 8.2 years. Cut-off values for defining low-risk and high-risk groups were established for each model as outlined in the 'Materials and Methods' section (Supplementary Figure S1A-C, available at https://doi.org/10.1016/j.esmoop.2025.105080). To get more confidence in the classification of patients into low- and high-risk groups, we assessed the recurrence rates based on the known clinical information regarding disease recurrence within each group. Our results demonstrate that, on average, 28% of patients classified as high-risk experienced disease recurrence, while only 8% in the low-risk group did so. (Supplementary Figure S1D-F, available at https://doi.org/10.1016/j.esmoop.2025.105080). Cox regression analysis for recurrence risk showed hazard ratio (HR) between 2.9 and 4.9 (P value <0.01) for high-risk groups across all models (Supplementary Figure S1G, available at https://doi.org/10.1016/j.esmoop.2025.105080). In addition, we examined the odds ratio for disease recurrence as performed in van't Veer et al., 2002[32]; in all models the high-risk group had odds ratio (OR) between

3.07 and 5.54 to experience disease recurrence [Fisher's exact test, OR 5.54, 95% confidence interval (CI) 2.14-17.04, $P = 7.36E\text{-}05$ for the Oncotype-like model, OR 3.99, 95% CI 2.84-5.66, $P < 2.2E\text{-}16$ and OR 3.07, 95% CI 1.76-5.4, $P = 3.67E\text{-}05$ for the MammaPrint-like train and test sets, respectively, and OR 5.17, 95% CI 3.43-7.99, $P < 2.2E\text{-}16$ for the PAM50-ROR model] (detailed statistical values in Supplementary Table S4, available at https://doi.org/10.1016/j.esmoop.2025.105080). These results highlight the predictive efficacy of the developed models and confirms their use in this study.

Next, we used our models to classify 1780 patients with luminal A and luminal B BC subtypes, with accessible information regarding mutations and CNA from the TCGA and METABRIC datasets. Patients were classified into risk groups for each model (Figure 1A-C). Figures 1D and E show the overlap in patient classification among different models. In total, 206 patients were consistently classified as low-risk, and 336 patients were consistently classified as high-risk across all three models. Notably, we observed 296 samples exclusively belonging to the MammaPrint-like low-risk group and 90 samples exclusively in the MammaPrint-like high-risk group. The observed differences in patient classification across models occur because each model is based on a unique set of genes (Supplementary Figure S2A and B, available at https://doi.org/10.1016/j.esmoop.2025.105080).

While differences exist in the proportion of luminal A, and luminal B patients across risk groups, we observed a substantial subset (20%-30%) of luminal A patients classified as high-risk in the different models (Supplementary Figure S3, available at https://doi.org/10.1016/j.esmoop.2025.105080).

To further investigate whether the identified risk groups accurately reflect the corresponding risk for recurrence, we employed Cox regression analysis on the TCGA and METABRIC dataset samples, to assess the HR of survival and recurrence between the high- and low-risk groups in each model. As expected, the results demonstrate that the high-risk groups exhibit higher HR than the low-risk groups indicating shortened disease-free survival and overall survival (Figure 1F). To account for potential confounders, we also conducted multivariable Cox regression analyses, incorporating TNM stage, age, and grade as covariates (Supplementary Figure S4, available at https://doi.org/10.1016/j.esmoop.2025.105080). In this analysis the strong associations between the high-risk group and recurrence remained significant after adjusting for confounders (P value <0.01), confirming their independent prognostic value. Among the covariates, the TNM stage was a significant predictor of recurrence, whereas age and grade were not significant.

Association between recurrence risk and mutations in luminal BC

We conducted an enrichment analysis to investigate the association between different mutations and recurrence
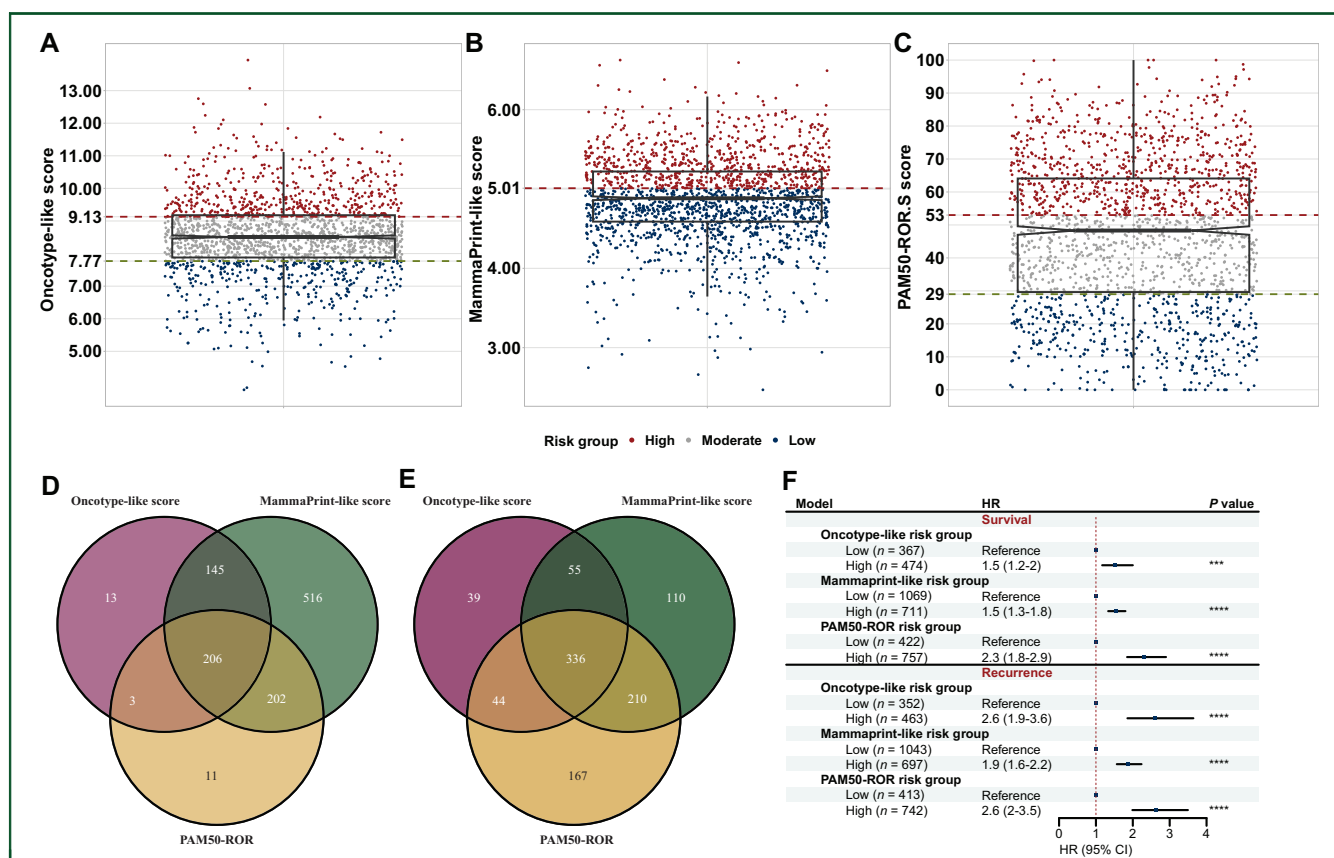
Figure 1. Classification of patients with luminal BC from TCGA and METABRIC databases into risk groups according to the various models. (A-C) Distribution of patients' risk scores in the different models. (A) Oncotype-like score model, (B) MammaPrint-like score model, and (C) PAM50-ROR model. Red indicates patients classified as high-risk, gray indicates moderate risk, and blue indicates low-risk patients. (D, E) Venn diagrams representing the intersection of samples included in the (D) low-risk groups and (E) high-risk groups in each model. (F) Hazard ratio (HR) for patients with luminal BC survival and recurrence based on the risk groups in the different models. The data were analyzed using the univariate Cox proportional hazards regression model. (***$P \leq 0.001$, ****$P \leq 0.0001$).
CI, confidence interval.

risk groups. We used four published curated lists of cancer driver genes including one pan-cancer list[58] and three BC lists.[20,21,59] (Supplementary Table S5, available at https://doi.org/10.1016/j.esmoop.2025.105080). Our findings revealed that mutations in *PIK3CA*, *MAP3K1*, *CDH1*, *KMT2C*, CBFB, and AKT1 were enriched in the low-risk groups, while mutations in *TP53*, *RB1*, and *PTPN22* were enriched in the high-risk groups (FDR < 0.05, detailed statistical values in Supplementary Table S5, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 2A).

We next examined the enrichment of co-occurring mutations and their association with distinct risk groups. In line with the enrichment pattern of individual mutations, we observed co-occurrence of mutations in *PIK3CA* and *MAP3K1*, *KMT2C* and *MAP3K1*, and *CBFB* and *PIK3CA* within the low-risk groups (FDR < 0.05, detailed statistical values in Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 2B). Conversely, co-mutations involving *PIK3CA* and *TP53* as well as *KMT2C* and *TP53* were enriched in high-risk patients, despite the individual mutations of *PIK3CA* and *KMT2C* being more prevalent in low-risk patients (FDR < 0.05, detailed statistical values in Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 2B).

As different mutation frequencies were observed between the high-risk and low-risk groups using various models, this prompted us to identify genes that exhibited at least a threefold difference in mutation frequencies between the risk groups across the different models. For each mutation, we calculated the mutation frequency rate as the ratio between the mutation frequency in the high- and low-risk groups. Our results consistently demonstrated that *TP53* and *RB1* had at least three times higher mutation rates in the high-risk group compared with the low-risk group (Fisher's exact test, *P* value <0.05 across all models). In contrast, *BRCA2* and *MEN1* exhibited higher mutation rates in the high-risk group only within the Oncotype-like score model (Fisher's exact test, *P* value <0.05, detailed statistical values in Supplementary Table S7, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 2C).

We used an independent cohort to validate the finding of differences in *BRCA2* mutation frequency between high- and low-risk according to the Oncotype-like score model (Sourasky). We collected information on primary luminal tumors, including Oncotype DX score, Oncotype DX risk group, and the presence of *BRCA2* mutations. Our dataset is composed of 55 patients, with 29 classified as low-risk and
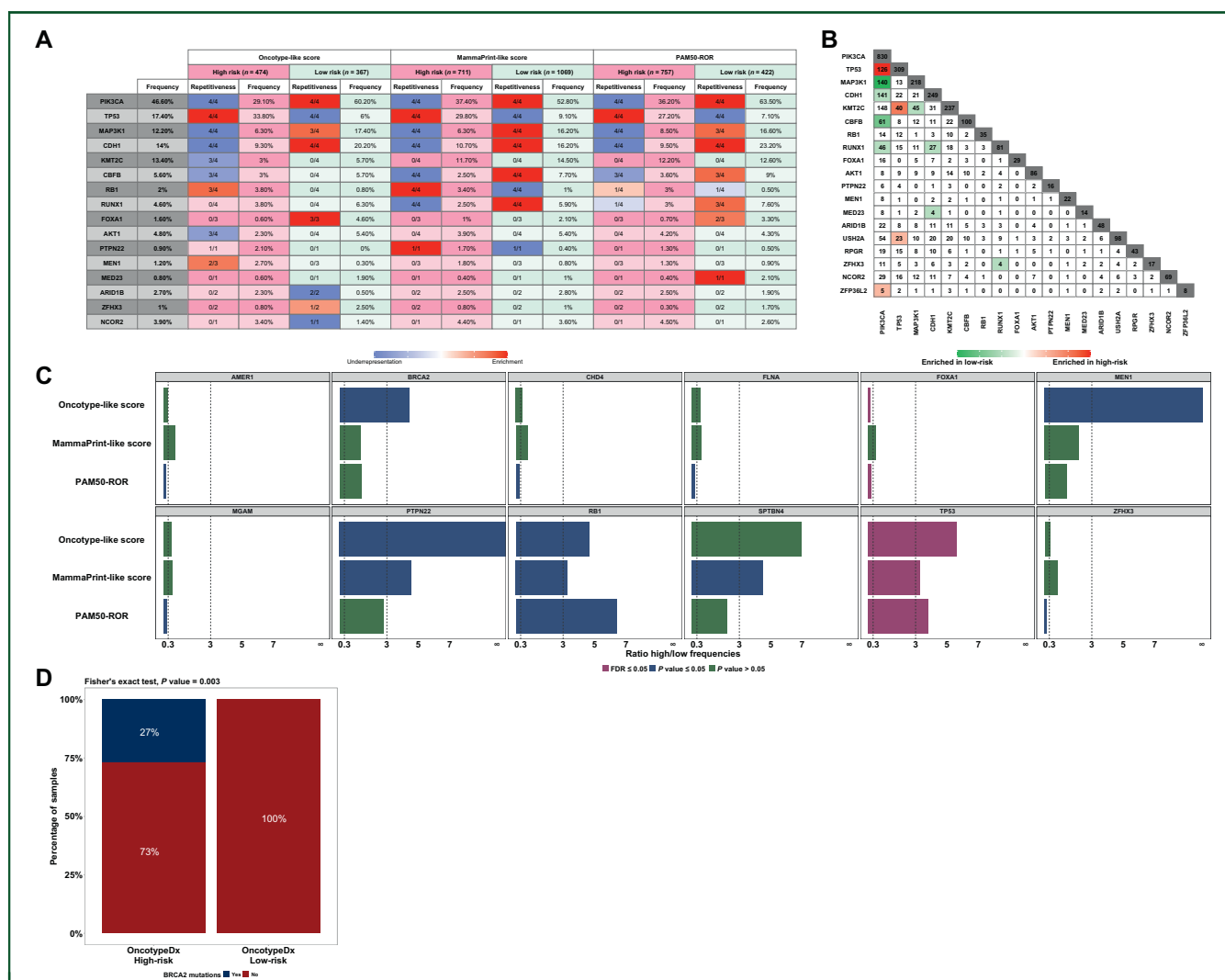
Figure 2. Mutations enrichment according to risk groups. (A) Enrichment and frequency of individual mutations within each risk group. The color range indicates whether the mutation is underrepresented (blue) or enriched (red). For each gene in every model, in every risk group, we counted the number of times it was significantly (FDR < 0.05) enriched or underrepresented across the lists that gene was included in (repetitiveness column). (B) Enrichment of co-occurring mutations in each risk group was determined by Fisher's exact test with FDR < 0.05. The red and green colors indicate significant enrichment in the high-risk group and in the low-risk group, respectively (in at least one out of the three modules. As the color intensity deepens, the enrichment becomes significant in more models). The number displayed in the gray tiles represents the total mutation count in a single gene across all patients. (C) Bar graph representing frequency ratio between high- and low-risk groups across different models of the selected 12 known cancer driver mutations with at least threefold frequency ratio between risk groups in one or more of the models. The color indicates the significance obtained from Fisher's exact test between the risk groups, with FDR < 0.05 (dark pink), P values <0.05 (blue), and not-significant (green). (D) The frequency of BRCA2 mutations in low- and high-risk groups as classified by the patients' Oncotype DX score (Oncotype DX score <11 is defined as low-risk and >26 is considered high-risk).
FDR, false discovery rate.

26 as high-risk. Our results demonstrate that *BRCA2* mutations are more frequent in Oncotype DX high-risk patients (Fisher's exact test, *P* = 0.003) (Figure 2D). This result supports our findings regarding *BRCA2* mutation frequency in the Oncotype-like score model.

Collectively, these results provide strong evidence of a distinguishable mutation pattern associated with the examined risk groups.

Accumulation of copy number alterations across low- and high-risk groups

To identify CNAs associated with the different risk groups, we reanalyzed the segment data file of TCGA patients using GISTIC 2.0.[42] Specifically, we focused on previously published CNAs associated with patients with luminal BC by the TCGA network[21] and the BIG-1-98 clinical trial.[46]

We found varying degrees of overlap between the previously described CNAs and the alterations we identified for the different risk groups. Specifically, we identified amplification peaks in a specific region of 14q21.1, which includes the genes *FOXA1*, *NKX2-1*, and *NFKBIA*, previously associated with patients with luminal B BC.[21] These alterations were consistently identified in the high-risk groups across all three models (chr14:37968637-38182419), while no amplification peaks were detected in the low-risk groups for these genomic locations (FDR < 0.0001 across all models, detailed statistical values in Supplementary Table S8, available at https://doi.org/10.1016/j.esmoop.2025.105 080, Figure 3A). Conversely, across all the models, the
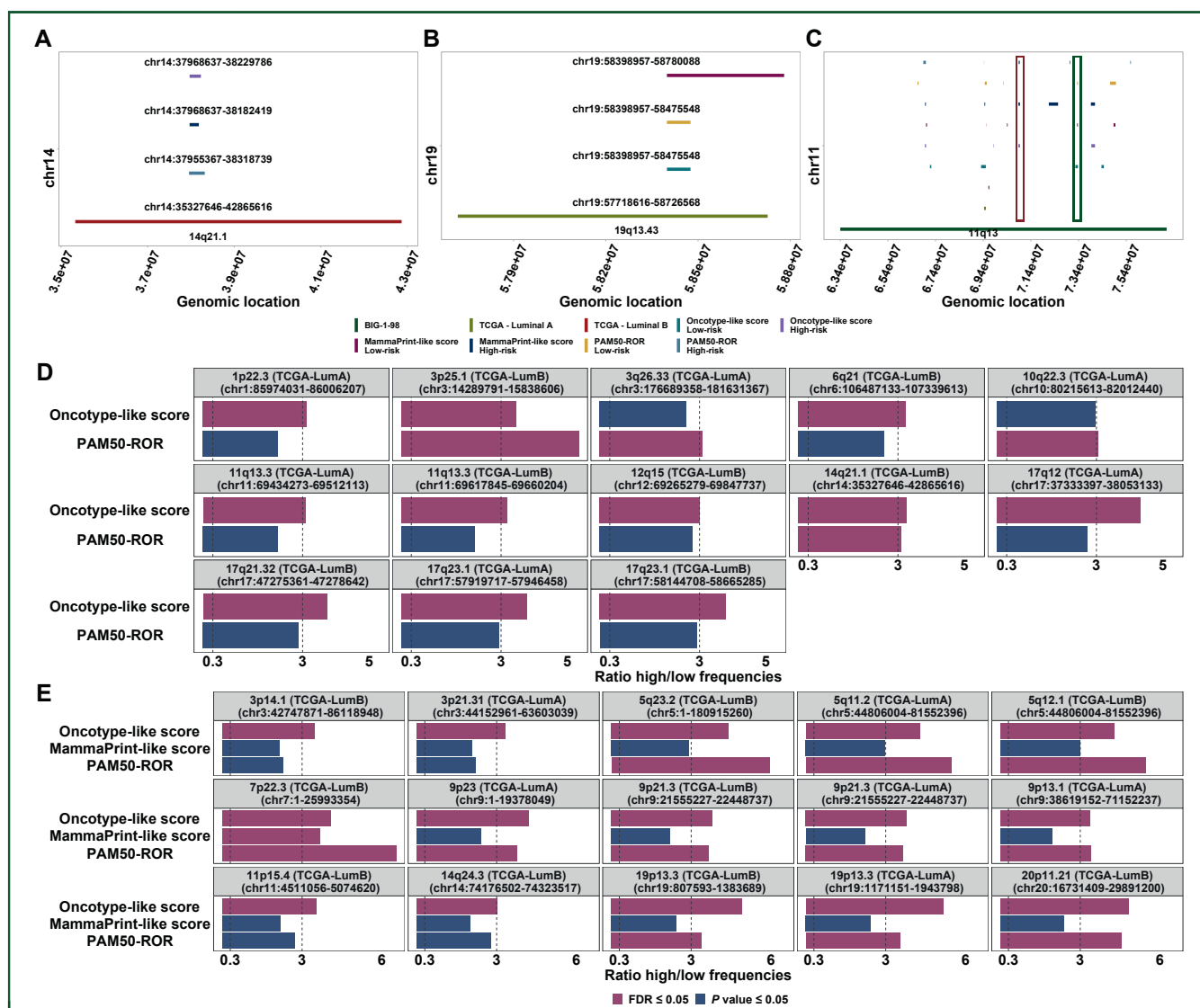
Figure 3. Copy number alterations associated with the different risk groups. (A, B) Overlapping of amplification peaks identified in the different risk groups with previously published CNAs associated with luminal BC across all models. (A) Cytoband 14q21.1 peak associated with high-risk groups. (B) Cytoband 19q13.43 is associated with the low-risk group. (C) Cytoband 11q13 contains different amplification regions, one is amplified exclusively in high-risk groups (highlighted in red square), and the other is amplified exclusively in low-risk groups (highlighted in green square). (D-E) Bar graphs representing the frequency ratio of amplification (D) and deletion (E) peaks between high- and low-risk groups across different models. Peaks were selected if they have at least a threefold frequency ratio between risk groups in one or more of the models. The color indicates the significance obtained from Fisher's exact test between the risk groups, with FDR < 0.05 (dark pink), and P values <0.05 (blue).

CNAs, copy number alterations; FDR, false discovery rate.

amplification peak in cytoband 19q13.43, linked to luminal A tumors,[21] showed partial overlap (chr19: 58398957-58475548) with amplification peaks only in the low-risk groups (FDR < 0.05 across all models, detailed statistical values in Supplementary Table S8, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 3B). Similarly, we identified specific patterns of deletion regions unique to either the high- or low-risk group (Supplementary Figure S5, available at https://doi.org/10.1016/j.esmoop.2025.105080).

While some of the CNAs showed association exclusively with specific risk groups, others exhibited overlaps in part with high-risk groups and in different regions with low-risk groups. For instance, in the cytoband 11q13, which includes the genes *CCDN1*, *FGF3*, *FGF4*, and *FGF19*, which were

linked to an increased risk of distant recurrence in luminal BC,[46] we identified a region that was amplified exclusively in high-risk groups across all models (chr11: 70877100-70943301) and another region that was amplified exclusively in low-risk groups across all the models (chr11:73342272-73375868) (FDR < 0.0001 across all models, detailed statistical values in Supplementary Table S8, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 3C).

Next, we detected CNA regions known to be associated with luminal BC (as reported by TCGA and BIG-1-98[21,46]) that displayed at least a threefold difference in frequencies between the high- and low-risk groups. Using the Oncotype-like score model, we identified 11 amplification peaks meeting this criterion, whereas using the PAM50-model

revealed only four amplification peaks with such distinction. However, no amplification peaks meeting this cut-off were detected with the MammaPrint-like score model. Two amplification peaks, 3p25.1 and 14q21.1, were consistently identified in both the Oncotype-like score and PAM50-models (Fisher's exact test, FDR < 0.05 across all models, detailed statistical values in Supplementary Table S9, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 3D).

As for deletion peaks, we identified 15 regions with at least a threefold difference in frequencies between the high- and low-risk groups according to the Oncotype-like score model. Out of these deletion peaks, 11 were found to be significant according to the PAM50-model. Additionally, one deletion peak, 7p22.3, was found to be significant across all three models (Fisher's exact test, FDR < 0.05 across all models, detailed statistical values in Supplementary Table S10, available at https://doi.org/10.1016/j.esmoop.2025.105080, Figure 3E).

Characterizing patients sharing the same risk group across all models

Once we demonstrated the association of various mutations and CNAs with high- or low-risk groups using different models, we aimed to profile patients who were consistently categorized as either low-risk or high-risk across all models (206 and 336 patients, respectively), defined as 'common low-risk' and 'common high-risk', respectively (clinical information represented in Supplementary Table S11, available at https://doi.org/10.1016/j.esmoop.2025.105080). Our analysis revealed a notable enrichment of TP53 mutations and co-mutations involving *TP53-PIK3CA* among high-risk patients. Furthermore, mutations in *PIK3CA*, *CDH1*, *MAP3K1*, *CBFB*, and *FOXA1* were found to be enriched within the low-risk patient subset, along with co-mutations involving *PIK3CA* and the other five mutated genes (Fisher's exact test, FDR < 0.01, Figure 4A-C, detailed statistical values in Supplementary Table S12, available at https://doi.org/10.1016/j.esmoop.2025.105080).
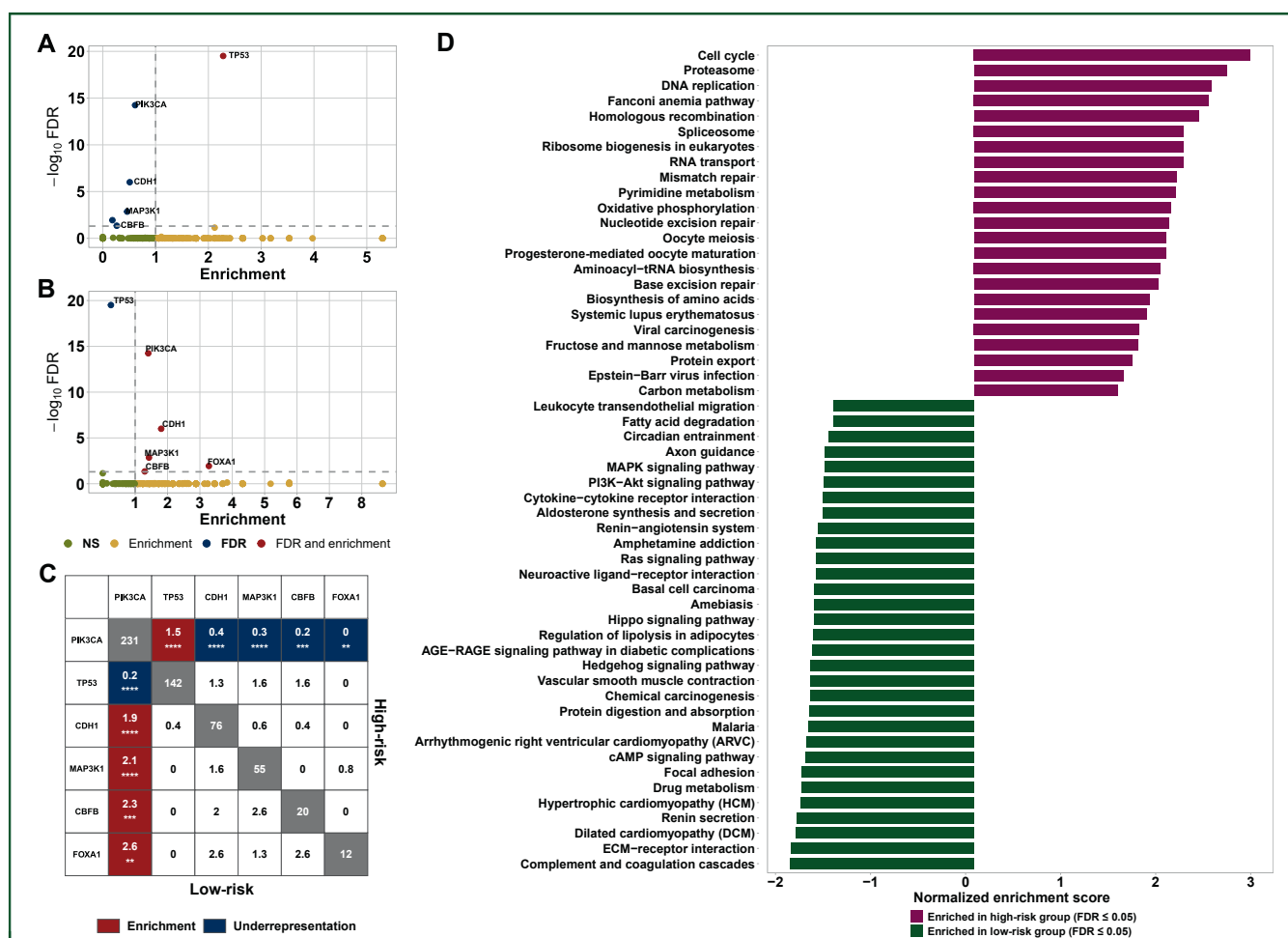

Figure 4. Characterization of common high- and low-risk patients. (A, B) Enrichment of mutations in patients classified as (A) high-risk or (B) low-risk across all three models. The *P* value for the enrichment was calculated using Fisher's exact test. (C) Enrichment of mutations coexistence in patients classified into the same risk group in all three models. The upper panel (the triangle above the diagonal) represents enrichment of the co-mutations calculated in the common high-risk group, and the lower panel (the triangle below the diagonal) represents enrichment of the co-mutations calculated in the low-risk group. The colors represent statistically significant enrichment (red), or significant underrepresentation (blue) calculated using Fisher's exact test (FDR < 0.05). The numbers along the diagonal represent the frequency of mutations in the particular gene across the entire cohort. (D) A gene set enrichment analysis (GSEA) of the differentially expressed genes (DEGs) sampled at the comparison between patients classified as high-risk across all three models versus patients classified as low-risk across all three models. ECM, extracellular matrix; FDR, false discovery rate; NS, not significant.

Conducting a GISTIC analysis on these patients' groups identified several distinct CNAs exclusive to the high- or low-risk categories. For instance, the amplification peak at cytoband 14q21.1, previously associated with luminal B BC[21] and identified in Figure 3A as associated with high-risk, was linked to the high-risk group. On the other hand, the amplification peak at cytoband 19q13.43, previously associated with luminal A BC[21] and shown in Figure 3B as associated with low-risk, was specifically associated with the low-risk group. We also identified the deletion peak at cytoband 17p13.3, previously associated with luminal B BC[21] overlapped with peaks found solely in the high-risk groups (FDR < 0.0001, Supplementary Table S8, and Figure S5, available at https://doi.org/10.1016/j.esmoop.2025.105080).

To gain deeper insights into the differences between the high- and low-risk groups, we carried out a differential gene expression analysis comparing common high-risk samples to common low-risk samples. We identified 139 up-regulated genes and 254 downregulated genes. GSEA against the KEGG dataset indicated an up-regulation of cell-cycle-related pathways in high-risk patients, whereas low-risk patients displayed up-regulation of the MAPK signaling pathway and the PIK3-Akt signaling pathway (Figures 4D).

Survival of high-risk patients is affected by different mutations

To assess the prognostic implications of mutations associated with specific risk groups on survival outcomes, we focused on mutations within the TP53 and PIK3CA genes, which were the most prevalent across our three models. Firstly, a univariate Cox regression model was used to examine the effect of these on patients' recurrence and survival. As shown in Figure 5A and B, TP53 mutations were associated with an increased risk of recurrence and low survival (HR 1.5, 95% CI 1.2-1.7, P value 6.2e-05) for all patients, while PIK3CA mutations did not significantly impact the outcome (recurrence HR 1.1, 95% CI 0.9-1.3, P = 0.536; survival HR 1.0, 95% CI 0.9-1.2 P = 0.625). Secondly, we explored the combined effect of mutations and risk classification on patients' outcome. Interestingly, we observed an unexpected increase in HR for recurrence and lower survival when high-risk patients harbored PIK3CA mutations in all models. Next, to further validate these findings, we conducted a stratified univariate and multi-variate Cox regression analysis within risk groups, as shown in Supplementary Figures S6 and S7, available at https://doi.org/10.1016/j.esmoop.2025.105080. PIK3CA mutations consistently correlate with worse outcomes in high-risk patients across all models, even after adjusting to TP53 mutations (P value <0.05, Supplementary Figures S6 and S7, available at https://doi.org/10.1016/j.esmoop.2025.105080), reinforcing their potential role as prognostic bio-markers. Kaplan–Meier analysis of high-risk patients according to the Oncotype-like score model revealed that the coexistence of PIK3CA-TP53 mutations (termed 'PIK3CA+/TP53+'), as well as PIK3CA mutation alone (termed

'PIK3CA+/TP53−'), were associated with decreased survival in high-risk patients (HR 1.9, 95% CI 1.3-2.9, P value = 2.2E-03; HR 1.6, 95% CI 1.1-2.3, P value = 1.6E-02, respectively). This trend was not observed in low-risk patients (Figures 5C and D). These results provide strong evidence that TP53 and PIK3CA mutations may be valuable for guiding risk-adapted treatment strategies, particularly in high-risk patients.

To understand the biological impact of PIK3CA+/TP53− mutation, we analyzed gene expression profiles of high-risk patients from the Oncotype-like score model with PIK3CA+/TP53− mutation, compared with high-risk samples without mutations in the TP53 and PIK3CA genes (termed 'PIK3CA-/TP53−'). We identified four down-regulated genes and 15 up-regulated genes (logFC = 1, FDR ≤ 0.05) (Supplementary Table S13, available at https://doi.org/10.1016/j.esmoop.2025.105080). Among the up-regulated genes, some are known to be overexpressed in PIK3CA-mutated ERα-positive BC[60] (such as LTF, CYP4Z1, and TFAP2B), while others are associated with poor prognosis across various cancers[61-63] (including CLEC3A, PIP, and CALML5). These DEGs were used to conduct survival analysis based on their expression levels. As shown in Figure 5E and G, high expression levels of CLEC3A, CALML5, and TFAP2B were associated with poor overall survival (P value <0.05, FDR = 0.155). Taken together, these results suggest that PIK3CA mutation is associated with unfavorable outcomes in high-risk patients, potentially linked to high expression of CLEC3A, CALML5, and TFAP2B.

DISCUSSION

In this study, we investigated the association between specific genetic alterations with genomic prognostic scores and the likelihood of recurrence in patients with luminal BC. Utilizing publicly available data from 11 BC studies, we developed three transcriptome-based predictive models based on widely used clinical tests: Oncotype DX®, MammaPrint®, and PAM50-ROR. These models were employed to categorize TCGA[21,64,65] and METABRIC[31,66,67] patients with luminal BC into high- and low-recurrence-risk groups, revealing significant differences in recurrence rates between these groups.

Our findings demonstrated that high-risk patients as identified by our models exhibited significantly higher average recurrence rates (28%) compared with low-risk patients (8%). This aligns with recurrence rates observed in clinical trials, such as TAILORx for Oncotype DX[33,34] and MINDACT for MammaPrint,[36] albeit with some discrepancies potentially arising from variations in gene sets used for risk scoring in each prediction model.

When comparing our results with the existing literature,[68-70] our chosen cut-offs effectively distinguished between high- and low-risk patients, as evidenced by higher HRs in the high-risk groups across all models. This consistency underscores the robustness of our classification approach in predicting recurrence. Historically, mutations in genes like TP53[71,72] and RB1,[73,74] and CNAs such as amplifications in cytoband 14q21.1 including the FOXA1
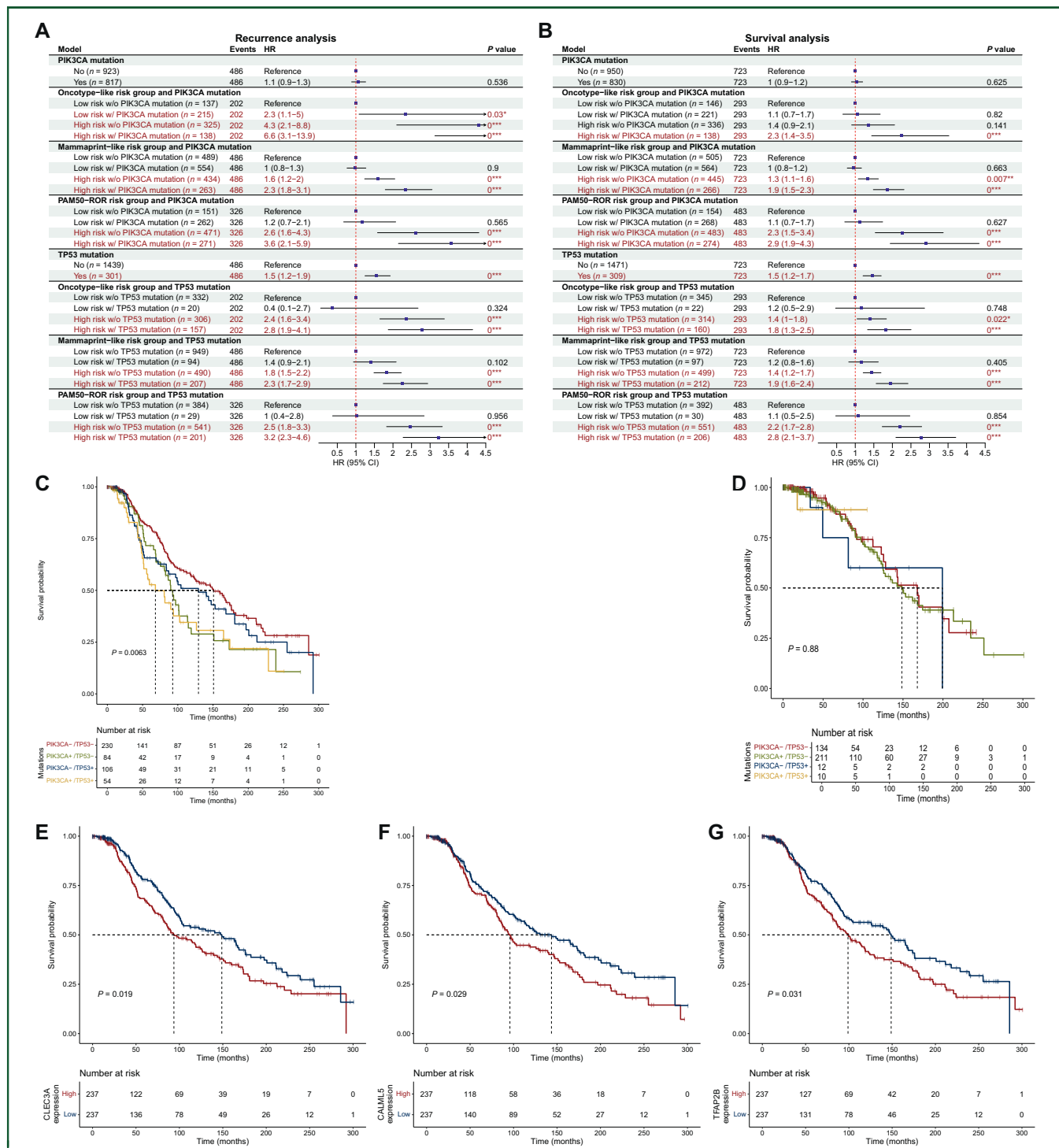
Figure 5. Survival of high-risk patients is affected by TP53 and PIK3CA mutations. (A, B) Univariate Cox proportional hazard regression analysis for distant recurrence and overall survival. HR was calculated for (A) recurrence, and (B) overall survival among patients with luminal BC from TCGA and METABRIC datasets, based on the mutations and risk groups in the different models. (C, D) Kaplan–Meier analysis of (C) high-risk and (D) low-risk patients in the Oncotype-like score model according to the presence of *TP53* and *PIK3CA* mutations. (E-G) Kaplan–Meier analysis for assessing overall survival in high-risk Oncotype-like model patients according to the expression of the (E) *CLEC3A*, (F) *CALML5*, and (G) *TFAP2B* genes. The groups were defined according to the median expression level of the gene in this cohort (*P ≤ 0.05, ***P ≤ 0.001).
CI, confidence interval; HR, hazard ratio.

gene[75,76] and deletions in cytoband 17p13.3 including the gene *RPH3AL*,[77,78] have been linked to poor prognosis in BC. Our study corroborated these findings and identified additional genomic aberrations associated with high relapse risk, including mutations in *PTPN22* and *MEN1* and

deletions in regions such as 10q26 and 22q11.1. Conversely, mutations in genes such as *PIK3CA*[79,80] and *MAP3K1*[80,81] were associated with low risk of recurrence. Notably, our observation that *PIK3CA* mutations were enriched in low-risk luminal samples suggests that they may serve as a

protective marker for overall survival, consistent with previous studies.[79,82] Our analysis also highlighted the coexistence of certain mutations (e.g. TP53-KMT2C, TP53-USH2A) that were more prevalent in high-risk groups, suggesting additive effects that could exacerbate disease progression. Similarly, mutations such as PIK3CA-CBFB and PIK3CA-CDH1, and amplification in 19q13.43, were linked to low risk, indicating a complex interplay of genetic factors that mitigate recurrence.

Typically, investigations into mutations or CNAs and their association with cancer recurrence risk use specific genomic aberrations or combinations to predict disease-free survival or examine frequently observed genomic aberrations in patient cohorts with known recurrence status.[25,71,80,83] However, our approach of focusing on the prevalence of specific mutations and CNAs in high- versus low-risk groups offers a novel perspective compared with traditional studies. For instance, we found TP53 and deletion CNA at 7p22.3 to have higher frequency ratios in the high-risk group across all three models.

Our findings reveal a context-dependent prognostic effect of PIK3CA mutations on survival. While PIK3CA mutations were not associated with prognosis in the overall luminal cohort, they significantly correlated with worse survival outcomes only in high-risk patients, particularly within the Oncotype-like model. This suggests that the negative impact of PIK3CA mutations on survival becomes evident in the presence of other high-risk factors. The underlying biological mechanisms remain unclear. The upregulation of genes such as CLEC3A, CALML5, and TFAP2B, associated with poor prognosis in various cancers, in patients with only PIK3CA mutations, suggests a complex genomic landscape influencing prognosis. In addition, multiple studies in metastatic BC demonstrate that PIK3CA mutation is associated with worse outcome.[84] Collectively, these factors underscore the multifaceted role of PIK3CA mutations in BC outcomes.

The potential clinical implications of our observations are noteworthy. Most of the genomic tests described in this study are well validated, with established cut-offs for clinical management and chemotherapy recommendations.[16] However, for premenopausal women or those younger than 50 years, the data are less definitive.[16] In this group, adjuvant chemotherapy is estimated to reduce recurrence risk by about 5%, but clear cut-offs are less established.

For young patients with TP53 mutations and ambiguous genomic test scores, it might be beneficial to escalate to adjuvant chemotherapy. Conversely, young patients with PIK3CA mutations may achieve better outcomes with hormonal treatment, combining an aromatase inhibitor and luteinizing hormone-releasing hormone agonist. This approach is based on evidence that these patients have lower benefits from chemotherapy but may gain more from letrozole compared with tamoxifen.[46]

Additionally, we validated here and in a previous study[85] the higher prevalence of BRCA2 mutations in Oncotype DX high-risk patients and their lower prevalence in Oncotype DX low-risk patients. While BRCA2 mutations are more frequent in the luminal B subtype,[86] our findings show a substantial proportion of luminal A patients (20%-30%) within the high-risk group. This suggests that risk classification based on our models extends beyond intrinsic subtyping and captures additional factors that contribute to recurrence risk.

To assess the robustness of our findings and eliminate potential bias from including model-building samples in the main analysis (337 and 643 samples out of 1780 were included in model building in Oncotype and MammaPrint-like models, respectively), we conducted a re-analysis excluding TCGA and METABRIC samples used for model development. This step was crucial to ensure that our results were not influenced by data leakage. The re-analysis showed that both the Oncotype-like and MammaPrint-like models maintained the same overall trend as in the original analysis, reinforcing their predictive value. The results, presented in Supplementary Figures S8-S16 and Table S14-S16, available at https://doi.org/10.1016/j.esmoop.2025.105080, confirm that our models generalize well beyond the training dataset. This consistency strengthens confidence in their applicability and suggests that the observed associations are not merely artifacts of the training data.

Our study provides valuable insights into the genetic factors contributing to the risk of recurrence in luminal BC. The identification of mutations and CNAs associated with high and low recurrence risks offers avenues for further research and the development of more refined prognostic models. Future studies should explore the functional impact of these genetic aberrations and their interactions, potentially leading to novel therapeutic targets for managing recurrence in patients with luminal BC.

DISCLOSURE

AS has received research grants from Novartis and Roche companies, and has an advisory role at Eli Lilly, Pfizer, Novartis, Roche, MSD, and AstraZeneca; travel expenses from Celgene, Medison, Roche, and MSD; was a speaker bureau at Teva, Roche, Pfizer, Novartis, and Eli Lilly; is supported by the Israel Science Foundation [grant number 3755/21] within the Israel Precision Medicine Partnership Program (IPMP). All other authors have declared no conflicts of interest.

DATA SHARING

The datasets used in this study are publicly available through the MetaGxBreast R package[37] (https://bioconductor.org/packages/release/data/experiment/html/MetaGxBreast.html), and cBioPortal[40,41] (https://www.cbioportal.org/study/summary?id=brca_tcga_pub2015%2Cbrca_tcga%2Cbrca_tcga_pub%2Cbrca_tcga_pan_can_atlas_2018%2

Cbrca_metabric). On request, the corresponding author can provide the analyzed data generated during this study.

REFERENCES

1. Xu S, Liu Y, Zhang T, et al. The global, regional, and national burden and trends of breast cancer from 1990 to 2019: results from the Global Burden of Disease Study 2019. *Front Oncol*. 2021;11:689562.
2. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209-249.
3. Henry NL, Shah PD, Haider I, Freer PE, Jagsi R, Sabel MS. Cancer of the breast. In: Niederhuber JE, Armitage JO, Kastan MB, Doroshow JH, Tepper JE, editors. *Abeloff's Clinical Oncology (Sixth Edition)*. Philadelphia: Elsevier; 20201560-1603.e12.
4. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73:17-48.
5. Ignatov A, Eggemann H, Burger E, Ignatov T. Patterns of breast cancer relapse in accordance to biological subtype. *J Cancer Res Clin Oncol*. 2018;144:1347-1355.
6. Kennecke H, Yerushalmi R, Woods R, et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol*. 2010;28:3271-3277.
7. Fina E, Cleris L, Dugo M, et al. Gene signatures of circulating breast cancer cell models are a source of novel molecular determinants of metastasis and improve circulating tumor cell detection in patients. *J Exp Clin Cancer Res*. 2022;41:78.
8. Tuttle M, Morris LF, Haugen B, et al. Thyroid-differentiated and anaplastic carcinoma. In: Amin MB, Edge SB, Greene FL, et al., editors. *AJCC Cancer Staging Manual.*. 8th ed. New York: Springer International Publishing; 2017.
9. Wallden B, Storhoff J, Nielsen T, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. 2015;8:54.
10. Spitale A, Mazzola P, Soldini D, Mazzucchelli L, Bordoni A. Breast cancer classification according to immunohistochemical markers: clinicopathologic features and short-term survival analysis in a population-based study from the South of Switzerland. *Ann Oncol*. 2009;20:628-635.
11. Tang P, Wang J, Bourne P. Molecular classifications of breast carcinoma with similar terminology and different definitions: are they the same? *Hum Pathol*. 2008;39:506-513.
12. Dai X, Li T, Bai Z, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*. 2015;5:2929-2943.
13. Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast*. 2015;24(suppl 2):S26-S35.
14. Ades F, Zardavas D, Bozovic-Spasojevic I, et al. Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *J Clin Oncol*. 2014;32:2794-2803.
15. Győrffy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res*. 2015;17:11.
16. Piccart MJ, Kalinsky K, Gray R, et al. Gene expression signatures for tailoring adjuvant chemotherapy of luminal breast cancer: stronger evidence, greater trust. *Ann Oncol*. 2021;32:1077-1082.
17. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994;266:66-71.
18. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature*. 1995;378:789-792.
19. Melchor L, Honrado E, García MJ, et al. Distinct genomic aberration patterns are found in familial breast cancer associated with different immunohistochemical subtypes. *Oncogene*. 2008;27:3165-3175.
20. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534:47-54.
21. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61-70.
22. Hu X, Stern HM, Ge L, et al. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. *Cell Growth Differ*. 2009;7:511-522.
23. Tripathy D, Harnden K, Blackwell K, Robson M. Next generation sequencing and tumor mutation profiling: are we ready for routine use in the oncology clinic? *BMC Med*. 2014;12:140.
24. Li G, Guo X, Chen M, et al. Prevalence and spectrum of AKT1, PIK3CA, PTEN and TP53 somatic mutations in Chinese breast cancer patients. *PLoS One*. 2018;13:e0203495.
25. Wang C-Y, Chang Y-C, Kuo Y-L, et al. Mutation of the PTCH1 gene predicts recurrence of breast cancer. *Sci Rep*. 2019;9:16359.
26. Weisman PS, Ng CKY, Brogi E, et al. Genetic alterations of triple negative breast cancer by targeted next-generation sequencing and correlation with tumor morphology. *Mod Pathol*. 2016;29:476-488.
27. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*. 2006;10:529-541.
28. Li Y, Zou L, Li Q, et al. Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat Med*. 2010;16:214-218.
29. Tofigh A, Suderman M, Paquet ER, et al. The prognostic ease and difficulty of invasive breast carcinoma. *Cell Rep*. 2014;9:129-142.
30. Schmidt M, Böhm D, von Törne C, et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res*. 2008;68:5405-5413.
31. Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486:346-352.
32. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-536.
33. Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res*. 2007;13:3207-3214.
34. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol*. 2007;25:1239-1246.
35. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A*. 2005;102:13550-13555.
36. Wang Y, Klijn JGM, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365:671-679.
37. Zon M, Gendoo DMA, Haibe-Kains B. MetaGxBreast: Transcriptomic Breast Cancer Datasets.
38. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
39. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401-404.
40. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:pl1.
41. Gendoo DMA, Ratanasirigulchai N, Schroeder MS, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics*. 2016;32:1097-1099.
42. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12:R41.
43. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010;463:899-905.
44. Beroukhim R, Getz G, Nghiemphu L, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A*. 2007;104:20007-20012.
45. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38:500-501.
46. Luen SJ, Asher R, Lee CK, et al. Association of somatic driver alterations with prognosis in postmenopausal, hormone receptor—positive,

HER2-negative early breast cancer: a secondary analysis of the BIG 1-98 randomized clinical trial. *JAMA Oncol*. 2018;4:1335-1343.

47. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841-842. https://doi.org/10.1093/bioinformatics/btq033.

48. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139-140.

49. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 2019;47:W199-W205.

50. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2022.

51. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.

52. Sievert Carson. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. New York: Chapman and Hall/CRC; 2020.

53. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. 2011;12:35.

54. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847-2849.

55. Therneau T. A Package for Survival Analysis in R. 2021. Available at https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf.

56. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using "ggplot2". 2021. Available at: https://cran.r-project.org/web/packages/survminer/survminer.pdf.

57. Dayimu A. Forestploter: Create Flexible Forest Plot. 2023. Available at https://cran.r-project.org/web/packages/forestploter/.

58. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;173:371-385.e18.

59. Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. *Nat Rev Cancer*. 2020;20:555-572.

60. Cizkova M, Cizeron-Clairac G, Vacher S, et al. Gene expression profiling reveals new aspects of PIK3CA mutation in ERalpha-positive breast cancer: major implication of the Wnt signaling pathway. *PLoS One*. 2010;5:e15647.

61. Ni J, Peng Y, Yang F-L, Xi X, Huang X-W, He C. Overexpression of CLEC3A promotes tumor progression and poor prognosis in breast invasive ductal cancer. *OncoTargets Ther*. 2018;11:3303-3312.

62. Naderi A. Prolactin-induced protein in breast cancer. *Adv Exp Med Biol*. 2015;846:189-200.

63. Jia G, Song Z, Xu Z, Tao Y, Wu Y, Wan X. Screening of gene markers related to the prognosis of metastatic skin cutaneous melanoma based on Logit regression and survival analysis. *BMC Med Genomics*. 2021;14:96.

64. Ciriello G, Gatza ML, Beck AH, et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*. 2015;163:506-519.

65. Liu J, Lichtenberg T, Hoadley KA, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018;173:400-416.e11.

66. Pereira B, Chin S-F, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*. 2016;7:11479.

67. Rueda OM, Sammut S-J, Seoane JA, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature*. 2019;567:399-404.

68. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med*. 2018;379:111-121.

69. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817-2826.

70. Cardoso F, Van'T Veer LJ, Bogaerts J, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375:717-729.

71. Ji JH, Bae SJ, Kim K, et al. Association between TP53 mutation and high 21-gene recurrence score in estrogen receptor-positive/HER2-negative breast cancer. *NPJ Breast Cancer*. 2022;8:19.

72. Fumagalli C, Ranghiero A, Gandini S, et al. Inter-tumor genomic heterogeneity of breast cancers: comprehensive genomic profile of primary early breast cancers and relapses. *Breast Cancer Res*. 2020;22:107.

73. Di Fiore R, D'Anneo A, Tesoriere G, Vento R. RB1 in cancer: different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis. *J Cell Physiol*. 2013;228:1676-1687.

74. Ahn S-M, Jang SJ, Shim JH, et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology*. 2014;60:1972-1982.

75. Katoh M, Igarashi M, Fukuda H, Nakagama H, Katoh M. Cancer genetics and genomics of human FOX family genes. *Cancer Lett*. 2013;328:198-206.

76. Fu X, Pereira R, De Angelis C, et al. FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. *Proc Natl Acad Sci*. 2019;116:26823-26834.

77. Emi M, Yoshimoto M, Sato T, et al. Allelic loss at 1p34, 13q12, 17p13.3, and 17q21.1 correlates with poor postoperative prognosis in breast cancer. *Genes Chromosomes Cancer*. 1999;26:134-141.

78. Putcha B-DK, Jia X, Katkoori VR, et al. Clinical implications of rabphillin-3A-like gene alterations in breast cancer. *PLoS One*. 2015;10(6):e0129216.

79. Kalinsky K, Jacks LM, Heguy A, et al. PIK3CA mutation associates with improved outcome in breast cancer. *Clin Cancer Res*. 2009;15:5049-5059.

80. Griffith OL, Spies NC, Anurag M, et al. The prognostic effects of somatic mutations in ER-positive breast cancer. *Nat Commun*. 2018;9:3476.

81. Carene D, Tran-Dien A, Lemonnier J, et al. Association between FGFR1 copy numbers, MAP3K1 mutations, and survival in axillary node-positive, hormone receptor-positive, and HER2-negative early breast cancer in the PACS04 and METABRIC studies. *Breast Cancer Res Treat*. 2020;179:387-401.

82. Loi S, Michiels S, Lambrechts D, et al. Somatic mutation profiling and associations with prognosis and trastuzumab benefit in early breast cancer. *J Natl Cancer Inst*. 2013;105:960-967.

83. Park JH, Kwon MJ, Seo J, Kim HY, Min SK, Kim LS. Somatic mutations of TP53 identified by targeted next-generation sequencing are poor prognostic factors for primary operable breast cancer: a single-center study. *J Breast Cancer*. 2022;25:379-386.

84. Fillbrunn M, Signorovitch J, André F, et al. PIK3CA mutation status, progression and survival in advanced HR + /HER2- breast cancer: a meta-analysis of published clinical trials. *BMC Cancer*. 2022;22:1002.

85. Halpern N, Sonnenblick A, Uziely B, et al. Oncotype Dx recurrence score among BRCA1/2 germline mutation carriers with hormone receptors positive breast cancer. *Int J Cancer*. 2017;140:2145-2149.

86. Waddell N, Arnold J, Cocciardi S, et al. Subtypes of familial breast tumours revealed by expression and copy number profiling. *Breast Cancer Res Treat*. 2010;123:661-677.