Molecular Therapy
Nucleic Acids
Original Article

Molecular classification of geriatric breast cancer displays distinct senescent subgroups of prognostic significance

Xia Wu,[1,2,11] Mengxin Chen,[1,11] Kang Liu,[3,11] Yixin Wu,[1,11] Yun Feng,[1] Shiting Fu,[1] Huaimeng Xu,[1] Yongqi Zhao,[1] Feilong Lin,[1] Liang Lin,[1] Shihui Ye,[1] Junqiang Lin,[1] Taiping Xiao,[1] Wenhao Li,[1] Meng Lou,[1] Hongyu Lv,[1] Ye Qiu,[4] Ruifan Yu,[1] Wenyan Chen,[5] Mengyuan Li,[6] Xu Feng,[7] Zhongbing Luo,[3] Lu Guo,[8] Hao Ke,[9,10] and Limin Zhao[1]

[1]Human Aging Research Institute (HARI) and School of Life Science, Nanchang University, and Jiangxi Key Laboratory of Human Aging, Nanchang 330031, China; [2]Ningbo Clinical Pathology Diagnosis Center, Ningbo, Zhejiang 315021, China; [3]Ganzhou People's Hospital, Ganzhou 341000, China; [4]Huankui Academy, Nanchang University, Nanchang, Jiangxi 330031, China; [5]Department of Medical Oncology, Nanchang People's Hospital, Nanchang 330008, China; [6]Department of Gynaecology and Obstetrics, Chongqing General Hospital, Chongqing 401147, China; [7]Xianlin High School, Weinan 714000, China; [8]Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou 310000, China; [9]Key Laboratory of Women's Reproductive Health of Jiangxi Province, Jiangxi Maternal and Child Health Hospital, Nanchang, Jiangxi 330006, China; [10]School of Life Science, Nanchang University, Nanchang 330031, China

Breast cancer in the elderly presents distinct biological characteristics and clinical treatment responses compared with cancer in younger patients. Comprehensive Geriatric Assessment is recommended for evaluating treatment efficacy in elderly cancer patients based on physiological classification. However, research on molecular classification in older cancer patients remains insufficient. In this study, we identified two subgroups with distinct senescent clusters among geriatric breast cancer patients through multi-omics analysis. Using various machine learning algorithms, we developed a comprehensive scoring model called "Sene_Signature," which more accurately distinguished elderly breast cancer patients compared with existing methods and better predicted their prognosis. The Sene_Signature was correlated with tumor immune cell infiltration, as supported by single-cell transcriptomics, RNA sequencing, and pathological data. Furthermore, we observed increased drug responsiveness in patients with a high Sene_Signature to treatments targeting the epidermal growth factor receptor and cell-cycle pathways. We also established a user-friendly web platform to assist investigators in assessing Sene_Signature scores and predicting treatment responses for elderly breast cancer patients. In conclusion, we developed a novel model for evaluating prognosis and therapeutic responses, providing a potential molecular classification that assists in the pre-treatment assessment of geriatric breast cancer.

INTRODUCTION

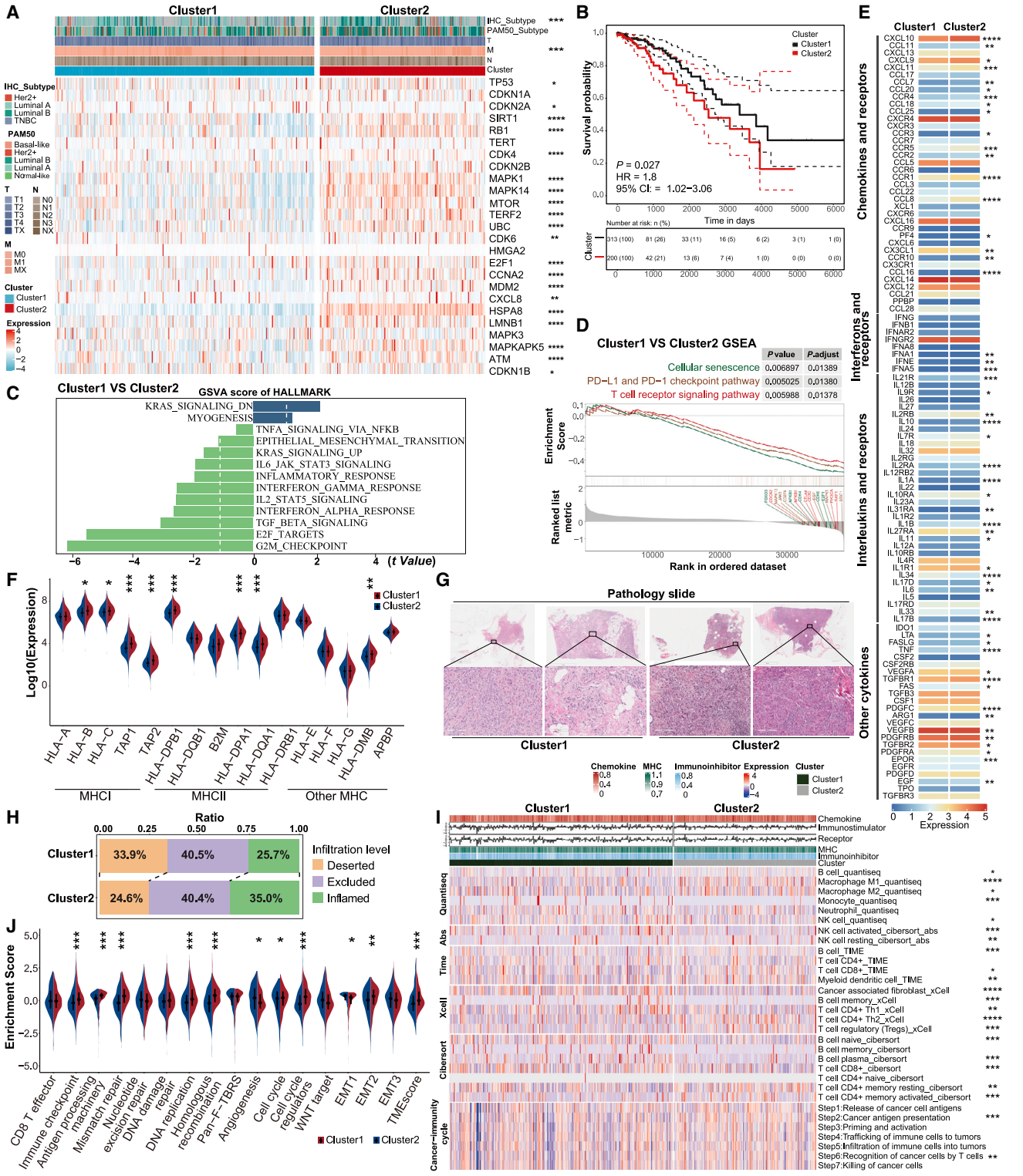
With an aging population and increased life expectancy, cancer prevalence among the elderly has increased significantly. Elderly patients often exhibit distinct biological characteristics and treatment responses compared with younger patients,[1] as well as a diminished tolerance to cancer therapies with age.[2] To address the historic failure in recognizing the impact of these differences on treatment decisions, prognosis, and tolerance, the National Comprehensive Cancer Network and the International Society of Geriatric Oncology have emphasized the importance of Comprehensive Geriatric Assessment (CGA) and other screening tools for elderly cancer patients,[3–5] which have greatly enhanced clinical treatment outcomes.[6,7] However, while CGA can effectively evaluate various physiological and physical health metrics, such as body mass index and the Groningen Frailty Indicator, it provides limited insights into molecular and cellular characteristics, which are crucial for predicting prognosis and drug responses in elderly cancer patients.

Breast cancer is among the prevalent malignancies in women. Significant advancements in treatments such as surgery, radiation therapy, chemotherapy, hormone therapy, and targeted therapy have greatly enhanced overall prognosis for breast cancer patients. Given the expanding cohort of elderly breast cancer patients, developing more precise classification criteria is essential for effective clinical treatment. Nevertheless, physiological age should not be the sole factor determining treatment protocols for geriatric breast cancer.[8,9] Cellular senescence, which reflects "age" at the molecular and cellular levels, is related to physiological age but also exhibits important differences.[10,11] Thus, using senescence as a molecular characteristic to

Received 5 March 2024; accepted 12 August 2024;
https://doi.org/10.1016/j.omtn.2024.102309.

[11]These authors contributed equally

Correspondence: Hao Ke, Key Laboratory of Women's Reproductive Health of Jiangxi Province, Jiangxi Maternal and Child Health Hospital, Nanchang, Jiangxi 330006, China.
E-mail: kehao@ncu.edu.cn
Correspondence: Limin Zhao, Human Aging Research Institute (HARI) and School of Life Science, Nanchang University, and Jiangxi Key Laboratory of Human Aging, Nanchang 330031, China.
E-mail: zhaolimin@ncu.edu.cn



(legend on next page)

classify elderly breast cancer patients may improve the accuracy of prognosis evaluation and drug sensitivity.

Understanding the molecular characteristics of elderly breast cancer patients is essential for making informed clinical treatment decisions. In this study, we developed the Sene_Signature model using large-scale databases, including the Cancer Genome Atlas (TCGA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), applying advanced machine learning techniques. Notably, the Sene_Signature model demonstrated superior performance compared with existing methods, providing robust prognosis predictions for geriatric breast cancer, as validated by multiple published datasets. Moreover, Sene_Signature proved to be a reliable indicator of therapeutic response, offering valuable theoretical guidance for personalized treatment strategies in elderly breast cancer patients (Figure S1). To enhance its practical application, we developed an accessible web platform (http://zhaoliminlab.cn:8080/GBC/index.jsp) that enables users to evaluate survival prognosis and drug responses for elderly breast cancer patients using their transcriptomic data.

RESULTS

Unsupervised clustering analysis reveals two distinct senescent subgroups in elderly breast cancer patients

To investigate the molecular characteristics of senescence in elderly breast cancer patients, we first identified key genes associated with cellular senescence. By analyzing their frequency of occurrence across various databases,[12–16] we identified 25 essential senescence-related genes (see methods for details). We analyzed the expression patterns and mutations of these 25 genes (Figures S2A–S2E) and found that most exhibited marked changes in geriatric breast cancer (Figure S2B). These findings suggest that these senescence-related genes may play crucial roles in the progression of geriatric breast cancer.

We next conducted nonnegative matrix factorization analysis on data from 513 elderly breast cancer patients in the TCGA database, focusing on the 25 senescence-related genes identified earlier, which divided the patients into 2 distinct senescent clusters via unsupervised clustering (Figures S3A–S3K; Table S1). Principal-component analysis further confirmed the differentiation of these clusters based on the expression patterns of the senescence-related genes (Figure S3L). Analysis showed that the expression of most senescence-related genes was higher in cluster 2 compared with cluster 1 (Figure S3M). A heatmap based on these genes demonstrated two distinct clusters, with higher gene expression in cluster 2, consistent with the violin plot results (Figure 1A). In terms of clinical breast cancer subtypes, cluster 1 predominantly included luminal subtype patients, while cluster 2 contained a higher proportion of triple-negative breast cancer (TNBC) and HER2-postive patients. Furthermore, the incidence of distant metastasis was higher in cluster 2 patients than in cluster 1 patients (Figure 1A). Survival analysis indicated poorer prognosis outcomes in cluster 2 patients (Figure 1B), as confirmed through univariate and multivariate Cox analyses (Figures S4A and S4B). In summary, we identified two distinct patterns of senescence in elderly breast cancer patients, highlighting significant differences in gene expression, clinical subtypes, and survival outcomes between the clusters.

Two senescent clusters exhibit distinct gene enrichment and immune landscapes

To investigate the disparities between the two senescent clusters, we performed gene set variation analysis (GSVA) using hallmark gene sets. Results revealed that cluster 2 was enriched in cell-cycle-related pathways, including G2M checkpoint, E2F targets, as well as in the NF-κB signaling pathway, TGF-β signaling pathway, and JAK-STAT signaling pathway (Figures 1C and S4C–S4E). In addition, the cellular senescence pathway was upregulated in cluster 2 (Figure 1D), consistent with the previously observed high expression of senescence-related genes in this cluster (Figure 1A). Cluster 2 was also enriched in immune-related pathways, such as the PD-L1 and PD-1 immune checkpoint pathways, and the T cell receptor signaling pathway (Figure 1D). Further examination of immune-related genes indicated higher expression of chemokines (e.g., CXCL10, CXCL11, and CXCL9) as well as interleukins and their receptors (e.g., IL-21R, IL-9R, and IL-10) in cluster 2 compared with cluster 1 (Figure 1E). In terms of tumor antigen presentation capacity, cluster 2 exhibited higher expression of MHC class I, MHC class II, and other MHC-related antigen presentation molecules (Figure 1F). Hematoxylin and eosin (H&E) staining of samples from elderly breast cancer patients (n = 440) in the TCGA cohort (Figure 1G) confirmed these findings, showing significantly higher levels of tumor-infiltrating lymphocytes in cluster 2 compared with cluster 1 (Figure 1H).

Using the deconvolution algorithm, we assessed the level of immune infiltration in elderly breast cancer patients based on TCGA

Figure 1. Unsupervised clustering analysis identified two distinct senescent clusters with different gene enrichment pathways in elderly breast cancer

(A) Complex heatmap showing expression levels of 25 genes for two senescent clusters. Clinical information corresponding to the top of the heatmap is also presented (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). (B) Kaplan-Meier survival curves for two distinct senescent clusters, with dashed lines indicating confidence intervals. (C) GSVA of hallmark pathways in clusters 1 and 2. T values are from a linear model, corrected for effects from patient origin. (D) Differential enrichment GSEA between clusters 1 and 2 (NES < 0, $p < 0.05$). (E) Expression levels of immune-inhibitory factors, interferons, chemokines and their receptors, and other cytokines in two senescent clusters (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). (F) Differential expression levels of MHC class I, MHC class II, and other MHC-related antigens in two distinct senescent clusters (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). (G) H&E staining was obtained for two distinct senescent clusters, corresponding to the following patient TCGA IDs: TCGA-E2-A14O, TCGA-AN-A0FW, TCGA-AR-A1AX, TCGA-E2-A14P. (H) Tumor-infiltrating lymphocytes between two distinct senescence clusters revealed significant differences. (I) Heatmap showing predicted proportions of immune cells using various immune infiltration algorithms (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). (J) Differential scores of immune-related gene sets, mismatch repair, and stromal scores between two senescent clusters using ssGSEA (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

transcriptome data. Results indicated that cluster 2 was enriched in infiltrating immune cells associated with immune exhaustion or suppression, including M2 macrophages, regulatory T (Treg) cells, resting natural killer cells, and resting memory T cells (Figure 1I). We also compared immune-related gene set scores, mismatch repair scores, and stromal scores between the two senescent clusters.[17] Consistent with previous findings, cluster 2 exhibited higher immune checkpoint features, mismatch repair, and tumor microenvironment (TME) scores (Figure 1J). These results indicate that two senescent clusters in elderly breast cancer display distinct patterns of gene enrichment and immune landscape, with cluster 2 showing a more pronounced immune response and higher expression of senescence-related pathways.

Development of the Sene_Signature model based on machine learning

To construct a scoring model to assess the degree of senescence in elderly breast cancer, DESeq2 was applied to identify differentially expressed genes (DEGs) ($|\log FC| > 1$ and false discovery rate (FDR) < 0.05) associated with the two distinct senescent clusters (Figure 2A), which are crucial for distinguishing between the two different senescence patterns. To avoid overfitting, accelerate training, and improve model interpretability, we performed feature selection using random forest, variance threshold, select from model, SelectKBest, and sequential forward selection. We also applied different machine learning algorithms, including GradientBoostingRegressor, RandomForestRegressor, OneClassDetection, LinearRegressor, KNeighborsRegressor, and SupportVectorRegressor, to determine the optimal model for evaluating molecular senescence in elderly cancer patients.

Based on various evaluation metrics, including area under curve, Matthews correlation coefficient, F1 score, positive predictive value, mean-square error, root mean-square error, mean absolute error, R-squared and adjusted R-squared (Table S2), the OneClassDetection senescence scoring model, Sene_Signature exhibited outstanding performance (Figures 2B and 2C). To evaluate the performance of the Sene_Signature model in predicting outcome in elderly cancer patients, univariate Cox analysis was conducted to compare the model with other currently published models, including SERPINE1_Signature,[18] 13_Epigenetic_Signature,[19] ASF1B_Signature,[20] SKP2/ORC6_Signature,[21] 8_DNA_Repair Signature,[22] and Immune_Score[23] (Figures 2D and S5). Notably, Sene_Signature demonstrated superior performance for elderly breast cancer patients. To further verify the feasibility of using senescence as a predictive measure for geriatric breast cancer prognosis, the scores of the Sene_Signature model derived from the senescence cluster were compared with those obtained from models based on established tumor hallmarks (angiogenesis, WNT target, CD8 T effector, DNA damage repair, cell cycle, and immune checkpoint),[17,24] confirming the marked impact of senescence on geriatric breast cancer prognosis (Figure 2D). Furthermore, the Sene_Signature model outperformed other models related to the senescence pathway[12–16] in predicting survival outcomes for elderly breast cancer patients (Figure 2D). A similar high performance of the Sene_Signature model was observed for the METABRIC database (Figure 2E).

In summary, we developed a machine learning-based model to assess senescence levels in elderly breast cancer patients. This model demonstrated superior prognostic performance, highlighting its robust applicability for this specific patient population.

Sene_Signature score correlates with clinical features

Comparing the Sene_Signature levels between the two distinct subtypes indicated that patients in cluster 2 had a significantly higher Sene_Signature score than those in cluster 1 (Figure 3A), consistent with the enrichment of senescence pathways in cluster 2 (Figure 1D). Further analysis revealed that 14 of the 25 senescence-related genes were positively correlated with the Sene_Signature in the TCGA database (Figure S6A), as also observed in the METABRIC dataset (Figure S6B). These results suggest that the Sene_Signature score is an effective measure of the two distinct molecular senescent clusters.

To further assess the clinical relevance of Sene_Signature in elderly breast cancer patients, the R package "survminer" was utilized to evaluate survival differences between high and low Sene_Signature groups. Results showed that patients with high scores had significantly worse survival outcomes than those with low scores (Figure 3B). This finding was confirmed in elderly cancer patients from the METABRIC and GSE2990 datasets (Figures 3C, 3D, and S6C). In addition, we found that high Sene_Signature scores were associated with more complex tumor subtypes (Figure 3E). Specifically, patients with higher scores tended to have a TNBC subtype characterized by increased heterogeneity (Figures 3E, 3F, S6D, and S6E). In addition, the high-scoring group had a higher proportion of patients in advanced T and M stages (Figures 3F and S6D). Overall, these findings indicate that the Sene_Signature score has clinical significance for elderly breast cancer patients.

Sene_Signature is positively correlated with TME infiltration

To explore the correlation between Sene_Signature and immunotherapy prediction, we analyzed differences in tumor mutational burden (TMB), which are indicative of immunotherapy effectiveness.[25–27] Results indicated that elderly breast cancer patients with a high Sene_Signature score exhibited a high TMB in both the TCGA and METABRIC datasets (Figures 3G and S7A). Further analysis indicated that Sene_Signature was significantly positively correlated with TMB (Figure 3H). Analysis of single-nucleotide site mutations in elderly breast cancer patients showed that those with higher scores displayed greater mutational diversity (Figures S7B and S7C). We utilized GISTIC2.0 to calculate the G-score and frequency in elderly breast cancer and found that patients with high scores exhibited a higher frequency of copy-number variation (CNV) mutations compared with those with low scores (Figure 3I).

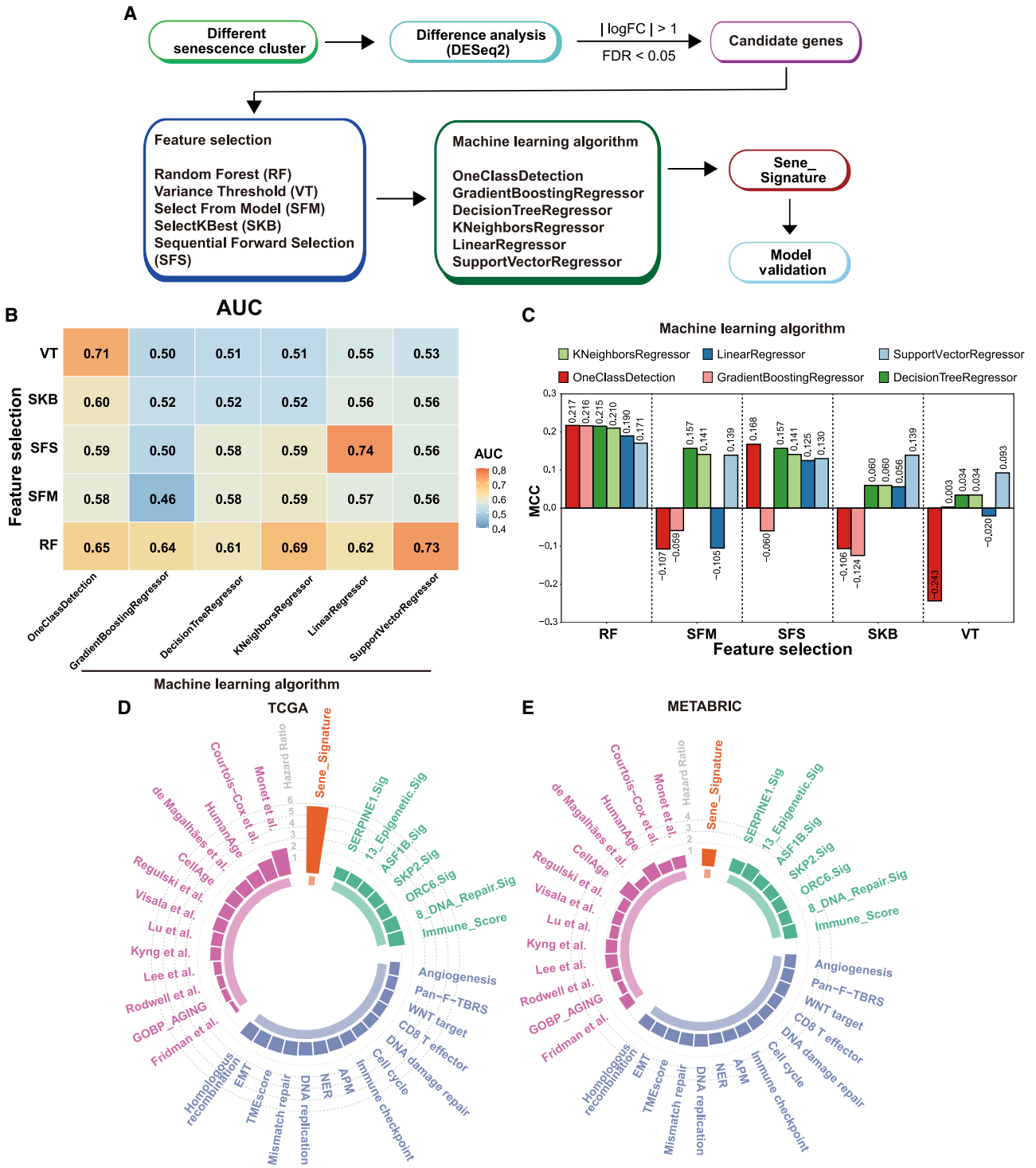


Figure 2. Construction of Sene_Signature model based on various machine learning methods

(A) Feature selection and model training flowchart. (B) Area under curve (AUC) curves were utilized to assess performance of various machine learning models. (C) Matthews correlation coefficient (MCC) was employed to evaluate model performance. (D and E) Sene_Signature model was compared with signature scores, risk assessment criteria, and senescence signature scores to assess impact on survival in the TCGA (D) and METABRIC (E) datasets.

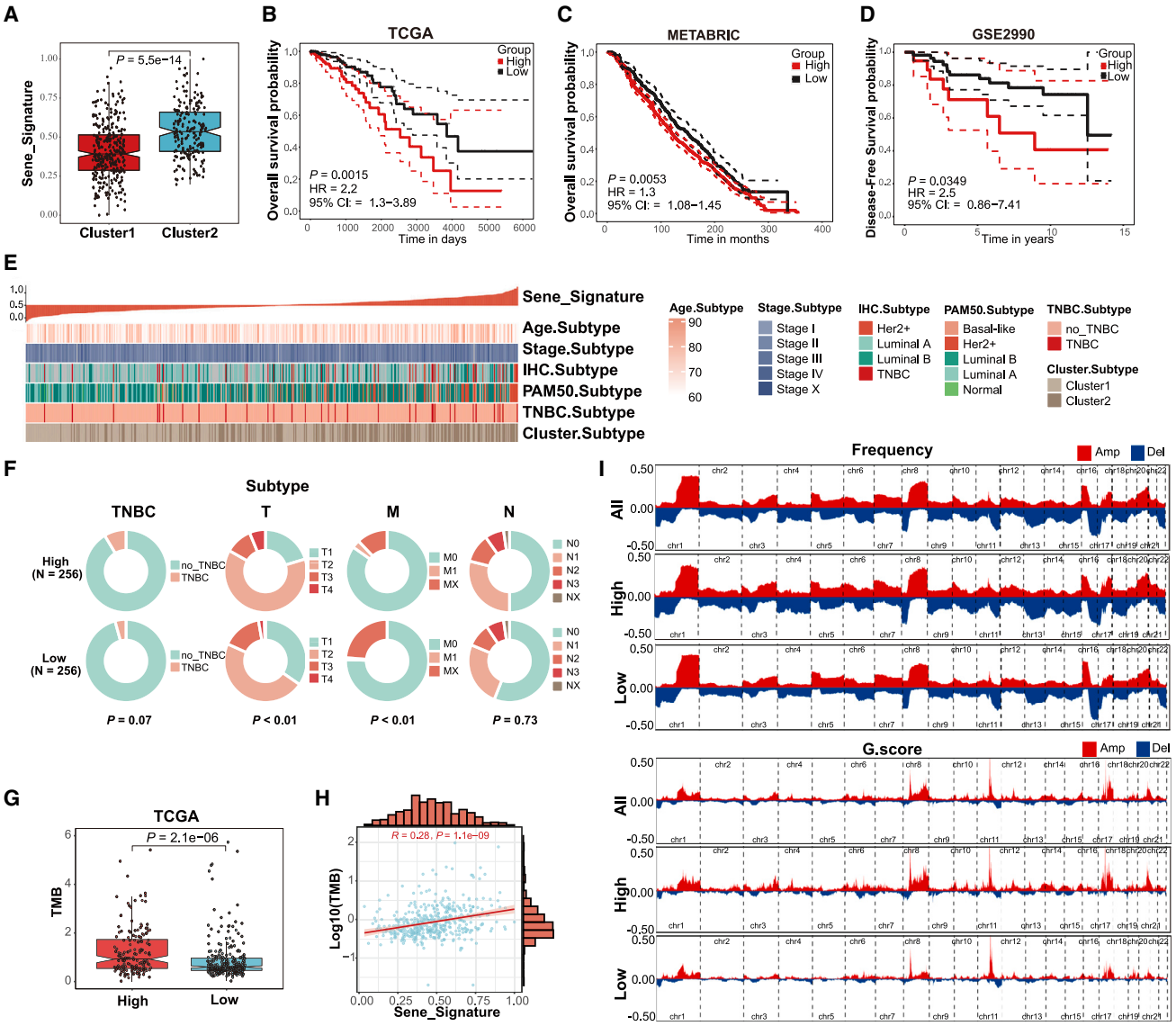


Figure 3. Sene_Signature score correlated with clinical features

(A) Boxplots of Sene_Signature for clusters 1 (red) and 2 (blue), indicating significant differences between both groups ($p < 0.05$). (B–D) Survival differences between high and low evaluation groups in the TCGA (B), METABRIC (C), and GSE2990 (D) datasets. Solid lines indicate median survival curves, while dashed lines represent 95% confidence intervals for these estimates, providing a measure of the precision around survival probability at any given time point. (E) Heatmap displays clinical information of patients with different Sene_Signature. Top is the numerical variable of Sene_Signature, which increases from left to right. Bottom is the heatmap of clinical data, including age subtype, stage subtype, IHC subtype, PAM50 subtype, triple-negative breast cancer (or not), and senescent clusters. (F) Clinical information differences between high- and low-score groups following division of TCGA patients based on median score. (G) Boxplot of TMB in the high- (red) and low-score (blue) groups in the TCGA dataset ($p < 0.05$). (H) Correlation between Sene_Signature and tumor mutational burden (TMB) ($p < 0.05$). (I) Analysis of copy-number variation (CNV) in high- and low-score groups of elderly breast cancer patients in TCGA dataset. The y axis represents frequency of CNVs and GISTIC score, the x axis represents genomic positions. Amplifications are denoted in red, and deletions are denoted in blue.

To further clarify the relationship between Sene_Signature and TME, the xCell algorithm[28,29] was applied to determine the proportions of immune cells in elderly breast cancer patients from the TCGA and METABRIC datasets. A positive correlation was observed between Sene_Signature score and immune cell infiltration, including immunosuppressive Treg cells (Figures 4A and 4B). To validate this observation, single-sample gene set enrichment analysis (ssGSEA) was performed to calculate enrichment scores for various pathways in patients.[17,24] Results demonstrated that Sene_Signature exhibited a positive correlation with TMEscore and with the enrichment score of immune checkpoint genes (Figures 4C and 4D). In addition, immune infiltration status of

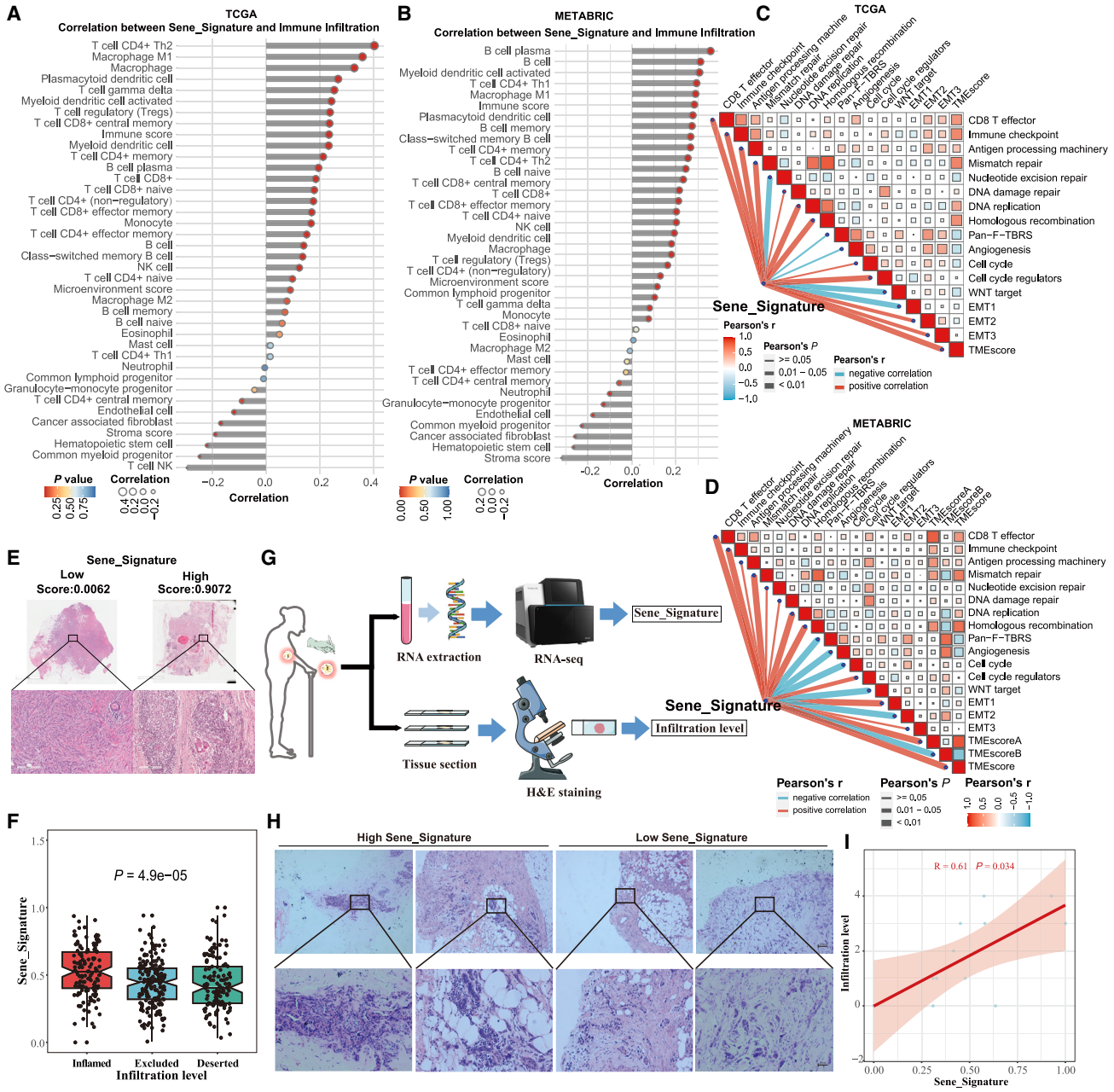
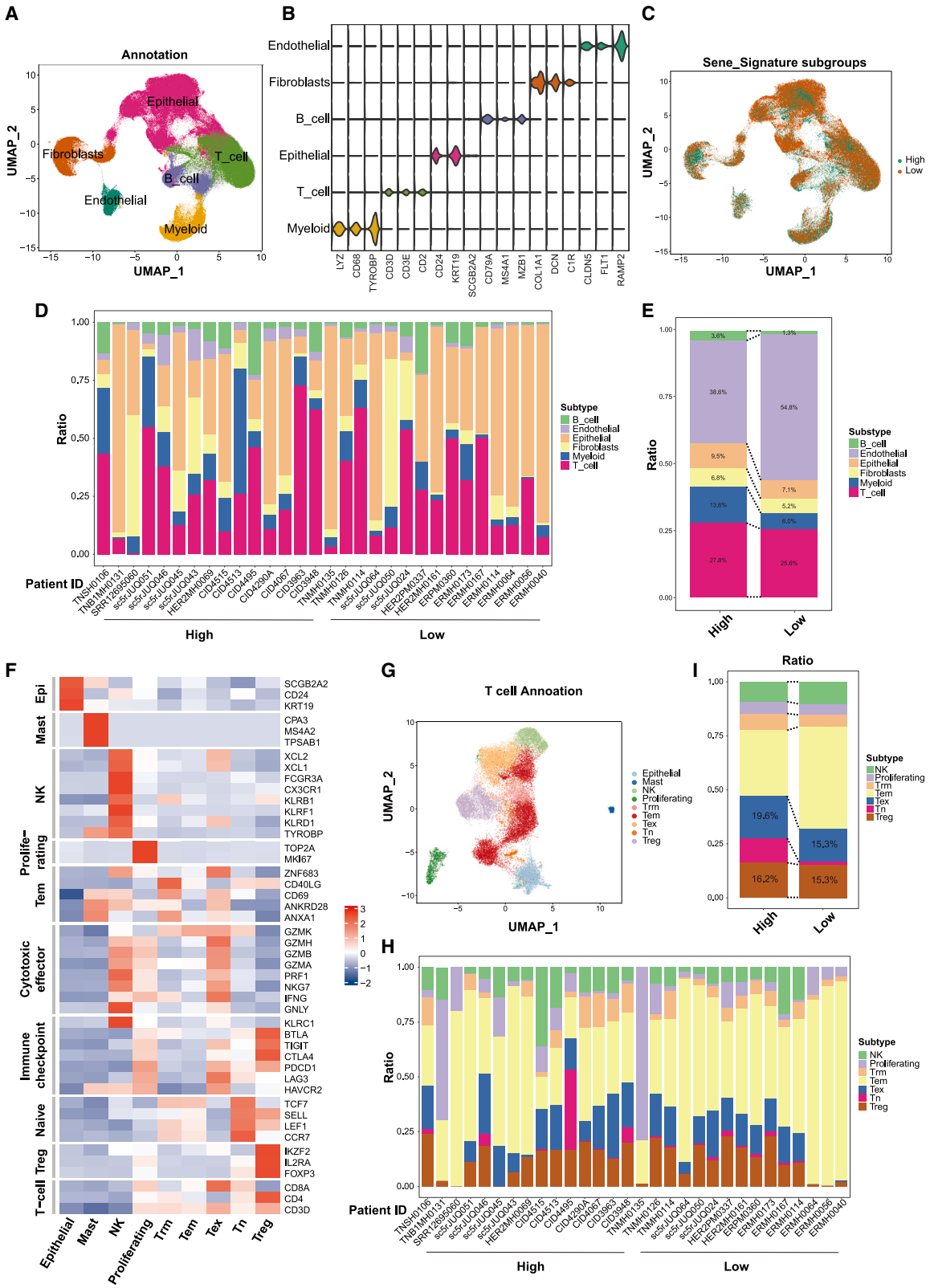


Figure 4. Sene_Signature positively correlated with TME infiltration

(A and B) Lollipop plot showing correlation between Sene_Signature and number of immune infiltrating cells in the TCGA (A) and METABRIC (B) datasets. Values greater than 0 indicate positive correlation, values less than 0 indicate negative correlation. Circle size represents correlation coefficient (R^2), and color represents significance (p value) of correlation. (C and D) Correlation between Sene_Signature and scores related to immune-related gene sets, mismatch repair, and stromal scores in the TCGA (C) and METABRIC (D) datasets. Upper-right part represents correlations between pathways, lower-left part represents correlations between Sene_Signature and pathways. Red indicates positive correlation, blue indicates negative correlation, and line thickness reflects Pearson correlation significance (p value). (E) Representative patients (TCGA-AC-A3QP, TCGA-D8-A1JG) were selected based on H&E staining results corresponding to Sene_Signature in elderly patients in the TCGA dataset. (F) Degree of immune infiltration in patients was categorized as inflamed, excluded, or deserted ($p < 0.05$). (G) Elderly patient breast cancer samples ($n = 12$) were collected for H&E staining and RNA-seq. (H) H&E-stained pathological slides were collected from eligible elderly breast cancer patients. (I) Pathology experts evaluated the H&E-stained pathological slides to quantify extent of immune infiltration. Correlation between Sene_Signature and immune infiltration was calculated by Spearman's correlation ($p < 0.05$).



(legend on next page)

patients was examined using H&E staining results from the TCGA, which showed that patients with a high Sene_Signature had more tumor-infiltrating lymphocytes than those with low scores (Figures 4E and 4F).

To confirm these findings, tumor tissues of elderly breast cancer patients (Figure 4G) were collected for RNA extraction and transcriptome sequencing to calculate the Sene_Signature for each patient. The corresponding samples were also preserved with 4% paraformaldehyde for histological staining to measure immune infiltration (Figure 4G). Results demonstrated that patients with a high Sene_Signature exhibited higher immune infiltration than patients with a low Sene_Signature (Figures 4H and 4I).

In summary, results indicated that Sene_Signature is closely related to the tumor immune microenvironment in elderly breast cancer patients.

Correlation between Sene_Signature and immune infiltration in scRNA-seq data

In addition to using the deconvolution method based on bulk-RNA sequencing (RNA-seq) results, we collected 78 single-cell transcriptome samples (totaling 352,450 cells) from the GEO database to determine immune cell distribution. These data were integrated using the "harmony" package and cell annotation was performed using the "Seurat" package to identify six cell types from breast cancer tumor tissues (Figures 5A, 5B, S8A, and S9A). Focusing on breast cancer samples from patients over the age of 60 years (30 patients), Sene_Signature subgroups were categorized based on gene expression (Table S3). Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction enabled clear visualization of cell data distribution from patients in different score subgroups (Figure 5C). We further analyzed the distribution of cell subtypes in each patient (Figure 5D) and found that the high-score group had a higher percentage of immune cells and an increased proportion of T cells compared with the low-score group (Figure 5E), consistent with previous findings (Figures 4A and 4B). We further classified T cell subtypes using UMAP dimensionality reduction, segregating the data into 12 clusters (Figures S9B and S9C) and identifying nine distinct T cell subtypes based on gene expression profiles, including naive, Treg, effector/memory, tissue-resident memory, exhausted (Tex), and proliferating T cells (Figures 5F and 5G). Results showed that patients in the high-score group had a higher proportion of Tex and Treg cells compared with those in the low-score group (Figures 5H and 5I). Overall, the single-cell transcriptome data suggest that patients with a higher Sene_Signature are associated with a greater percentage of immune-infiltrating cells, including immunosuppressive T cells.

Sene_Signature predicts drug sensitivity of elderly breast cancer

To investigate the impact of different senescence patterns on drug treatment, we analyzed the relationship between drug sensitivity and the expression of 25 senescence-related genes using Genomics of Drug Sensitivity in Cancer (GDSC) data. Results indicated that high expression of senescence-related genes was correlated with sensitivity to cell-cycle inhibitors and kinase inhibitors (Figure S10A; Table S4).

We next analyzed how the Sene_Signature model could guide drug treatment in elderly breast cancer patients. Patients were categorized into high and low Sene_Signature groups, and the "oncoPredict" package was used to predict the half-maximal inhibitory concentrations (IC_{50}). Differences in drug sensitivity between the high- and low-score groups were then assessed using limma. Results demonstrated that the high Sene_Signature group had lower IC_{50} values for cell-cycle and epidermal growth factor receptor (EGFR) inhibitors, suggesting greater sensitivity in the TCGA geriatric breast cancer patients to cell-cycle and EGFR pathway-targeted drugs (Figures 6A and 6B). In contrast, the low-score group exhibited lower IC_{50} values for ERK/MAPK inhibitors, indicating stronger responsiveness toward these inhibitors (Figures 6C, 6D, and S11A). Similar trends were observed in the METABRIC dataset (Figures 6E–6H and S11B). To further explore the relationship between Sene_Signature and drug sensitivity, we analyzed the correlation between predicted IC_{50} values and Sene_Signature scores, finding a significant negative correlation for cell-cycle inhibitors (AT7519 and THZ-2-49) (Figure 6I) and EGFR inhibitors (gefitinib and afatinib) (Figure 6J), implying higher responsiveness in the Sene_Signature group to treatments targeting the cell-cycle and EGFR pathways. Conversely, we found a significant positive correlation with ERK/MAPK inhibitors (AZ628 and selumetinib) (Figures 6K and S11C). These findings were consistent in the METABRIC dataset (Figures 6L–6N and S11D). To validate the clinical efficacy of drug treatments, we analyzed GEO data (GSE33658) from eight breast cancer patients over 60 years treated with the EGFR inhibitor gefitinib. We calculated the Sene_Signature for

Figure 5. Single-cell transcriptomic analysis of immune infiltration variations in patients with different Sene_Signature scores

(A) Integration of single-cell RNA-seq data from 78 breast cancer patients, followed by UMAP dimensionality reduction to obtain cell clustering results. (B) Annotation of cell subtypes, with violin plots showing expression of marker genes for each cell type. (C) Patients were grouped into high- and low-scoring categories based on Sene_Signature score model and mapped onto the UMAP results. (D) Bar chart depicting proportion of different cell types in each patient. Rows represent patient abbreviations and grouping information, while columns represent proportion of different cell types in total cells of each patient. (E) Bar chart showing cell proportions for different Sene_Signature score groups. Specific ratios are indicated on bar chart. (F) Annotation of cell subtypes, with a heatmap displaying mean expression of marker genes for each cell type. Higher expression tends toward red, while lower expression tends toward blue. (G) UMAP scatterplot displaying T cell subtypes, with distinct colors indicating various cell subclasses. (H) Bar chart illustrating proportion of different T cell subclasses in each patient. Patient identifiers and their respective groupings are shown as rows, with bar lengths representing relative abundance of each T cell subclass. (I) Bar chart showing cell proportions for different Sene_Signature score groups among T cell subclasses. Rows represent patient grouping information, while columns represent proportion of different T cell subclasses in total cells of each patient. Bar chart is annotated with specific ratio values.

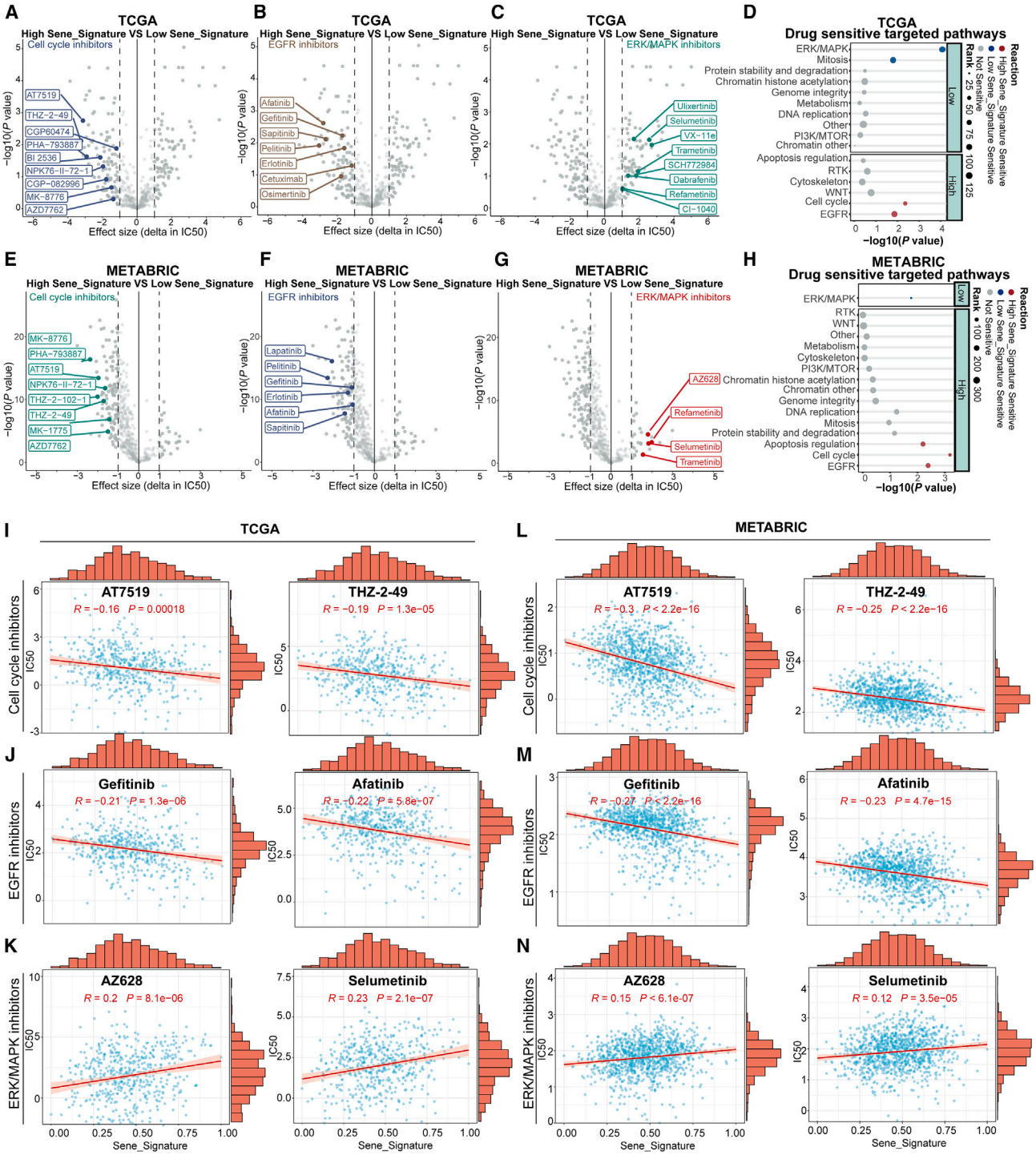


Figure 6. Sene_Signature predicted drug sensitivity in elderly breast cancer

(A–C) Differential analysis of IC_{50} values between high and low Sene_Signature groups in elderly breast cancer patients from the TCGA dataset. The x axis represents the delta in IC_{50} values, the y axis represents –log10(*p* value). Specifically marked points represent cell cycle inhibitors (A), EGFR inhibitors (B), and ERK/MAPK inhibitors (C) with (abs (delta in IC_{50}) > 1 and *p* value < 0.05). (D) Enrichment analysis of drug target pathways in high and low Sene_Signature groups in elderly breast cancer patients from the TCGA dataset. Drugs sensitive in the high-score group are shown in red, drugs sensitive in the low-score group are shown in blue, and non-significant enrichment results are shown

(legend continued on next page)

these patients and observed their response to the drug. Results showed that all patients in the high Sene_Signature group responded to EGFR inhibitor treatment, while only half of the low Sene_Signature group showed a response (Figure S12A). In conclusion, the Sene_Signature model has significant potential for guiding drug treatment in elderly breast cancer patients.

Development and use of the Sene_Signature website

We also developed a website (http://zhaoliminlab.cn:8080/GBC/index.jsp) to facilitate the application of Sene_Signature across different patient groups (Figure 7). The core feature of the site is the "Sene_Signature" module, which provides an input interface for external datasets. This module calculates the Sene_Signature for patients, allowing for analysis of immune infiltration, mutations, and clinical characteristics of patients based on model predictions. In addition, the site includes a "Drug" module that predicts drug sensitivity in elderly patients based on uploaded transcriptomic data, providing a potential basis for personalized treatment in older breast cancer patients. Overall, this platform offers a practical tool for researchers and clinicians to assess geriatric scores and explore potential treatment options.

DISCUSSION

In our study, we determined that genes associated with senescence can differentiate elderly breast cancer patients into two distinct senescent clusters through unsupervised clustering. Based on the characteristic genes of these subgroups, we constructed a regression model using machine learning, resulting in the creation of the Sene_Signature for these patients. We found a consistent correlation between the unsupervised clustering results and the high or low Sene_Signature scores (Figure 3A). Notably, patients in the high-score group exhibited a poorer survival prognosis, a higher tendency toward malignant subtypes, and a higher TMB compared with those in the low-score group. We also evaluated the relationship between Sene_Signature scores and immune infiltration in elderly breast cancer patients. Using the robust xCell method to deconvolute immune cell proportions from bulk RNA-seq,[28,29] elderly breast cancer patients with higher scores demonstrated a higher degree of immune infiltration, particularly of immunosuppressive cells (Figures 4A and 4B). These findings were corroborated by subsequent single-cell data analysis (Figure 5I), which revealed an increase in Tex and Treg cells in high-score patients. The higher presence of these cells likely contributes to an immunosuppressive TME, creating a senescence barrier around tumor cells and resulting in a worse prognosis for these patients.

In recent years, machine learning has emerged as a powerful tool, increasingly utilized for analyzing and predicting clinical disease data across a wide range of conditions,[30] including various cancers, chronic illnesses such as rheumatoid arthritis[31] and Alzheimer's disease,[32] and rare diseases such as shoulder dystocia.[33] OneClassDetection, a machine learning method based on data sample pairing,[34] has been used to construct a stemness feature that effectively assesses the dedifferentiated oncogenic state in tumor samples, revealing a novel correlation between cancer stemness and the TME.[35] In our study, we compared multiple machine learning approaches to evaluate prognosis of elderly cancer patients, ultimately finding that single-class logistic regression provided the most accurate results (Figure 2).

Based on drug response analysis, we found that elderly breast cancer patients with a high Sene_Signature may benefit from EGFR and cell-cycle inhibitors. Conversely, patients with a low Sene_Signature score may respond better to ERK/MAPK inhibitors. Although our study supported the potential application of EGFR inhibitors (Figure S12A), it is important to note that we lack validation data for other types of drugs. This limitation is primarily because drug response data often come from clinical trials, and much of this information is not publicly available. Many researchers are emphasizing the need for data sharing from clinical trials,[36,37] advocating for the development of standardized procedures and collaborative efforts to facilitate data sharing and acquisition. This would substantially accelerate research progress that relies on large data sources. Further studies and clinical data analysis are required to provide more robust guidance for drug therapy in elderly breast cancer patients.

In this study, we assessed mRNA expression features to construct a molecular predictor for elderly breast cancer. Other factors, such as immune TME,[38] mutations,[39] proteomics,[40,41] and pathology,[42] are also valuable for a more comprehensive understanding of cancer biology. Future research will focus on incorporating both clinical and genomic features to create a more precise classification system for elderly breast cancer.

In summary, we developed a scoring model that predicts prognosis and drug response in elderly breast cancer patients based on molecular senescence patterns, providing valuable insights into clinical treatment decisions in this demographic.

METHODS

Elderly breast cancer sample collection

Formaldehyde-fixed, paraffin-embedded BRCA tissue samples were obtained from the Ganzhou People's Hospital (Jiangxi, China) in 2023 from elderly breast cancer patients who had not received any therapy at the hospital. Based on previous studies,[43–45] tumor samples were collected from breast cancer patients aged 60 years and above.

in gray. (E–G) Differential analysis of IC_{50} values of cell-cycle inhibitors (E), EGFR inhibitors (F), and ERK/MAPK inhibitors (G) between high and low Sene_Signature groups in elderly breast cancer patients from the METABRIC dataset. (H) Enrichment analysis of drug target pathways in high and low Sene_Signature groups in elderly breast cancer patients from METABRIC. (I–N) Correlation analysis was conducted between drug sensitivity predictions and Sene_Signature scores in the TCGA (I–K) and METABRIC (L–N) datasets. Datasets were categorized into three major classes: cell cycle inhibitors (I and L), EGFR inhibitors (J and M), and ERK/MAPK inhibitors (K and N) ($p < 0.05$).

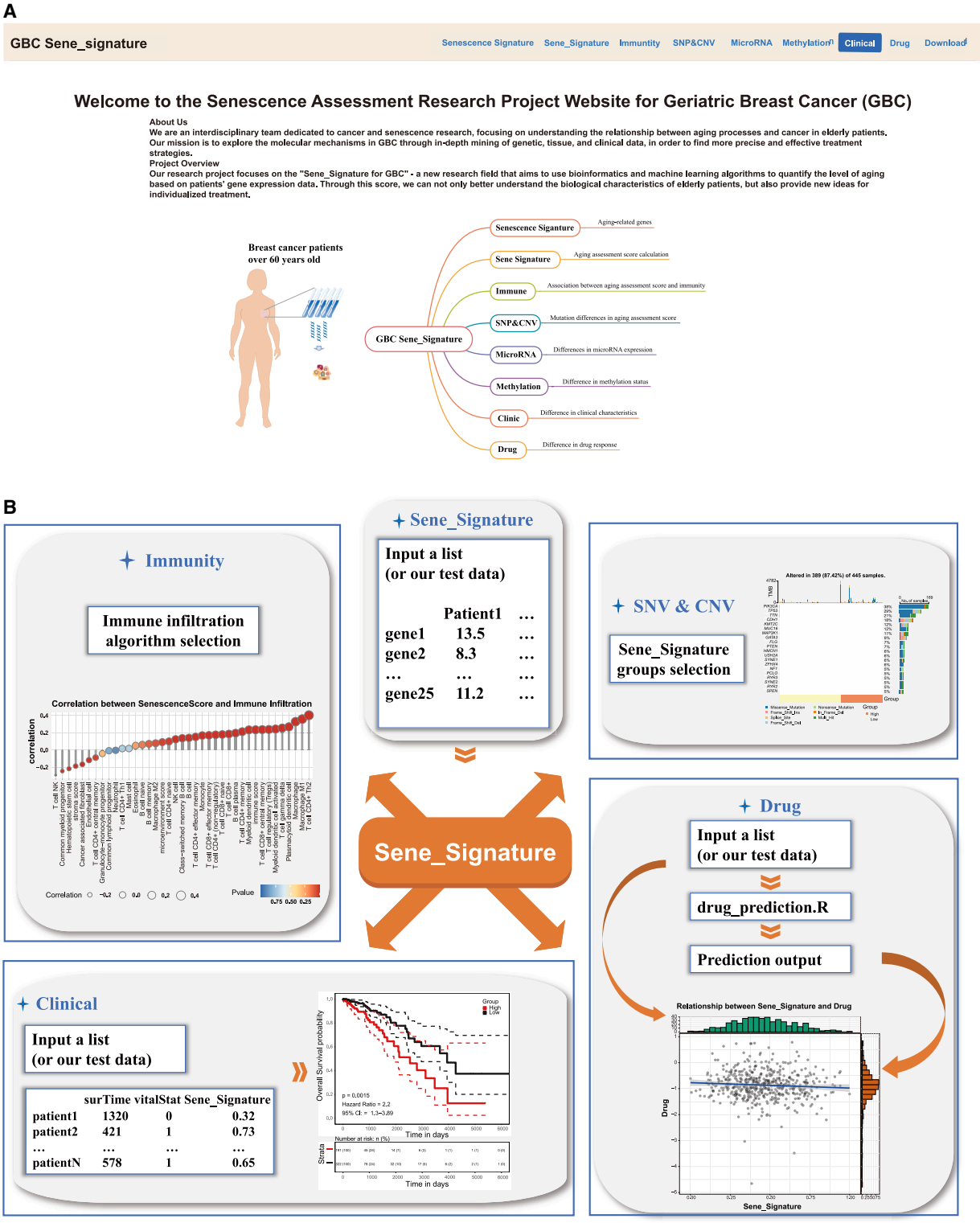


Figure 7. Development of our Sene_Signature website

(A) The main interface of our website related to Sene_Signature for elderly cancer patients. (B) The connection between the data modules of each interface, including immune infiltration, mutations, clinical characteristics and drug prediction.

The use of clinical samples was approved by the Ethics Committee of the Ganzhou People's Hospital (TY-ZKY2024-064-01).

RNA extraction and RNA-seq

Total RNA from tumor tissues was extracted using TRIzol reagent (TaKaRa), then converted to cDNA using a PrimeScript RT Reagent Kit (TaKaRa, containing gDNA Eraser). RNA integrity and purity were checked prior to library preparations. A NEBNext Ultra RNA Library Prep Kit for Illumina was used according to the manufacturer's recommendations to create sequencing libraries, which were then sequenced in 150-bp paired-end reads using the Illumina NovaSeq 6000 platform. The cleaned reads of each sample were aligned to their respective reference genomes (human: GRCh38) using HISAT2 software with default parameters.

H&E staining

Breast cancer tissue was fixed in formalin and embedded in paraffin. Tissue blocks were then sectioned (5 μm thickness) and stained with H&E. For immunostaining, sections were dehydrated using graded alcohol and subjected to antigen retrieval by boiling in 10 mM sodium citrate for 20 min, referring to previous research.[46] A microscope was used to observe the distribution of immune infiltration in patients at 5× and 20× magnification.

Data sources

In total, 1,094 samples were obtained from the Genomic Data Commons (GDC) portal (https://portal.gdc.cancer.gov/) and sorted according to the respective sample information sheet. According to published studies,[43–45] breast cancer patients older than 60 years were categorized as elderly, yielding 513 elderly breast cancer samples. Transcriptomic data (including counts and TPM values), clinical information, and pathological image data were downloaded using the R package "TCGAbiolinks." Furthermore, we obtained transcriptomic data and related clinical information from 1,105 elderly breast cancer patients from the METABRIC database (https://www.mercuriolab.umassmed.edu/metabric) for validation purposes. We also downloaded data from 75 elderly breast cancer patients in the GSE2990 dataset to validate model reliability. Single-cell transcriptomic data were collected from the GEO database, including GSE176078, GSE161529, GSE180286, GSE158399, and E-MTAB-8107 from the ArrayExpress database (https://www.ebi.ac.uk/biostudies/arrayexpress). After integrating the data and excluding metastatic tumor samples, we obtained single-cell transcriptomic data from 78 breast cancer patients, 30 of whom were over the age of 60 years. Data integration was performed using the R package "harmony." Transcriptional data from the GSE33658 cohort in the GEO database were used to validate drug responses. Various analyses were performed using the GSCA website (http://bioinfo.life.hust.edu.cn/GSCA/#/). All mentioned databases are publicly available and open source, and this study adhered to the access policies of each database.

Identification of 25 senescence-related genes

Genes related to senescence were compiled and analyzed from various gene set databases, including MSigDB,[12] HAGR,[13] GeneCards,[14] PathCards,[15] and Biocarta,[16] with their frequency of occurrence then assessed and ranked to select the top 25 most frequently appearing genes. The final gene list included TP53, CDKN1A, CDKN2A, SIRT1, RB1, TERT, CDK4, CDKN2B, MAPK1, MAPK14, MTOR, TERF2, UBC, CDK6, HMGA2, E2F1, CCNA2, MDM2, CXCL8, HSPA8, LMNB1, MAPK3, MAPKAPK5, ATM, and CDKN1B.

Construction of senescence patterns

The expression levels of the 25 key genes were assessed to determine their impact on breast cancer in elderly individuals. Unsupervised clustering analysis based on the expression levels of these senescence genes was performed using the R package "ConsensusClusterPlus" (reps = 50, pItem = 0.8, pFeature = 1, clusterAlg = "km", distance = "Euclidean") to determine the optimal clustering solution for patients, resulting in the identification of two distinct senescence clusters.

Immune infiltration analysis in elderly breast cancer patients

Information on immune modulators, including MHC, chemokines and their receptors, interferons and their receptors, interleukins and their receptors, and other cytokines, was obtained from various studies.[47–49] Enrichment scores for characteristic genes of the tumor immune microenvironment and tumor immune phenotype were calculated using ssGSEA (http://biocc.hrbmu.edu.cn/TIP/index.jsp). Estimations of immune cell quantities from transcriptomic data were performed using the R package "immunedeconv" through deconvolution.

GSEA

To gain a deeper understanding of the biological characteristics of elderly patients, gene enrichment analysis was conducted. Initially, the "msigdbr" R package was used to obtain hallmark gene sets and perform GSVA enrichment analysis on the expression matrix of elderly patients, obtaining gene enrichment scores for 50 pathways. Limma was then used to analyze differences between clusters 1 and 2, considering pathways with $p < 0.05$ as differentially enriched. To further explore differences between the two unsupervised clusters, GSEA was conducted for Kyoto Encyclopedia of Genes and Genomes pathways using the "clusterProfiler" package, providing information for subsequent functional interpretation.

Sene_Signature

To effectively measure distinct senescence patterns, we constructed a senescence scoring model to obtain a Sene_Signature, using machine learning to quantitatively assess the senescence status of elderly patients with breast cancer.

Initially, differences between the two senescent clusters were analyzed. The transcriptomic data of elderly patients with breast cancer were analyzed based on these clusters, using "limma" for differential analysis. DEGs were identified with a filtering parameter of $|\log FC| > 1$ and FDR < 0.05, resulting in the identification of two DEGs crucial for distinguishing between the two senescence clusters.

Random forest[50] was applied to calculate the mean decrease Gini (MDG) for the identified DEGs. Key genes with MDG > 1[51,52] were analyzed using the importance function, resulting in 120 feature genes (Table S5) that distinguished between the two senescent clusters. Subsequently, TCGA elderly breast cancer patients were divided into training and test sets at a 6:4 ratio. Regression models were then constructed using various machine learning methods, including GradientBoostingRegressor, RandomForestRegressor, OneClass Detection, LinearRegressor, KNeighborsRegressor, and SupportVector Regression. Among these, the senescence scoring model constructed by OneClassDetection was determined to be the best.

Finally, a senescence scoring model was created using the OneClassDetection machine learning algorithms. A logistic regression model was created using the gelnet function in the R package "Gelnet".

$$R(w) = \lambda_1 \sum_j d_j |w_j| + \frac{\lambda_2}{2}(w - m)^T P(w - m)$$

The weight of each gene was obtained and the Sene_Signature model was created.

CNV and single-nucleotide polymorphism analysis

The Sene_Signature model categorized elderly breast cancer patients into high and low groups. "Maftools" was used to create mutation waterfall plots for elderly breast cancer patients to visualize their mutation profiles. CNV analysis utilized GISTIC2 to identify amplified and deleted genomic sequences. GISTIC2.0 from MATLAB was downloaded and installed (ftp://ftp.broadinstitute.org/pub/GISTIC2.0/) using the human hg38 genome sequence as a reference (https://gdc.cancer.gov/about-data/gdc-data-processing/gdc-reference-files).

Single-cell transcriptome analysis

Single-cell transcriptomic data were imported into the R package "Seurat" (v.4.1.1) for quality control and filtering. Low-quality cells (<400 genes/cell, <3 cells/gene, >25% mitochondrial genes, >3% blood cell genes) were filtered, with "DoubletFinder" used to remove doublets. The "harmony" package was used for data integration across multiple single-cell datasets. Cell types identified through unsupervised clustering with UMAP were annotated using known marker genes: LYZ, CD68, TYROBP (myeloid cells), CD3D, CD3E, CD2 (T cell), CD24, KRT19, SCGB2A2 (epithelial cells), CD79A, MS4A1, MZB1 (B cells), COL1A1, DCN, C1R (fibroblasts), CLDN5, FLT1, and RAMP2 (endothelial cells). T cell subtype annotation was used, as proposed in previous research.[53]

Drug sensitivity analysis

Drug sensitivity analysis was conducted using publicly available drug databases, specifically referencing the GDSC, to predict drug sensitivity in elderly breast cancer patients. Subsequently, the "oncoPredict" R package was applied to predict IC_{50} values for elderly breast cancer patients. In addition, the limma method was employed to determine differential drug sensitivity between the high and low Sene_Signature groups. The enriched compound target pathways in GDSC were identified using the provided target pathway annotations for the differential drugs obtained through limma ($|\text{delta in } IC_{50}| > 1$ and $p < 0.05$).

Statistical analysis

Various statistical analyses were employed to assess associations within the data. The chi-squared test was applied to examine associations between categorical variables, while the Wilcoxon test or Student's t test was applied to evaluate significant differences in categorical variables. Pearson correlation coefficients were calculated to measure correlations between variables. Significance levels were defined as $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$. Survival differences were determined using the "survminer" R package, including univariate Cox regression analysis to assess the impact of gene expression on survival prognosis. Both univariate and multivariate analyses were performed to study the combined effects of variables on prognosis. All statistical and bioinformatics analyses were carried out using R (v.4.2.0).

Highly visual interactive web application

Based on the analysis data used in this study, an interactive web application (http://zhaoliminlab.cn:8080/GBC/index.jsp) was developed using "Tomcat" to allow researchers to explore the potential mechanisms of Sene_Signature at the multi-omics level. The web application includes several Sene_Signature analysis modules, including the signature expression module, calculation of external dataset Sene_Signature module, somatic mutation module, clinical prognosis module, microRNA module, methylation module, and drug sensitivity module. The source codes generated to build the web application are available at the GitHub repository (https://github.com/HARI-Zhaolab/GBC).

DATA AND CODE AVAILABILITY

The datasets presented in this study can be found in various online repositories. All raw RNA seq data generated from this study can be accessed in the NGDC database under accession number HRA006220. Codes used during analysis are available at https://github.com/HARI-Zhaolab/Sene_Signature. Additional data related to this paper can be requested from the authors.

ACKNOWLEDGMENTS

We thank the patients and researchers who participated in the TCGA, METABRIC, and GEO studies and provided publicly available data. We thank Drs. Xiyin Li, Leilei Cui, and Zhenying Hu for constructive suggestions. This work was supported by the National Natural Science Foundation of China (32360164, 82260488, and 32200679), the "Double Thousand Plan" of Jiangxi Province (jxsq2023101075), the Science Fund for Distinguished Young Scholars of Jiangxi Province (20232ACB215001), the Jiangxi Provincial Natural Science Foundation (20224BAB205014 and 20224BAB216071), the Natural Science Foundation of Chongqing (CSTB2022NSCQ-MSX056), Beijing Science And Technology Innovation Medical Development Foundation (KC2023-JX-0082-06), and the College Students' Innovative Entrepreneurial Training Plan Program of Nanchang University.

AUTHOR CONTRIBUTIONS

X.W. and H.K. designed the study and wrote the manuscript. L.Z. designed the study and wrote and revised the manuscript. X.W., M.C., and Y.W. performed most data analyses and interpreted the data. K.L. and Z.L. performed clinical sample collection and data curation. W.C. provided patient diagnosis for classification review. Y.F., S.F., H.X., T.X., W.L., M. Lou, H.L., Y.Q., and R.Y. provided experimental technical assistance. Y.Z., F.L., L.L., S.Y., and J.L. offered critical feedback and helped shape the study and manuscript. M. Li, X.F., and L.G. helped perform analysis and provided constructive discussions. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2024.102309.

REFERENCES

1. Van Herck, Y., Feyaerts, A., Alibhai, S., Papamichael, D., Decoster, L., Lambrechts, Y., Pinchuk, M., Bechter, O., Herrera-Caceres, J., Bibeau, F., et al. (2021). Is cancer biology different in older patients? Lancet Healthy Longev. *2*, e663–e677.
2. Wildiers, H., Kunkler, I., Biganzoli, L., Fracheboud, J., Vlastos, G., Bernard-Marty, C., Hurria, A., Extermann, M., Girre, V., Brain, E., et al. (2007). Management of breast cancer in elderly individuals: recommendations of the International Society of Geriatric Oncology. Lancet Oncol. *8*, 1101–1115.
3. Biganzoli, L., Battisti, N.M.L., Wildiers, H., McCartney, A., Colloca, G., Kunkler, I.H., Cardoso, M.-J., Cheung, K.-L., de Glas, N.A., Trimboli, R.M., et al. (2021). Updated recommendations regarding the management of older patients with breast cancer: a joint paper from the European Society of Breast Cancer Specialists (EUSOMA) and the International Society of Geriatric Oncology (SIOG). Lancet Oncol. *22*, e327–e340.
4. Dotan, E., Walter, L.C., Browner, I.S., Clifton, K., Cohen, H.J., Extermann, M., Gross, C., Gupta, S., Hollis, G., Hubbard, J., et al. (2021). NCCN Guidelines® Insights: Older Adult Oncology, Version 1.2021. J. Natl. Compr. Cancer Netw. *19*, 1006–1019.
5. Biganzoli, L., Wildiers, H., Oakman, C., Marotti, L., Loibl, S., Kunkler, I., Reed, M., Ciatto, S., Voogd, A.C., Brain, E., et al. (2012). Management of elderly patients with breast cancer: updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). Lancet Oncol. *13*, e148–e160.
6. Kenis, C., Decoster, L., Van Puyvelde, K., De Grève, J., Conings, G., Milisen, K., Flamaing, J., Lobelle, J.P., and Wildiers, H. (2014). Performance of two geriatric screening tools in older patients with cancer. J. Clin. Oncol. *32*, 19–26.
7. Kenis, C., Bron, D., Libert, Y., Decoster, L., Van Puyvelde, K., Scalliet, P., Cornette, P., Pepersack, T., Luce, S., Langenaeken, C., et al. (2013). Relevance of a systematic geriatric screening and assessment in older patients with cancer: results of a prospective multicentric study. Ann. Oncol. *24*, 1306–1312.
8. Extermann, M., Brain, E., Canin, B., Cherian, M.N., Cheung, K.L., de Glas, N., Devi, B., Hamaker, M., Kanesvaran, R., Karnakis, T., et al. (2021). Priorities for the global advancement of care for older adults with cancer: an update of the International Society of Geriatric Oncology Priorities Initiative. Lancet Oncol. *22*, e29–e36.
9. Wildiers, H., and Brain, E. (2015). Different adjuvant chemotherapy regimens in older breast cancer patients? Ann. Oncol. *26*, 613–615.
10. Hernandez-Segura, A., Nehme, J., and Demaria, M. (2018). Hallmarks of Cellular Senescence. Trends Cell Biol. *28*, 436–453.
11. Di Micco, R., Krizhanovsky, V., Baker, D., and d'Adda di Fagagna, F. (2021). Cellular senescence in ageing: from mechanisms to therapeutic opportunities. Nat. Rev. Mol. Cell Biol. *22*, 75–95.
12. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.
13. Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E., and de Magalhães, J.P. (2013). Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Res. *41*, D1027–D1033.
14. Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., et al. (2010). GeneCards Version 3: the human gene integrator. Database *2010*, baq020.
15. Belinky, F., Nativ, N., Stelzer, G., Zimmerman, S., Iny Stein, T., Safran, M., and Lancet, D. (2015). PathCards: multi-source consolidation of human biological pathways. Database *2015*, bav006.
16. Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database *2016*, baw100.
17. Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature *554*, 544–548.
18. Foekens, J.A., Schmitt, M., van Putten, W.L., Peters, H.A., Kramer, M.D., Jänicke, F., and Klijn, J.G. (1994). Plasminogen activator inhibitor-1 and prognosis in primary breast cancer. J. Clin. Oncol. *12*, 1648–1658.
19. Bao, X., Anastasov, N., Wang, Y., and Rosemann, M. (2019). A novel epigenetic signature for overall survival prediction in patients with breast cancer. J. Transl. Med. *17*, 380.
20. Corpet, A., De Koning, L., Toedling, J., Savignoni, A., Berger, F., Lemaître, C., O'Sullivan, R.J., Karlseder, J., Barillot, E., Asselain, B., et al. (2011). Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer. EMBO J. *30*, 480–493.
21. Ibrahim, A., Toss, M.S., Alsaleem, M., Makhlouf, S., Atallah, N., Green, A.R., and Rakha, E.A. (2024). Novel 2 Gene Signatures Associated With Breast Cancer Proliferation: Insights From Predictive Differential Gene Expression Analysis. Mod. Pathol. *37*, 100403.
22. Zhang, D., Yang, S., Li, Y., Yao, J., Ruan, J., Zheng, Y., Deng, Y., Li, N., Wei, B., Wu, Y., et al. (2020). Prediction of Overall Survival Among Female Patients With Breast Cancer Using a Prognostic Signature Based on 8 DNA Repair-Related Genes. JAMA Netw. Open *3*, e2014622.
23. Sui, S., An, X., Xu, C., Li, Z., Hua, Y., Huang, G., Sui, S., Long, Q., Sui, Y., Xiong, Y., et al. (2020). An immune cell infiltration-based immune score model predicts prognosis and chemotherapy effects in breast cancer. Theranostics *10*, 11938–11949.
24. Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., Bin, J., Liao, Y., Rao, J., and Liao, W. (2019). Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. Cancer Immunol. Res. *7*, 737–750.
25. Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N. Engl. J. Med. *377*, 2500–2501.
26. Osipov, A., Lim, S.J., Popovic, A., Azad, N.S., Laheru, D.A., Zheng, L., Jaffee, E.M., Wang, H., and Yarchoan, M. (2020). Tumor Mutational Burden, Toxicity, and Response of Immune Checkpoint Inhibitors Targeting PD(L)1, CTLA-4, and Combination: A Meta-regression Analysis. Clin. Cancer Res. *26*, 4842–4851.
27. Yarchoan, M., Albacker, L.A., Hopkins, A.C., Montesion, M., Murugesan, K., Vithayathil, T.T., Zaidi, N., Azad, N.S., Laheru, D.A., Frampton, G.M., and Jaffee, E.M. (2019). PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. JCI insight *4*, e126908.
28. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.
29. Aran, D., and Butte, A.J. (2016). Digitally deconvolving the tumor microenvironment. Genome Biol. *17*, 175.
30. Norgeot, B., Glicksberg, B.S., and Butte, A.J. (2019). A call for deep-learning healthcare. Nat. Med. *25*, 14–15.
31. Norgeot, B., Glicksberg, B.S., Trupin, L., Lituiev, D., Gianfrancesco, M., Oskotsky, B., Schmajuk, G., Yazdany, J., and Butte, A.J. (2019). Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA Netw. Open *2*, e190606.

32. Rodriguez, S., Hug, C., Todorov, P., Moret, N., Boswell, S.A., Evans, K., Zhou, G., Johnson, N.T., Hyman, B.T., Sorger, P.K., et al. (2021). Machine learning identifies candidates for drug repurposing in Alzheimer's disease. Nat. Commun. *12*, 1033.

33. Tsur, A., Batsry, L., Toussia-Cohen, S., Rosenstein, M.G., Barak, O., Brezinov, Y., Yoeli-Ullman, R., Sivan, E., Sirota, M., Druzin, M.L., et al. (2020). Development and validation of a machine-learning model for prediction of shoulder dystocia. Ultrasound Obstet. Gynecol. *56*, 588–596.

34. Sokolov, A., Paull, E.O., and Stuart, J.M. (2016). One-class detection of cell states in tumor subtypes. Pac. Symp. Biocomput. *21*, 405–416.

35. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell *173*, 338–354.e15.

36. Geifman, N., Bollyky, J., Bhattacharya, S., and Butte, A.J. (2015). Opening clinical trial data: are the voluntary data-sharing portals enough? BMC Med. *13*, 280.

37. Butte, A.J. (2021). Trials and Tribulations-11 Reasons Why We Need to Promote Clinical Trials Data Sharing. JAMA Netw. Open *4*, e2035043.

38. Damrauer, J.S., Roell, K.R., Smith, M.A., Sun, X., Kirk, E.L., Hoadley, K.A., Benefield, H.C., Iyer, G., Solit, D.B., Milowsky, M.I., et al. (2021). Identification of a Novel Inflamed Tumor Microenvironment Signature as a Predictive Biomarker of Bacillus Calmette-Guérin Immunotherapy in Non-Muscle-Invasive Bladder Cancer. Clin. Cancer Res. *27*, 4599–4609.

39. Damrauer, J.S., Beckabir, W., Klomp, J., Zhou, M., Plimack, E.R., Galsky, M.D., Grivas, P., Hahn, N.M., O'Donnell, P.H., Iyer, G., et al. (2022). Collaborative study from the Bladder Cancer Advocacy Network for the genomic analysis of metastatic urothelial cancer. Nat. Commun. *13*, 6658.

40. Creighton, C.J. (2020). Proteomic signatures of clear cell renal cell carcinoma. Nat. Rev. Nephrol. *16*, 133–134.

41. Zhang, Y., Chen, F., Chandrashekar, D.S., Varambally, S., and Creighton, C.J. (2022). Proteogenomic characterization of 2002 human cancers reveals pan-cancer molecular subtypes and associated pathways. Nat. Commun. *13*, 2669.

42. Rakha, E.A., and Pareja, F.G. (2021). New Advances in Molecular Breast Cancer Pathology. Semin. Cancer Biol. *72*, 102–113.

43. Abdel-Razeq, H., Abu Rous, F., Abuhijla, F., Abdel-Razeq, N., and Edaily, S. (2022). Breast Cancer in Geriatric Patients: Current Landscape and Future Prospects. Clin. Interv. Aging *17*, 1445–1460.

44. San Miguel, Y., Gomez, S.L., Murphy, J.D., Schwab, R.B., McDaniels-Davidson, C., Canchola, A.J., Molinolo, A.A., Nodora, J.N., and Martinez, M.E. (2020). Age-related differences in breast cancer mortality according to race/ethnicity, insurance, and socioeconomic status. BMC Cancer *20*, 228.

45. Ma, C.-D., Zhou, Q., Nie, X.-Q., Liu, G.-Y., Di, G.-H., Wu, J., Lu, J.-S., Yang, W.-T., Chen, J.-Y., Shao, Z.-M., et al. (2009). Breast cancer in Chinese elderly women: pathological and clinical characteristics and factors influencing treatment patterns. Crit. Rev. Oncol. Hematol. *71*, 258–265.

46. Zhao, L., Ke, H., Xu, H., Wang, G.-D., Zhang, H., Zou, L., Xiang, S., Li, M., Peng, L., Zhou, M., et al. (2020). TDP-43 facilitates milk lipid secretion by post-transcriptional regulation of Btn1a1 and Xdh. Nat. Commun. *11*, 341.

47. Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. *18*, 248–262.

48. Pal, B., Chen, Y., Vaillant, F., Capaldo, B.D., Joyce, R., Song, X., Bryant, V.L., Penington, J.S., Di Stefano, L., Tubau Ribera, N., et al. (2021). A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J. *40*, e107333.

49. Xiao, Y., Ma, D., Zhao, S., Suo, C., Shi, J., Xue, M.-Z., Ruan, M., Wang, H., Zhao, J., Li, Q., et al. (2019). Multi-Omics Profiling Reveals Distinct Microenvironment Characterization and Suggests Immune Escape Mechanisms of Triple-Negative Breast Cancer. Clin. Cancer Res. *25*, 5002–5014.

50. Breiman, L. (2001). Random Forests. Mach. Learn. *45*, 5–32.

51. Shen, M., Xie, Q., Zhang, R., Yu, C., and Xiao, P. (2023). Metabolite-assisted models improve risk prediction of coronary heart disease in patients with diabetes. Front. Pharmacol. *14*, 1175021.

52. Liang, Y., Liang, Z., Huang, J., Jia, M., Liu, D., Zhang, P., Fang, Z., Hu, X., and Li, H. (2023). Identification and validation of aging-related gene signatures and their immune landscape in diabetic nephropathy. Front. Med. *10*, 1158166.

53. Bassez, A., Vos, H., Van Dyck, L., Floris, G., Arijs, I., Desmedt, C., Boeckx, B., Vanden Bempt, M., Nevelsteen, I., Lambein, K., et al. (2021). A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. Nat. Med. *27*, 820–832.